SPRINGER REFERENCE

Thomas Schramme
Steven Edwards
Editors

Handbook of the Philosophy of Medicine

Springer

Handbook of the Philosophy of Medicine

Thomas Schramme • Steven Edwards
Editors

Handbook of the Philosophy of Medicine

With 14 Figures and 8 Tables

Springer

Editors
Thomas Schramme
Department of Philosophy
University of Liverpool
Liverpool, UK

Steven Edwards
Philosophy, History and Law
Swansea University
Swansea, UK

ISBN 978-94-017-8687-4 ISBN 978-94-017-8688-1 (eBook)
ISBN 978-94-017-8689-8 (print and electronic bundle)
DOI 10.1007/978-94-017-8688-1

Library of Congress Control Number: 2016950589

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media B.V.
The registered company address is: Van Godewijckstraat 30, 3311 GX Dordrecht, The Netherlands

Preface

Philosophy of medicine is a subject that has been around since the beginning of medicine but has only fairly recently, roughly in the last 40 years, been professionally developed into a discipline in its own right. It has gained a stronger status in relation to medical ethics or bioethics, which focuses on moral issues in medicine, whereas philosophy of medicine has a broader and less applied remit, addressing metaphysical, epistemological, and other philosophical issues in medicine.

There are now dedicated societies and academic centers dealing with different topics in philosophy of medicine. This interest is continuously increasing, not least because it has become obvious that several issues in bioethics are based on more theoretical problems of medicine.

The *Handbook of the Philosophy of Medicine* is offered as an all-embracing reference work that analyzes and discusses philosophical issues in relation to medicine and health care. It does not directly focus on ethical issues in health care, which have been thoroughly discussed in the last few decades, but centers around the basic concepts and methodological problems in medicine, which often underlie the ethical debates in health care.

This is the first wide-ranging, multiauthored handbook in the field. It introduces and develops dozens of topics, concepts, and issues and is written by distinguished specialists from multiple disciplines. The *Handbook of the Philosophy of Medicine* aims to be the most thorough book of its kind, covering all major topics that have been discussed in this vibrant area. It provides a single source of information for this far-ranging and still developing field. The chapters also advance these debates and aim at setting the agenda for years to come. The handbook will provide essential reading for anyone who wishes to develop an in-depth understanding of the philosophy of medicine or any of its subfields. It will be an invaluable source for laypeople, academics with an interest in medicine, and health care specialists who want to be informed and up to date with the relevant discussions.

A book project of this scale is very much a team effort. We are immensely grateful for the support of so many friends and colleagues. Most importantly, our authors have been fantastic to work with. Their enthusiasm for the project and their desire to advance the discipline, as well as their level of scholarship in the relevant areas, have made our task very easy. The members of the Advisory Board, Ruth Chadwick, Wim Dekkers, Martyn Evans, Elselijn Kingma, Lennart Nordenfelt, and Pekka Louhiala,

were extremely supportive and helped us enormously in identifying relevant topics and suitable authors. Finally, editorial staff at Springer, Alexa Singh, Navjeet Kaur, and Abhijit Baroi, were a pleasure to work with. They diligently and speedily produced the submitted chapters. In addition, Mike Hermann at the New York office supervised the project from beginning to end and provided invaluable advice.

Department of Philosophy, University of Liverpool, Liverpool, UK — Thomas Schramme

Philosophy, History and Law, Swansea University, Swansea, UK — Steven Edwards

Contents

Volume 1

Volume 2

About the Editors

Thomas Schramme is Chair in Philosophy at the University of Liverpool. He was Professor of Philosophy at the University of Hamburg from 2009 to 2016. He had affiliations with Swansea University (2005–2009) and the University of Mannheim (1998–2004). He earned his Ph.D. at Free University Berlin with a thesis on the concept of mental illness. His main research interests are in philosophy of medicine, political philosophy, and moral philosophy. He has published in journals such as *Bioethics*, *Ethical Theory and Moral Practice*, *International Journal of Law and Psychiatry*, and *Theoretical Medicine and Bioethics*. He edited several special issues and sections: "Expanding the normative framework of public health ethics: Some results from an interdisciplinary research group," *Public Health Ethics* 1/2015 (with Stefan Huster); "Christopher Boorse and the Philosophy of Medicine," *Journal of Medicine and Philosophy* 39 (6), 2014; "New Trends in Philosophy of Psychiatry," *Theoretical Medicine and Bioethics* 31 (1), 2010; "Lennart Nordenfelt's theory of health," *Medicine, Health Care and Philosophy: A European Journal* 10 (1), 2007. In addition, Schramme published several edited collections: *New Perspectives on Paternalism and Health Care* (Springer 2015), *Being Amoral: Psychopathy and Moral Incapacity* (MIT Press), and *Philosophy and Psychiatry* (de Gruyter Verlag 2004; with Johannes Thome).

Prof. Steven D. Edwards After working in the fields of intellectual disability and psychiatric nursing in the 1970s and 1980s, Prof. Edwards left nursing to study philosophy at the University of Manchester, UK. From there he obtained degrees in philosophy at the levels of undergraduate, masters (M.Phil.) (Relativism), and Ph.D. (Philosophy of Mind). After studying and teaching in the Department of Philosophy at the University of Manchester, he worked as a Senior Lecturer in Philosophy at Buckinghamshire Chilterns University,

before moving to join the Centre for Philosophy and Healthcare at University of Wales, Swansea, in 1995. He obtained a University of Wales Personal Chair in Philosophy of Healthcare in 2004. His academic interests are varied and he has published in a wide range of areas of philosophy, but over the past 20 years he has specialized in ethics and philosophy in the context of healthcare. He was founding coeditor of the journal *Nursing Philosophy* and has published books in philosophy of nursing, nursing ethics, philosophy of disability, philosophy of mind, and relativism. In addition to these books, he has published approximately 70 academic papers in scholarly journals. With respect to course leadership, for several years he was Director of Swansea University's MA Ethics of Health Care, and also Program Director of the BSc Medical Science and Humanities. He has been actively involved in the establishment of Clinical Ethics Committees in South Wales and is currently Chair of ABMU Health Board Clinical Ethics Committee. His current research is focussed on linguistic rights in healthcare, as well as ethical issues regarding the treatment of extremely premature neonates.

Contributors

Peter Allmark Centre for Health and Social Care Research, Sheffield Hallam University, Sheffield, UK

Mahesh Ananth Department of Philosophy, Indiana University South Bend, South Bend, IN, USA

Nicanor Pier Giorgio Austriaco O.P. Department of Biology, Providence College, Providence, RI, USA

Kristine Bærøe University of Bergen, Bergen, Norway

Jerome Bickenbach Human Functioning Sciences, Swiss Paraplegic Research, Nottwil, Switzerland

Andrew Bloodworth Department of Inter-Professional Studies, College of Human and Health Sciences, Swansea University, Swansea, UK

Kenneth Boyd Biomedical Teaching Organisation, Edinburgh University Medical School, Edinburgh, Scotland, UK

Hillel D. Braude The Mifne Center, Rosh Pinna, Israel

Matthias Braun Department of Theology, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

Alena Buyx Institute of Experimental Medicine, Christian Albrechts University of Kiel, Kiel, Germany

Silvia Camporesi Department of Global Health and Social Medicine, King's College London, London, UK

Arthur Caplan NYU Langone Medical Center and School of Medicine, New York, NY, USA

Havi Carel Department of Philosophy, University of Bristol, Bristol, UK

Ruth Chadwick University of Manchester, Manchester, UK

Angus Clarke Institute of Medical Genetics Building, Cardiff University School of Medicine, Cardiff, UK

Clare Clement College of Medicine, Institute of Life Sciences 2, Swansea University, Swansea, Wales, UK

Peter Dabrock Department of Theology, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

Catherine Dekeuwer Jean Moulin University Lyon III, Lyon, France

Wim Dekkers IQ Healthcare, Radboud University Medical Center, Nijmegen, The Netherlands

Marcus Doel College of Science, Department of Geography, Wallace Building, Swansea University, Swansea, Wales, UK

Andrew Edgar Cardiff University, Cardiff, UK

Steven Edwards Philosophy, History and Law, Swansea University, Swansea, Wales, UK

Ashley Frawley Department of Public Health, Policy and Social Sciences, Swansea University, Swansea, UK

Eva Friedel Klinik für Psychiatrie und Psychotherapie Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Berlin, Germany

KWM Fulford Faculty of Philosophy, University of Oxford, Oxford, UK

S. Nassir Ghaemi Department of Psychiatry, Tufts University School of Medicine, Boston, MA, USA

Mood Disorders Program, Tufts Medical Center, Boston, USA

Cambridge Health Alliance, Harvard Medical School, Boston, USA

Jan-Christoph Heilinger University of Munich, Munich Center for Ethics, Munich, Germany

Andreas Heinz Klinik für Psychiatrie und Psychotherapie Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Berlin, Germany

Jeanette Hewitt Centre for Philosophy, History, and Law, Swansea University, Swansea, UK

Martin Hoffmann Philosophisches Seminar, Universität Hamburg, Hamburg, Germany

Bjørn Hofmann The Norwegian University of Science and Technology, Gjøvik, Norway

Centre for Medical Ethics, University of Oslo, Oslo, Norway

Stephen Holland Departments of Philosophy and Health Sciences, University of York, Heslington, York, UK

Søren Holm School of Law, The University of Manchester, Manchester, UK

Center for Medical Ethics, HELSAM, University of Oslo, Oslo, Norway

Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Jeremy Howick Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

Peter Hucklenbroich Institut für Ethik, Geschichte und Theorie der Medizin, Westfälische Wilhelms-Universität Münster, Münster, Germany

David Hughes Department of Public Health and Policy Studies, Swansea University, Swansea, UK

Katrina Hutchison Department of Philosophy, Macquarie University, Sydney, NSW, Australia

Ashrafunnesa Khanom College of Medicine, Institute of Life Sciences 2, Swansea University, Swansea, Wales, UK

Elselijn Kingma University of Southampton, Southampton, UK

Hans-Peter Krüger Institut für Politische Philosophie und Philosophische Anthropologie, Universität Potsdam, Potsdam, Germany

Martin Langanke Faculty of Theology, Systematic Theology, Ernst Moritz Arndt University of Greifswald, Greifswald, Germany

Institute for Ethics and History of Medicine, University Medicine Greifswald, Greifswald, Germany

Vic Larcher Honorary Consultant in Paediatrics and Ethics, Great Ormond Street Hospital, London, UK

Keekok Lee Faculty of Humanities, University of Manchester, Manchester, UK

Dorothée Legrand Archives Husserl, CNRS, Ecole normale supérieure, Paris Sciences et Lettres Research University, Paris, France

Christian Lenk Ulm University, Institute for History, Theory and Ethics of Medicine, Ulm, Germany

Henrik Lerner Department of Health Care Sciences, Ersta Sköndal University College, Stockholm, Sweden

Wenke Liedtke Faculty of Theology, Systematic Theology, Ernst Moritz Arndt University of Greifswald, Greifswald, Germany

Pekka Louhiala Department of Public Health, University of Helsinki, Finland

Heidi Maibom University of Cincinnati, Cincinnati, OH, USA

James A. Marcum Department of Philosophy, Baylor University, Waco, TX, USA

Eric Matthews The School of Divinity, History and Philosophy, University of Aberdeen, Aberdeen, UK

John McMillan Bioethics Centre, University of Otago, Dunedin, New Zealand

Mike McNamee College of Engineering, Swansea University, Swansea, UK

Alexander Mebius Royal Institute of Technology (KTH), Stockholm, Sweden
Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

Oliver Müller University of Freiburg, Department of Philosophy and BrainLinks-BrainTools, Freiburg, Germany

Lennart Nordenfelt Ersta Sköndal University College, Stockholm, Sweden

John Paley Sheffield Hallam University, Sheffield, UK

Raimo Puustinen Medical School, University of Tampere, Tampere, Finland

Frances Rapport Centre for Healthcare Resilience and Implementation Science, Macquarie University, Sydney, NSW, Australia

Jens Ried Department of Theology, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

Wendy Rogers Department of Philosophy and Department of Clinical Medicine, Macquarie University, Sydney, NSW, Australia

Federica Russo Department of Philosophy, Faculty of Humanities, University of Amsterdam, Amsterdam, The Netherlands

Matthew Sample University of Washington, Department of Philosophy, Seattle, WA, USA

John Saunders College of Human and Health Sciences, Swansea University, Swansea, UK

Jan Schildmann Faculty of Medicine, Institute for Medical Ethics and History of Medicine, Ruhr-Universität Bochum, Bochum, Germany

Jann E. Schlimme Department of Psychiatry and Psychotherapy, Charité University Medicine, Berlin, Germany

Thomas Schramme Department of Philosophy, University of Liverpool, Liverpool, UK

Derek Sellman Faculty of Nursing, Edmonton Clinic Health Academy, University of Alberta, Edmonton, AB, Canada

James Stacey Taylor Philosophy, Religion, and Classical Studies, The College of New Jersey, Ewing, NJ, USA

William E. Stempsey Department of Philosophy, College of the Holy Cross, Worcester, MA, USA

Melanie Storey College of Medicine, Institute of Life Sciences 2, Swansea University, Swansea, Wales, UK

Dorota Szawarska Warsaw, Poland

Tim Thornton College of Health and Wellbeing, University of Central Lancashire, Preston, Lancashire, UK

Hugh Upton College of Human and Health Sciences, Swansea University, Swansea, Wales, UK

Sridhar Venkatapuram Global Health and Social Medicine, King's College London, London, UK

Jochen Vollmann Faculty of Medicine, Institute for Medical Ethics and History of Medicine, Ruhr-Universität Bochum, Bochum, Germany

Carolin Wackerhagen Klinik für Psychiatrie und Psychotherapie Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Berlin, Germany

Jerome C. Wakefield Silver School of Social Work and Department of Psychiatry, School of Medicine, New York University, New York, NY, USA

Simon Walker Bioethics Centre/Te Pokapū Matatika Koiora, University of Otago, Dunedin, New Zealand

Simon Woods PEALS (Policy, Ethics and Life Sciences) Research Centre, Newcastle University, Newcastle upon Tyne, UK

Sarah Wright College of Medicine, Institute of Life Sciences 2, Swansea University, Swansea, Wales, UK

Peter Zachar Department of Psychology, Auburn University Montgomery, Montgomery, AL, USA

Part I

Core Concepts in Health Care

Philosophy of Medicine and Bioethics

1

Thomas Schramme

Contents

Abstract

The chapter discusses ways to understand the notion of philosophy of medicine, with a special focus on the relation between philosophy of medicine and bioethics. Philosophy *of* medicine has been distinguished from other associations between philosophy and medicine. These conceptual distinctions lead to an account that delineates bioethics from the realm of philosophy of medicine. It has often been argued that medicine itself is a normative practice in that it aims at the good of patients. This undermines a simple account of medicine as a purely empirical, natural science. Yet such a normative account of medicine does not show that philosophy of medicine needs to aim at normative guidance like bioethics.

T. Schramme (✉)
Department of Philosophy, University of Liverpool, Liverpool, UK
e-mail: t.schramme@liverpool.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_58

Introduction

Philosophy of medicine is now an established field of study. Whether it could also be called a discipline or at least a subdiscipline in its own right is a contested matter, depending on a particular understanding of what constitutes an academic discipline. In any case, philosophy of medicine has not always been acknowledged as a distinct field of study, and important debates about its status and, indeed, whether it exists at all took place in the fairly recent past. It is important to get some grasp of these debates and the issues surrounding it to appreciate the tasks and the scope of philosophy of medicine. So the main aspect of this chapter is to clarify the notion of philosophy of medicine.

Another central problem is the relation of philosophy of medicine and bioethics. Although ethics is a part of philosophy, it is sometimes equated with moral philosophy – and although bioethics is obviously closely related to medicine, bioethics is arguably not a part of philosophy of medicine. Philosophy of medicine is distinctive in focusing on conceptual, methodological, axiological, epistemological, metaphysical, and other philosophical issues regarding medicine from a theoretical point of view, i.e., in order to analyze, understand, or explain aspects of the theory and practice of medicine. Bioethics, in contrast, discusses normative problems in medicine from a practical point of view, i.e., in order to provide guidance as to how people should act. Philosophy of medicine and bioethics are here delineated by distinguishing between a theoretical and a practical perspective or stance, not by their scope. Both might focus on theoretical and practical issues in medicine, for instance, they might address medical research aimed at gaining knowledge about the functioning of organisms (an issue regarding the theory of medicine), or they might be concerned with the clinical encounter between patient and doctor (an issue regarding the practice of medicine), but they do this with different aims. Put briefly and somewhat crudely, philosophy of medicine aims at analysis, whereas bioethics aims at guidance.

Bioethics could also be described as a field of inquiry that does not aim at the specific and peculiar characteristics of medicine. It consists rather in an application of the instruments provided by ethics to a specific area, namely, to biomedicine. To be sure, there can be genuine normative issues in philosophy of medicine, but here the impetus is to understand the specific evaluative and normative aspects of medicine, for instance, the notion of suffering. Also, for a topic to belong to philosophy of medicine, the direct object of understanding must be an aspect of medicine, not a general issue that is pursued by merely using examples from medicine, for instance, when discussing the philosophy of scientific experiments or of causality.

To say that philosophy of medicine and bioethics are different fields of study obviously does not mean that there is no connection between them. For instance, in order to discuss the ethics of organ transplantation, one needs a clear understanding of the concept of death. To analyze this concept is a task for philosophy of medicine. To distinguish between bioethics and philosophy of medicine also does not imply that there are no issues regarding value or morality in philosophy of medicine.

Yet, these issues are discussed within philosophy of medicine from a theoretical point of view, i.e., not in order to solve practical problems but in order to understand and explain evaluative and normative aspects of medicine. To do bioethics properly, one needs some acquaintance with philosophy of medicine, and to do philosophy of medicine properly, one needs some knowledge of ethical problems in medicine. After all, theories to explain aspects of medicine should serve specific purposes, and these are usually determined by the need to solve normative problems. The latter connection also speaks against reducing philosophy of medicine to a mere subfield of philosophy of science. Although philosophy of medicine obviously is part of the philosophy of science, it is also not restricted to explaining medicine as a science (Pellegrino 1998, 326). Medicine is both a science and an art; it has theoretical as well as practical aspects. It is different from many other sciences in its interpersonal aspects, the encounter between patient and clinical personnel. Hence, there are accordingly philosophical aspects of medicine that are not usually found in other areas of philosophy of science. To be sure, whether philosophy of medicine can be fully subordinated to the philosophy of science obviously depends on the interpretation of the scope and perspectives of philosophy of science itself, a topic that falls out of the remit of this chapter.

It should be noted that the distinction between bioethics and philosophy of medicine, as it has just been introduced, is itself contested. There are certainly philosophers who would see bioethics, or medical ethics for that matter, as one element of philosophy of medicine. Yet the rationale for such subordination seems to be based on a myopic, if not entirely incorrect, vision on ethics as a subfield of philosophy. It is true, of course, that bioethics can be aligned with the philosophical subdiscipline called "ethics." Hence, bioethics is a field of study, in this perspective, that is itself part of philosophy as applied to medicine. Yet, this does not preclude other important differences between philosophy of medicine and bioethics, as has been stressed before, especially regarding their aims and purposes. It is also important to be clear about the relation of philosophy and medicine when speaking of the philosophy *of* medicine. As will be seen shortly, not every application of philosophical methods to medicine is an instance of philosophy of medicine.

Whether philosophy of medicine exists then depends on our understanding of it. The question can further be separated into two aspects. First, we might ask whether philosophy of medicine exists as a separate genuine discipline or field of study. We have already touched upon this issue and replied affirmatively. It should be noted, though, that for philosophy of medicine to exist in this way, there need not be any person doing it. It means an existence as a theoretical entity, as an idea. Second, it could be asked whether philosophy of medicine exists as an institution. This would require certain real entities such as persons doing philosophy of medicine, learned societies, academic journals, a canon of key book, textbooks, and possibly degrees in the field, etc. (Caplan 1992). Although philosophy of medicine in this respect is still a nascent endeavor, an answer can be given in the affirmative.

The Definition of Philosophy of Medicine

Philosophy and medicine are both academic disciplines as well as practices. They can be put in many different relations to each other. For instance, philosophical methods or approaches, such as conceptual analysis, phenomenology, or hermeneutics, may be applied to medical notions or medical practices. This alone does not make it an instance of philosophy of medicine. The genuine interest in understanding must focus on medicine. So to do philosophy of medicine, the philosophical methods and approaches are mere instruments in gaining knowledge about aspects of medicine, rather than gaining knowledge about methods and approaches in philosophy by using medicine as an area of application. Surely, this distinction lays out an ideal type that will not always be visible in real publications. Yet, to use two books as examples, there is a difference between a scholar using medical examples in order to philosophically discuss pain (Hardcastle 2001) – obviously a phenomenon that is common in medical settings – and a scholar analyzing the notion of suffering in relation to the goals of medicine (Cassell 1991). The first relation could be called, following the distinctions drawn by Edmund Pellegrino (Pellegrino 1976, 1986), philosophy in medicine, whereas the latter would be genuine philosophy of medicine. As with other "philosophy of" relations, the purpose of study here is to gain knowledge about the nature of medicine, about the specific aspects of that particular theory and practice. A similar difference can be seen between a study discussing autism in order to explore a theoretical issue in the philosophy of mind, namely, how we gain access to another person's mind, (Gordon and Barker 1994) and a philosophical investigation into what it means to live with chronic illness (Toombs 1992). Admittedly, these delineations are somewhat stipulative and probably contested. They are also, as has been said before, hard to draw in reality. Yet they should also help us in thinking about the proper domain of philosophy of medicine.

In antiquity, the relationship between philosophy and medicine was strong (Frede 1986). This of course applied to other disciplines as well, which have today become separated from their philosophical origins. Many philosophers, such as Pythagoras, Empedocles, and Democritus, were medical experts and some of them had advanced theories regarding the nature of disease. This was usually described as an imbalance of important elements of bodies, called humors. The attention of philosophers at the time revolved around a general theory of nature, especially human nature; hence, medical phenomena were of particular interest. These anthropological and ontological issues were complemented by a practical interest in giving advice for a good life, of which mental and somatic health was deemed part and parcel. Doctors developed a theoretical interest in medicine, roughly from the fifth century BC, because disease more and more came to be acknowledged as a condition that could be altered and possibly healed. Hence, the task of medical experts was to gain access to the necessary related knowledge. To be sure, the search for a systematic approach to study medicine was not so much caused by a theoretical attraction but by social pressure, which threatened the role and status of physicians. Doctors accordingly started to explore issues that were

oriented toward the practice of medicine, for instance, the way theoretical knowledge could be applied to individual cases and what were the inherent limitations of the art of healing. From this emerged different accounts of the role of philosophical methods in medicine and, roughly from the fourth century BC, competing theories regarding methods of treatment. Another topic of interest was ethical problems, especially regarding the relation of medical experts to patients. One aim was to defend their social status by drawing a rigid demarcation to quacks.

In many other countries and cultures, there were occasional strong links between philosophy and medicine, and also in later modernity, there were quite a few academic outputs that are in the field of philosophy of medicine (Temkin 1956; King 1977). Yet, the current more concerted and institutionalized debate regarding the philosophy of medicine within the Western civilization began in May 1974 (Potter 1991). It was then that the first interdisciplinary symposium on philosophy and medicine took place at Galveston, Texas. From this annual meeting, the important book series in philosophy and medicine (now published by Springer) emerged, which was originally edited by Tristram Engelhardt and Stuart Spicker. In 1976 the first issue of the *Journal of Medicine and Philosophy* was published. Its founding editor was Edmund Pellegrino. In the same year, a meeting of the American Philosophy of Science Association focused on epistemological issues in medicine. Other important journals in the fields were released later, for instance, *Theoretical Medicine and Bioethics* (originally called *MetaMed*, founded in 1977 by Kazem Sadegh-Zadeh) and *Medicine, Health Care and Philosophy* (founded 1998, official journal of the European Society for Philosophy of Medicine and Health Care). There are also numerous journals that publish in the field of philosophy of medicine, as well as specialized journals with an interdisciplinary bend, for instance, *Philosophy, Psychiatry and Psychology* (founded 1994, official journal of the Association for the Advancement of Philosophy and Psychiatry). Finally, there are a couple of learned societies and an important email distribution list, run by Jeremy Rosenbaum Simon.

Not surprisingly, maybe, there was a lot of optimism in the 1970s regarding the potentials of the discipline called philosophy of medicine. This can be clearly seen in a quote from Edmund Pellegrino, who wrote in 1976: "We are entering a new era of dialogue – perhaps one as promising as that between Greek medicine and philosophy" (Pellegrino 1976, 12 f.). Yet, this positive prognosis was later put into doubt, for instance, by Arthur Caplan in 1992, who claimed that philosophy of medicine does not exist, and by Heinrich Loewy, when he announced a new journal section on "Philosophy and its Role in Medicine": "Surprisingly little has been written in English about the philosophy of medicine: Physicians often see medicine as a purely technical occupation and can make little of the term 'philosophy of medicine'. Philosophers, likewise often feel that medicine is a merely technical discipline and that its philosophy is somehow not worthy of serious attention" (Loewy 1994, 201 f.).

Altogether, both the very optimistic and the bleak outlook seem wrong from today's point of view. There is now a genuine discipline philosophy of medicine. Still, there is scope for discussing what exactly it is or what it should be. The way

we conceive of philosophy of medicine obviously has repercussions on how it is faring in terms of its disciplinary viability. In the following, two influential accounts of the philosophy of medicine will be scrutinized more closely.

Edmund Pellegrino

In his important contribution "Philosophy of Medicine: Problematic and Potential," Edmund Pellegrino thoroughly discusses the possible relationships between medicine and philosophy. He categorizes these relations into three different types: Philosophy *and* medicine, philosophy *in* medicine, and philosophy *of* medicine.

> Philosophy *and* medicine comprises the mutual considerations by medicine and philosophy of problems common to both (. . .). Some of the recurrent problems of philosophy – the mind-body debate; the meanings of perception, consciousness, language; the special or nonspecial character of chemical and physical laws in living things – are susceptible to this type of collaborative attack. (. . .) Philosophy *in* medicine refers to the application of the traditional tools of philosophy – critical reflection, dialectical reasoning, uncovering of value and purpose, or asking first-order questions – to some medically defined problem. The problems can range from the logic of medical thought to the epistemology of medical science as science, the problem of causality, the limitations of observation and experiment, and of course, the whole range of vexing issues in the active field of biomedical ethics. (. . .) When philosophy turns to the meaning of medicine as clinical practice and examines its conceptual foundations, its ideologies, its ethos, and the philosophical bases for medical ethics, then it becomes the philosophy *of* medicine. The questions examined by philosophy *in* medicine are then carried to the unique realm of the clinical encounter with a human being experiencing health, illness, neurosis, or psychosis, in a setting which involves intervention into his existence. The philosophy of medicine seeks explanations for what medicine *is* and *ought* to be (. . .) These three types of engagement are rarely separable in actual fact, and philosophers can, and do, engage in all three. We have dissected them free to underscore the central importance of the philosophy of medicine: the philosophical issues imbedded in the theory of medicine as a practical human activity. Ultimately, the more proximate issues dealt with by philosophy *and* medicine, and philosophy *in* medicine, must rest on the philosophy *of* medicine. (Pellegrino 1976, 19 ff.)

Pellegrino explicitly excludes biomedical ethics from philosophy of medicine. He also describes as an essential part of the philosophy of medicine a practical component, the clinical encounter. This practical element is especially important, according to Pellegrino, because it is its unique feature, in contrast to, say, biology. Medicine here has a set aim, in contrast to natural sciences, namely, health or healing of living beings. The personal relationship between doctor and patient in pursuing this aim turns medicine into a value-laden, a moral, activity. Hence, medicine cannot be reduced to other sciences, for instance, to a mix of biology and psychology (cf. Shaffer 1975). Pellegrino claims: "Medicine, then, is an activity whose essence appears to lie in the clinical event which demands that scientific and other knowledge be particularized in the lived reality of a particular human, for the purpose of attaining health or curing illness, through the direct manipulation of the body, and in a value-laden matrix. It is in this sense that medical

theory is a theory of practical reality and not just the theory of the sciences which contribute to it" (Pellegrino 1976, 17).

There are several issues that can be queried in this account of the philosophy of medicine. For instance, it can be queried whether cure of illness and promotion of health are really the essential or only goals of medicine. Relatedly, there is also a worry that Pellegrino has resolved an issue by stipulation that should be first clarified by a debate within philosophy of medicine: to determine the "nature" of medicine. It seems wrong to restrict philosophy of medicine to the practical realm of medicine by claiming the clinical encounter as the essence of medicine. To clarify what role the practice of medicine has in relation to the theory of medicine is a genuine task of the philosophy of medicine itself and should not be excluded by restricting the scope of the nature of medicine to the clinical setting.

Ten years later, Pellegrino revisited his three-partite distinction of the relationships between medicine and philosophy. Here, he advances a definition of the philosophy of medicine that does not rely on a particular interpretation of the essence of medicine from the outset, but sees a determination of such an interpretation as the outcome of doing philosophy of medicine. "The third mode of relationship, philosophy *of* medicine, concentrates on a philosophical inquiry into medicine-qua-medicine. It seeks to define the nature of medicine as medicine, to elaborate some general theory of medicine and medical activities." (Pellegrino 1986, 10)

Later in the chapter, he gives a more substantial account of the distinctive problems discussed in philosophy of medicine:

> Philosophy of medicine is more than philosophizing about the phenomena peculiar to medicine, i.e., philosophy *in* medicine. It seeks to understand and define the conceptual substrata of medical phenomena. Its agenda is a broad ranging one – it deals with such crucial notions as the ideas of health, illness, normality and abnormality, healing cure, care, suffering, and pain. What do these concepts embrace? What is the nature of medical diagnosis, clinical judgment and discovery? (. . .) Does the end of medicine modify the logic and the epistemology of clinical judgments? (. . .) What are the values that structure medicine? (. . .) Is health a value and in what sense? (. . .) Questions of this sort provide the agenda for the philosophy of medicine as a discipline. (Pellegrino 1986, 14 f.)

When reading this list of topics for the philosophy of medicine, it becomes less clear in what way it differs from philosophy *in* medicine. After all, philosophy in medicine has just been described as "philosophizing about the phenomena peculiar to medicine." The key to Pellegrino's understanding of the philosophy of medicine is that he believes in a distinctive nature of medicine, "medicine-qua-medicine," that determines its agenda. This distinctive nature of medicine, for Pellegrino, is its practical focus with the related *telos* of health.

> Philosophy of medicine makes the specific method and matter of medicine the subject of study by the method of philosophy. Philosophy of medicine seeks philosophical knowledge of medicine itself. It seeks to understand what medicine is and what sets it apart from other disciplines, and from philosophy, itself. (. . .) Medicine *qua* medicine comes into existence in the clinical encounter or in public health when the knowledge of the sciences basic to

> medicine is employed for a specific end, i.e., for the cure, containment, amelioration, or prevention of human illness in individuals or in human societies. (. . .) Philosophy of medicine seeks to understand the nature and phenomena of the clinical encounter, i.e., the interaction between persons needing help of a specific kind relative to health and other persons who offer to help and are designated by society to help. (Pellegrino 1998, 326 f.)

In summary, for Pellegrino, philosophy of medicine is to be distinguished from other relationships between philosophy and medicine. For him, its focus is medicine as a distinctive discipline. On the one hand, he says that determining such nature of medicine is itself a task of philosophy of medicine; on the other hand, he repeatedly claims that indeed the distinctive feature of medicine is its practical nature, more specifically in the clinical encounter.

Tristram Engelhardt and Edmund Erde

Tristram Engelhardt and Edmund Erde coauthored an entry on "Philosophy of Medicine" for the 1978 edition of the influential *Encyclopedia of Bioethics* and later published another substantial article on the subject (Engelhardt and Erde 1980). Similarly to Pellegrino, they also distinguish between different types of relationships between philosophy and medicine:

> Philosophical activity concerning medicine can be focused through four major themes: Philosophy *for* medicine, philosophy *in* medicine, philosophy *about* medicine, and philosophy *of* medicine. (. . .) The first uses concepts speculatively to generate medical explanations. (. . .) The second theme can be styled 'philosophy in medicine'. Here, philosophy is a formal analytical tool, not in the direct service of medical theory or therapy, but rather employed to display logical structures in medicine. (. . .) The third theme (. . .) involves reflection on traditional philosophical issues (not logical issues in the strict sense) arising from the domain of medicine. (. . .) The fourth theme, 'philosophy of medicine', can be used to identify those epistemological and conceptual issues peculiar to medicine in a way analogous to the philosophy of any science (. . .). (Engelhardt and Erde 1978, 1049 f.)

Surely, this classification of various relations between medicine and philosophy cannot simply be taken to be descriptive. Engelhardt and Erde hence regard their definition of the philosophy of medicine to be the proper one. Here, they endorse a fairly narrow understanding of philosophy of medicine that they later gave up, because they appreciated that the field of "medicine" cannot easily be delineated. Still, in the quoted definition, they exclude bioethics from the realm of philosophy of medicine, like Pellegrino. Bioethics belongs to the category of "philosophy about medicine," because its problems are traditional ones, not specific to medicine, if raised here in a new way. Other problems of philosophy about medicine have also been introduced in other domains, such as philosophy of mind and philosophy of science, and then been applied to medicine. In contrast to such transfer of philosophical issues to medicine, philosophy of medicine deals with problems specific to medicine, according to Engelhardt and Erde. Examples are the analysis of basic medical concepts, such as "disease," "pathology," or "health." Engelhardt, at

another occasion, calls this subject area of study philosophy of medicine in a strong sense (Engelhardt (1977), 98 ff.). According to this understanding, "philosophy about medicine" would be equal to "philosophy of medicine" in a weak sense.

The Need for Philosophy of Medicine

Although one should surely be careful not to take these distinctions as carved in stone or to see no connections when pursuing either area, such a classification of different relations between philosophy and medicine serves an analytic purpose of sorting out a diverse field of study. Another, if contested, benefit of such a demarcation is to keep medical ethics and bioethics separate from philosophy of medicine. This is helpful because it becomes obvious that conceptual, metaphysical, epistemological, and value theoretical aspects underlie moral problems in medicine. For instance, in medical ethics the notion of quality of life is often used without analyzing the strongly related concept of health. Also, when justifying the often painful treatment of children with spina bifida by referring to the goal of "healing," one implicitly assumes an ideal of health (Hare 1986, 174). But how can such treatment be justified without first having clarified the notion of health? Another example is the question whether enteral nutrition and hydration are or ought to be part of basic medical care that cannot be rejected. An answer to this question obviously depends on an understanding of the notion of care.

Keeping philosophy of medicine and bioethics distinct might therefore advance the sense that bioethics, a highly professionalized discipline, actually needs the discussion of foundational philosophical questions (Thomasma 1985, 239; Lindahl 1990). It might also boost the number of publications in the philosophy of medicine, as a dearth in this area can easily be diagnosed, in contrast to the volume of publications in bioethics.

In summary, there is a viable discipline philosophy of medicine in its own right, and there is a need for academic contributions in this area. Surely, there are still open questions regarding its contours, especially when we consider medicine's distinctive entanglement of theoretical and practical issues. Here, issues of normative significance are raised and hence the relation between ethics and philosophy of medicine becomes the subject of inquiry on another level. Is medicine and its foundations intermingled with values and norms in a way that turns it into a "moral science"? Such an interpretation has already been mentioned in the case of Pellegrino's account. It is an important problem of philosophy of medicine itself, whether medicine is diffused by values and norms, and if so, of what kind they are.

Values and Norms in Medicine

Medicine can be interpreted as a value-laden practice insofar as every science is charged with values and norms. Marx Wartofsky, for instance, identified two norms that are presumed in natural sciences (Wartofsky 1977, 111). First, the norm of mathematical rationality, second the norm of empirical testability to scrutinize the

truth of statements. These norms determine what is accepted as proper and "good" methodology in science. Such a methodology at the same time delineates the area of science. Only those issues that can be discussed by using the methodology of hypothesis and deduction and that can be checked against empirical facts are then supposed to be matters of science. A well-known corollary of such thinking is the distinction between facts and values, including the scientific ideal of staying value-free. In contrast, Wartofsky claims that science is, despite its plea for neutrality, normative in its pursuit and content.

> Since methodology is normative and prescriptive (indeed proscriptive) in the way just described, the very choice of methodological canons, and their elaboration over the past century, certainly marks the enterprise as normative; and normative in the sense that a particular historically evolved norm serves to exclude or replace others. (ibid. 112)

According to Wartofsky, there are numerous limitations and shortcomings of a value-free model that become obvious on different levels. First, the vicinity of science is too restricted. Many scientific issues in, for instance, thermodynamics or quantum physics cannot be neatly subsumed under this model. Similar considerations apply to human sciences and psychology. Second, possible ways of scientific judgment, argumentation, and deduction actually break open the borders of the prescribed methodology. Third, the model obscures the relation between scientific theory and practice. Fourth, the historical and social context of science is amputated. Fifth, by presuming the "objectivity" of science, there is an unhelpful distinction implied between pursuing truth and considering the use of such truth.

Many, if not all, of Wartofsky's claims are admittedly discussed by many other scholars within philosophy of science, and several of his theses are contested. The value of his contribution for philosophy of medicine lies mainly in the fact that he goes on to assign a characteristic of medicine to each of the listed problems in the traditional scientific model, to show how the philosophy of science might be put on a new basis by scrutinizing medicine from a philosophical point of view. First, health and disease – phenomena on which medicine is based – do not only allow for a description as biological conditions. Rather, they necessarily imply social contexts and subjective states, as disease never occurs isolated but in a system of relationships. Health and disease are therefore normative ideas that transcend the contours of the model of scientificity portrayed earlier. Here, it is important to stress that Wartofsky does not want to claim that only those conditions are instances of disease, which have been identified such by medicine or affected persons. He wants to point out that the identification of diseases is itself done within a social practice. This transcends the area of medicine over and above the realm of, say, biology and grounds it in a historically developed life form.

Second, clinical judgment, especially diagnosis, is far too complex to be put under the rubric of experimental and theoretical deduction. Third, the relation between theory and practice is special in medicine, as scientific research here is determined by the practical possibilities of knowledge. Although medical science and medical art of healing can be distinguished, the former is still oriented toward

the latter, as medicine aims at the advancement of human well-being. Fourth, as already stressed in the first point, basic medical notions are influenced by social and historical dimensions. It would destroy the complexity of such a relation if issues of scientific quality and non-scientificity would be separated. Fifth, the application of medical knowledge in the encounter between doctor and patient is a central element of medicine. Diagnosis and therapy in medical practice alter the standpoint of the examiners; they themselves become an element of the system to be explored. Hence, the idea of objectivity, which is common in the natural sciences, is here undermined. Moral issues become pertinent: "Indeed, moral questions, and the social facts of life and death, weal and woe, are not peripheral to medicine but central to it" (ibid. 120).

For Wartofsky it follows that there ought to be a radical revision of what is considered scientific in order to arrive at a richer notion of science. Now, it seems that – if we follow his considerations – we would have to say that moral issues are after all part and parcel of philosophy of medicine, in contrast to what has been established before. This would be due to the fact that medicine is first and foremost a morally determined practice. Yet, the fact that a certain practice is normatively charged, maybe in a moral way, does not make the philosophical analysis of this normative practice itself an endeavor in ethics. Whether and in what way medicine is infused by values is an important issue for philosophy of medicine. But such a question is not itself a normative question – it does not lead to statements about how to act or evaluations what is a good practice, for example. It rather concerns epistemological and metaphysical aspects of elements of medicine that can indeed be normatively charged.

> There need not be a strict divide between values and facts in understanding epistemological questions about medicine. On the contrary, I believe that fact and value blur in important and unavoidable ways in the realm of medicine. But the recognition that fact and value, morality and methodology are inextricably wed when the subject is medicine does not obviate the claim that the philosophy of medicine is and ought to address different questions than those pursued by those doing bioethics. Bioethics tries to answer questions that are normative. The philosophy of medicine concerns itself with questions that are primarily either epistemological or metaphysical. (Caplan 1992, 69)

Conclusion

There is some disagreement within philosophy of medicine about its proper definition. This also affects the delineation between bioethics and philosophy of medicine. The point of view developed here states that philosophy of medicine deals with meta-medical problems from a theoretical perspective that aims at explanation or analysis. Bioethics, in contrast, aims at guidance or recommendation. Philosophy of medicine analyzes concepts such as "health," "disease," or "care," and it tries to identify the values and norms underlying medicine. In addition, it deals with epistemological questions, for instance, regarding the

status of clinical judgment and the methods of gaining medical knowledge. Genuine ethical topics play an important role in the public debate regarding medicine. It therefore seems especially important for philosophers to discuss the foundations of such discussions.

Definition of Key Terms

Philosophy of medicine	A field of study that aims at analyses of metaphysical, epistemological, methodological, conceptual, and other philosophical issues regarding medicine.
Philosophy and medicine	A perspective on problems common to both philosophy and medicine.
Philosophy in medicine	The application of philosophical methods or theories to the realm of medicine.
Medicine	Medicine consists of a theoretical (science) as well as a practical aspect (art). It predominantly aims at restoring or improving the health of patients. It is a contested matter – itself a topic of philosophy of medicine – whether medicine has a specific nature or essence.
Discipline	An established field of study fulfilling certain formal requirements, such as the existence of learned societies, textbooks, and journals.
Bioethics	An area of applied ethics that focuses on normative issues in biomedicine and aims at guiding decisions how to act.

Summary Points

- Medicine is both an art and a science; medicine is diffused with values.
- This makes medicine a normative discipline though it does not necessarily include a certain aim or telos.
- Philosophy of medicine can be delineated from other relations between these disciplines.
- Philosophy of medicine is a field of study that aims at analyses of metaphysical, epistemological, methodological, conceptual, and other philosophical issues regarding medicine.
- Philosophy of medicine has become an established discipline in its own right.
- Philosophy of medicine is different from bioethics or medical ethics in that the former aims at analysis, the latter at guidance.

References

Caplan AL (1992) Does the philosophy of medicine exist? Theor Med 13:67–77
Cassell EJ (1991) The nature of suffering and the goals of medicine. Oxford University Press, Oxford
Engelhardt HT Jr (1977) Is there a philosophy of medicine? In: Asquith P, Suppe F (eds) PSA (Philosophy of Science Association) 1976, vol 2. Philosophy of Science Association, East Lansing, pp 94–108
Engelhardt HT Jr, Erde EL (1978) Philosophy of medicine. In: Reich W (ed) Encyclopedia of bioethics, vol 3–4. Macmillan/Free Press, New York, pp 1049–1054
Engelhardt HT Jr, Erde EL (1980) Philosophy of medicine. In: Durbin PT (ed) A guide to the culture of science, technology, and medicine. Free Press, New York, pp 364–461
Frede M (1986) Philosophy and medicine in antiquity. In: Donagan A, Perovich AN Jr, Wedin MV (eds) Human nature and natural knowledge. Kluwer, Dordrecht, pp 211–232
Gordon RM, Barker JA (1994) Autism and the 'theory of mind' debate. In: Graham G, Stephens GL (eds) Philosophical psychopathology. MIT Press, Cambridge, MA, pp 163–181
Hardcastle VG (2001) The myth of pain. MIT Press, Cambridge, MA
Hare RM (1986) Health. J Med Ethics 12:174–181
King L (1977) The philosophy of medicine: the early eighteenth century. Harvard University Press, Cambridge, MA
Lindahl BI (1990) Editorial. Theor Med 11:1–3
Loewy EH (1994) Philosophy and its role in medicine: inaugurating a new section. Theor Med 15 (2):201–205
Pellegrino ED (1976) Philosophy of medicine: problematic and potential. J Med Philos 1(1):5–31
Pellegrino ED (1986) Philosophy of medicine: towards a definition. J Med Philos 11(1):9–16
Pellegrino ED (1998) What the philosophy *of* medicine *is*. Theor Med Bioeth 19:315–336
Potter RL (1991) Current trends in the philosophy of medicine. Zygon 26(2):259–276
Shaffer J (1975) Roundtable discussion. In: Engelhardt HT Jr, Spicker SF (eds) Evaluation and explanation in the biomedical sciences, vol 1, Philosophy and medicine. Reidel, Dordrecht, pp 215–218
Temkin O (1956) On the interrelationship of the history and the philosophy of medicine. Bull Hist Med 30:241–251
Thomasma DC (1985) Editorial: philosophy of medicine in the U.S.A. Theor Med 6:239–242
Toombs SK (1992) The meaning of illness. A phenomenological account of the different perspectives of physician and patient. Kluwer, Dordrecht
Wartofsky MW (1977) How to begin again: medical therapies for the philosophy of science. In: Asquith P, Suppe F (eds) PSA (Philosophy of Science Association) 1976, vol 2. Philosophy of Science Association, East Lansing, pp 109–122

Normality as Convention and as Scientific Fact

2

Ruth Chadwick

"The normal itself is an abnormality."
G.K. Chesterton

Contents

Abstract

The concept of the normal is central in modern societies in general and in medicine in particular. Norms are established for body measurements such as cholesterol and body temperature. There are several interpretations of "normal" however. The statistical concept of "normal" is a relatively recent phenomenon historically and some argue that it is a mechanism of power and control. On the

R. Chadwick (✉)
University of Manchester, Manchester, UK
e-mail: ruth.chadwick@manchester.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_9

other hand, a concept of the normal is arguably necessary to science, medicine, and the possibility of diagnosis.

Introduction

The word "normal" derives from the Latin "norma," the square which stonemasons used to standardize their patterns, thus testing the accuracy of a 90° angle in the stone. In contemporary debates, there are multiple interpretations of the concept. As John Dupré has noted, there is an unsophisticated usage according to which "normal" is what is familiar (Dupré 1998). C. Daly King remarked, however, that "in the whole field of psychology and to a lesser but increasing extent in the biological fields, we find a prevalent misuse and misapplication of the basic term, normal" (King 1945, p. 493). Broadly speaking, there might appear to be, first, a twofold division between those who argue that what is normal is a matter of scientific discovery and fact, on the one hand, and those who argue that it is an evaluative concept, on the other. It is also part of contemporary deployment of the term that it is used to bridge the fact-value gap. The mason's square also came to be used in that way – the square became a symbol of honesty, leading to phrases such as "on the square" and "square deal."

On closer inspection it appears that the situation is much more complicated than one binary distinction: there are other distinctions and divisions to be drawn. On the side of those who argue that it is a scientific concept, there is disagreement between those who interpret the normal in a statistical way and those who argue that there is another, biological, interpretation which relates to function. Some accounts include elements of both interpretations. Within the evaluative camp, there are those who point out first that the discipline of statistics (and thus the statistical interpretation of normality) is not value-free and those who go further than simply saying that the normal is evaluative and argue that the "normal" is used as a mechanism of power and control. Ian Hacking, for example, claims that normality "uses a power as old as Aristotle to bridge the fact/value distinction, whispering in your ear that what is normal is also right" (Hacking 1990, p. 160). There is also a distinction to be made between the use of the normal as used to refer to the "normal" state of an *individual*, rather than in relation to a reference class. In the context of medicine, in particular, this usage may be particularly relevant in relation to the notion of the "recovery" of an individual (see below).

Norms as Scientific Facts

Statistical Norm

The statistical concept of normality, although it is well established in modern discourse, has had a relatively short history. According to Hacking (1990) and to Lennard Davis, it appeared around 1830, alongside the development of the discipline

of statistics itself. A major reason for the emergence of statistics at that time was the perceived need for well-informed state policy. Davis points out the ways in which statistics quickly became taken up in relation to human bodies as well as nation states, through medical statistics and new understandings of disability (Davis 1995).

The statistical data distribution pattern known as the bell curve, the so-called normal distribution, occurs in many natural phenomena. It is also called the Gaussian distribution after Carl Friedrich Gauss (1777–1855), who used it to model errors in astronomical observations. The peak of the curve occurs at the mean of the data, with 50 % of the distribution to the left and 50 % to the right. In a normal distribution, 68 % of all observations fall within a range of $\pm$ 1 standard deviation from the mean.

In medicine normal distributions are used in relation to measurements such as blood pressure and cholesterol levels. They also include data about life expectancy and age of onset of puberty and of menopause. Although it is known that these timespans and ages are subject to change along with environmental conditions and lifestyle, a normal distribution pattern recurs, albeit with a different mean. But how did the shift occur from observing this pattern to using it as a mechanism of evaluation of human beings?

Davis argues that the emergence of statistics proved attractive to Galton and the emerging eugenics movement. He writes: "The rather amazing fact is that almost all the early statisticians had one thing in common: they were eugenicists" (Davis 1995, p. 30). The possible identification of some members of the population as disabled and "abnormal" offered an opportunity to argue for the importance of trying to maintain the quality of the gene pool. When the eugenic enterprise became associated with statistics, however, it had to face the problem that however much a eugenic plan is put in place, the law of the bell curve dictates that there will always be people at the extremes. Furthermore, the concept of abnormality, though in principle applicable to those at each extreme, came to be applied negatively to those at the extreme that was not admired. This is an example of the bridging of facts and values through the use of the concept.

Davis argues that prior to the emergence of statistics, people tended to compare themselves to an ideal, rather than to the norm. The implication of comparing oneself to an ideal, of course, is that everyone falls short, although all can aspire. The development of the concept of normality, according to Davis, marked a paradigm shift in how human life is regarded. Both of these points are strongly reminiscent of themes in Plato's *Republic*. The notion of an ideal was taken to an extreme by Plato, through the Theory of Forms. Indeed, for Plato everything in the world that we perceive was only a representation of the ideal forms. He also drew an analogy between the individual and the state, arguing that both had a tripartite structure. Interestingly, prior to the emergence of the concept of the norm, Plato's *Republic* also contained ideas describable as eugenic, in their attempt to influence the quality of children born, by the mechanism of controlled breeding programs. At the time of writing, there is considerable discussion of the possibilities of human enhancement. Arguably, the move toward discussing enhancement in the context of medicine suggests a return toward aspiring to an ideal (c.f., Wiesing 2009). The much disputed distinction between therapy and enhancement is perhaps a reflection of that.

Normality and Biological Function

When King, as noted above, suggested that the word normal is misused, he said that the statistical average may be, and very often is, abnormal. The use of normal in the sense of the statistical sense was incorrect: "normal . . .is objectively, and properly, to be defined as *that which functions in accordance with its design*" (King 1945, p. 494). He also claims that the term, normal, was originally invented in this context. In particular, he argues that medicine could not operate with a statistical interpretation of the normal but needs to think of health and disease in terms of function and malfunction.

What we have at the peak of the bell curve, according to King, is what is typical, but the average is not "normal": the former judgment is quantitative, the latter qualitative. King despairs, however, of the possibility of dissociating, now, the use of the word normal from statistics and proposes another term, "paradic," from paradigm, for the functional meaning (King 1945, p. 500).

Robert Wachbroit has also argued for a way of understanding normality that is related to biological function. He says:

> Consider one of the favorite examples of the philosopher of biology, "the function of the heart is to circulate the blood". That statement is clearly not about any *particular* heart....Nor is it a statement about *all* hearts....What the statement is about is the *normal* heart, with the understanding that a particular heart . . .may not be normal. (Wachbroit 1994, p. 580)

According to Wachbroit, there is no concept in the physical sciences corresponding to this sense of normality in the biological sciences – it makes no sense, for example, to talk of "the normal electron." He also argues that accounts of biological functions cannot *explain* the concept of biological normality because they *presuppose* it. Only if this is the case can we understand the distinction between functions and malfunctions. Statistical data may provide e*vidence* for normality in this biological sense, but it is a different concept of normality from that as expressed in the bell curve.

Arguments Against Normality as Biological Function

There are, however, several problems in understanding normality in terms of biological function. There is, first, the fact of considerable variation in animal and human populations. Indeed, in the biological sciences, particularly in the aftermath of the Human Genome Project, the scientific concept of interest is *variation* rather than *normality*: research is ongoing into the variations in the genome that underlie differences in areas such as susceptibility to disease and responses to pharmaceutical products. What is more important for the success, or otherwise, of a functional explanation of normality, however, is the extent of the differences between individual members of a species that are apparently compatible with survival. Examples discussed include Slijper's goat, which was born without front legs and with multiple

other deformities but which learned to walk on its hind legs (see Amundson 2000; Cooper 2012).

Discussing these and other examples, Ron Amundson argues that the concept of normality, like that of race, is a biological error. Claiming that "Diversity of function is a fact of biology" (Amundson 2000, p. 33), he points out that the explanations of the "functional determinists" are not supported by evolutionary theory. There is no evidence that particular organs are designed to have particular functions. On the contrary, "[T]he disadvantages experienced by people who are assessed as 'abnormal' derive not from biology, but from implicit social judgments about the acceptability of certain kinds of biological variation" (Amundson, ibid, p. 33).

Wachbroit, however, supporting a functional account, acknowledges that variation has long been recognized and in fact is in itself normal as in blood group, for example, but suggests that the aim is to explain the *constraints on*, rather than the *presence of*, variation. Amundson suspects that the kind of variation that is allowed and acknowledged by the functionalists is functionally equivalent variation. He draws a distinction between the *level* of an individual's functional performance and the *mode or style* by which that performance is achieved. Functional determinists can allow for variation in the level of performance, but although they distinguish their view from the concept of the normal associated with the statistical average, the two come together: "Whatever the hierarchical level, functional determinism states that functions take place in a uniform mode at a relatively uniform performance level by a statistically distinctive portion of the members of a species. These are the normals" (Amundson 2000, p. 36).

Normality as Convention

In arguing against functional determinism, Amundson supports the view that social judgments construct ideas about what is normal. Turning to explicit examination of the view that normality is a social convention, there are again, at least two aspects to the claim that normality is a convention. It is possible to identify a "thin" account which simply says that what is normal varies with time and place – that what is normal today may be regarded as abnormal tomorrow: nothing is fixed. Beyond this a "thick" account sees an agenda behind the use of the normal to control.

With regard to both, it is important to consider the notion of a reference class for the application of the concept. "[S]omething cannot be normal without being a normal something or other" (Dupré 1998, p. 222). Whereas Amundson discussed different species, Dupré is interested in the idea of normal *people*. Further, the concept of normality "is a concept that relates individuals to a paradigm for the kind – the normal member of the kind" (Dupré 1998, p. 224), which may be determined statistically. Along with other commentators, for Dupré the drive behind our interest in *normality* in relation to human beings is in fact an interest in identifying *abnormality*. Variation exists not only in bodily measurements such as height and weight but also in our abilities; hence, it might be thought that there is a range of "normal" capacities, the lack of any of which could underlie a judgment of abnormality and/or disability. This might be the case particularly in a medical

context, where the identification of an abnormality or disability is typically the prelude to some ameliorative or curative intervention or at least management. Concepts of disability have been the focus of considerable debate in recent decades, and it is important to investigate the use of the normal/abnormal in this context.

There are very strong arguments for the view that disability is at least partly constructed by society and the environment: it is not just social *judgments* which construct our view of what is normal/abnormal or disability but the way in which society organizes itself, including in areas such as architecture and transport, which affect both the ways in which and the extent to which individuals can move about in the world. Dupré agrees that statements of normality and abnormality presuppose some state of the environment – in other words, it is necessary to look beyond variation between individuals: "most of the significant capacities of even fully able people are contextually determined" (Dupré 1998, p. 230).

When we turn to psychological characteristics of people and behavior, the situation is even more complicated. Homosexuality was notoriously once regarded as abnormal behavior and classified as a disease, with attempts made at "curing" it. Whether or not homosexuality has a biological basis or is a lifestyle choice, homosexuals have been successful in taking control of their self-definition. Dupré argues that "we can say that certain practices are normal for homosexuals or bisexuals, though not normal for heterosexuals. To ask whether they are normal for people is to make something akin to a category mistake" (Dupré 1998, p. 233).

As Dupré further points out, it does not follow that any sort of variation in human behavior can be rescued from classification as abnormal in the way that homosexuality has. He cites the example of "multiples" discussed by Hacking (1995) as a difficult case, because being a multiple cannot be accommodated within the range of behaviors assimilable by current modes of social organization. While it is not inconceivable that this might change, there is little prospect of it. Dupré concludes that for behavior to be normal, it must fall within the boundary that social organization and norms permit (Dupré 1998, p. 234). This is a normative decision: "normality in behaviour depends mainly on what we, as a society, decide to accept as normal" (Dupré 1998, p. 243), and this is culturally relative (ibid, pp. 244–245).

Normality as an Instrument of Power and Control

The idea that there is a boundary around what social organization and norms *permit* hints at the thicker view that the norm is an instrument of power. The ways in which norms become controlling vary in the extent to which they restrict the freedoms of those to whom they are applied. Historically, one of the most obvious freedom-limiting uses has been the different behaviors conventionally expected of men and women in various social contexts. An example which has had less significant effects on freedom of action, and which may seem trivial by comparison, has been the development of the norm of the western business suit in a large number of situations.

Arguably of more relevance to philosophy of medicine is the later work of Foucault in which he extended his ideas to the concept of biopower, by which populations, as well as individuals, are regulated at the biological level (Foucault 1976; Taylor, 2009). Statistical analysis, through the use of the bell curve, offers a technique for managing populations, in medicine as well as in other areas of social life, such as education, crime, and punishment.

Normality and Disease

Just as debates about the normal are associated with notions of disability, they are also implicated in concepts of health and disease. Jiri Vácha notes the association between the frequent and the normal and beyond that, the healthy (Vácha 1978). But the dichotomies of health/illness are also notoriously controversial.

In the context of illness and disease, the understanding of what is normal is an important background to the practice of *diagnosis*. Diagnosis is potentially a very powerful tool, which can trigger interventions of different degrees of intrusiveness. The fact that diagnostic practice changes over time, however, especially as regards mental illness, has lent credence to the view that what is normal is simply a social construction. The example of homosexuality has already been mentioned. Cases of young women being declared insane and detained because of promiscuity are less cited but constitute further evidence for that position, as does the historic abuse of mental illness diagnoses by political regimes. These all support the view that classifications of abnormality resulting in a diagnosis of mental illness can amount to a use of the normal as a mechanism of power and control.

Whereas there is a rich literature about the problematic status of mental illness and the use of diagnosis in that context to control, it might be thought that in the arena of physical disease, there would be less likelihood of this. Wendy Rogers, however, has shown how concepts of physical disease can operate to the disadvantage of specific population groups.

Rogers suggests that the ideal human being of medical textbooks is a 70 kg male with 32 teeth, no mental disorders, and a clean genetic slate (Rogers 1999). This can cause problems for women: an example is regarding menopause as an illness. The establishment of physiological norms through the use of statistics gives rise here, and elsewhere, to serious difficulties. In the "descriptive" account of disease, with "normal" being equivalent to "average":

> Physiological norms are derived from measuring a large number of parameters in a population and converting these into tabular forms with cutoff levels for the "normal". Usually, the limits of normal are defined as lying within two standard deviations of the mean. Anything outside this is by definition abnormal, irrespective of whether or not the person with the abnormal values feels ill. (Rogers p. 204)

But Rogers makes the point that any bodily parameter is always relative to individual conditions so that a level of hemoglobin is a level *in a certain person*, *of a certain gender*, *and of a certain age*.

According to Rogers one consequence of defining disease in this quantitative way is that the goal of treatment is then defined as returning the pathological parameters to normal levels. The subjective state of the individual is irrelevant (ibid, p. 205). But Rogers further argues that "even if we do accept statistical norms as the basis for some diseases, for other bodily parameters the statistical norm will not define the healthy" (ibid.).

On an alternative, overtly normative concept of disease, where people's subjective states are acknowledged to have a role, the goal of medicine is to restore people to their own version of normal (ibid, p. 207). If this analysis is applied to menopause, on the latter version menopause may be considered a disease because women experience it in that way. On a quantitative understanding of disease, however, all postmenopausal women are diseased. Rogers makes a convincing case for this being inappropriate: "In particular, conflating postmenopause, estrogen deficiency, and osteoporosis cannot be justified" (ibid, p. 214).

Obesity

The example of obesity demonstrates some of the difficulties of definition, the potential for control at a public health level, and the relationship between the statistical norm and "healthy." At the time of writing, there is what has been called an obesity "epidemic," with the population in many different countries becoming steadily larger. Surveys have illustrated that people who are regarded by the medical profession as overweight may not consider themselves to be so, because the average size has increased. The fashion industry has allegedly changed the definitions of dress sizes so that what counted as a size 12 in 1960, for example, was smaller than a dress marked size 12 today. Body shape is also changing. So what counts as normal weight? The medical profession has developed measures such a body mass index (BMI) and waist circumference as indicators of "healthy" weight, but these are not only controversial (e.g., they are problematic in relation to certain subpopulations) but also as the population increases in size these measures are less and less "normal" from a statistical point of view. Watching news footage from the 1950s, at a time of postwar rationing, it is very noticeable that the average body shape was very slim.

It is difficult to uphold the view, moreover, that there is an equivalence between "healthy" and "normal" from a statistical point of view. As the population develops higher and higher incidences of diabetes on the one hand and dementia on the other, these become statistically normal.

Normality and the Individual

Problems such as these add support to the view already hinted at, in the discussion of Rogers, that in the case of health and illness, and also for normal and abnormality, the idea of a "reference class" might be inappropriate. A similar point is made in the fictional work *A Cunning Man* by Canadian novelist Robertson Davies:

> . . .the popular idea is of health as a norm to which we must all seek to conform. . .But are there not as many healths as there are bodies? If whatever we are demands certain physical frailties, why struggle to get rid of them? (Davies 1994, p. 248)

Vácha also suggests that the concept of species normality be replaced by the concept of individual normality or responsiveness (Vácha 1978).

It is important then to consider the individual, and indeed it is a common way of speaking, when someone has suffered illness, to talk of them getting "back to normal," when they recover. Again, as stated above, one interpretation of disease understands the goal of treatment to be to restore an individual to their own version of the normal. A moment's thought reveals that this is problematic for a number of reasons. Recovery does not take one back to the situation before the illness: there will be scars, traces, antibodies – the illness is within the memory of the body. The person may be able to function without any noticeable difference from how they did before, but they are not, nevertheless, the same. This raises questions for what we mean when we speak of being "back to normal." Clearly, when we are speaking in this way with reference to a single individual, the idea of the norm as a statistical average ceases to apply, unless we mean that after recovery, the individual is able to function in ways akin to the average person in the population.

This raises a further question, however, as to whether, when we are speaking of being back to normal, we should properly refer to the function of the organism as a whole or to the organ(s) or parts of the body affected. The type of illness episode will surely be relevant here. If a patient has had an injury to their right hand, what being "back to normal" might mean is that he or she can use the hand again, freely, to type, to play tennis, and so on. Rachel Cooper, however, has spoken of the inappropriateness, in many cases, of referring only to the organ affected: as she points out, examples such as Slijper's goat suggest we should consider the individual as a whole (Cooper 2012). The functioning of the hand may not be the only issue – it may look different and give rise to occasional pain, even though to all intents and purposes it is functioning without any problems as a hand should. This discussion is reminiscent of the argument in Amundson about the extent of variation in mode of performance as well as level. Individual members of a species can in some cases accommodate extensive variation, through compensation for the poor functioning of one organ within the context of a whole body.

The case of the patient who had a hand transplant only to ask later that it be removed is an interesting illustration of some of the difficulties of "bodily integrity" following such a transplant (see e.g., Slatman and Widdershoven 2010). What counts as normal here is open to a number of different interpretations. It is normal, in terms of statistical average, for humans to have two hands that function. It is not normal, however, for humans to have a hand that is transplanted from another human being. When we speak of what is normal for the individual, understood in terms of function, the best approximation to what was normal for the individual who had lost a hand could be interpreted as supplying a transplanted hand which could function as a hand should. But for the individual concerned, the hand may not be accepted as part of them – it may not fit into their understanding of what was normal for them.

Here we have competing versions of getting "back to normal" for an individual. For the individual it may be better, ultimately, to have no hand than a transplanted hand. "Back to normal" frequently if not always requires adjustment to the new situation, with forms of compensation provided by the body and mind which enable the individual as a whole to function.

Normality and Health-Care Ethics

It has become apparent that the use of the concept of normality has ethical implications, whichever interpretation of the concept is adopted. Even if it is not used in ways that explicitly disadvantage particular groups, or to attempt to control behavior, it is commonly used as a bridge for the fact-value gap. It has also been prominent in discussions of the goals of medicine, appropriate interventions, and distributive justice in relation to health care, notably in the work of Norman Daniels. Daniels has argued that the preservation and restoration of normal function is a primary goal of health care (Daniels 1985). In his work on distributive justice, the concept of the "normal opportunity range" also plays a major part. For Daniels what is important is the link between normality/abnormality and access to the normal opportunity range open to humans, and so justice requires efforts to facilitate such access.

There are issues about the link, if any, between this normal opportunity range, as defined by Daniels, and judgments of quality of life. Critics suggest that people who apparently suffer reduction in access to the normal opportunity range, due to illness or disability, can report experiencing high quality of life. Once again the issue of construction of disability and the relation between internal and external perspectives arises. The idea of the normal opportunity range has to compete with other criteria for just distribution of resources, and even if this concept is not appealed to explicitly, there are other ways in which the "normal" can influence distribution decisions – the idea of "premature death," for example, presupposes a certain expectation of normal life span.

Can Medicine Dispense with the Concept of the Normal?

Despite the problems noted by Rogers and others, a future in which the "normal" might not be used in medicine appears challenging. The distribution curve for parameters, such as body temperature, BMI, and cholesterol levels, gives at least a starting point for assessment. Otherwise, how is a physician to begin assessing a patient? It is important, however, to be sensitive to the extent of variation that is possible in human beings, not only in level but in mode, as Amundson pointed out (Amundson 2000), and to recognize that there may be significant differences not only between population groups but also between individuals.

It may be the case that as society moves further down the track toward the implementation of personalized or precision medicine, the concept of the normal will

lose some of its traction. The idea that medicine can and should be "tailored" to the individual, which requires precise measurements of that individual, involving not only genomic but multi-omic data, perhaps suggests this. Even this prospect, however, may require the use of a reference class, built up from the "big data" that are now being collected. The concept of the normal may therefore have an ongoing application but becomes potentially dangerous when it is used to downplay the importance of individual variation, to evaluate such variation negatively where there is no need to do so, and to try to mould individuals to the perceived norm rather than embracing difference.

Definition of Key Terms

Bell curve	A graph with a bell-shaped curve, identified with the normal distribution of a characteristic in a reference class
Functional determinism	The view that biological organs in a species have a specific function
Normality	The state of being normal, whether in relation to physiological or psychological characteristics, where normal is commonly defined in terms of the statistically normal distribution
Reference class	The class to which an individual is deemed to belong for the purpose of making judgments of normality

Summary Points

- There are two broad categories of definition of the normal, scientific, and normative.
- Within each category there are subdivisions.
- Scientific interpretations of the normal include the statistical and the functional.
- Normative interpretations include a thin and a thick version.
- A thin normative interpretation points to the fact of variation.
- A thick normative interpretation appeals to evidence that the normal has been used as an instrument of power and control.
- The use of the normal clearly has ethical implications, for example, in distributive justice.
- Although the importance of the normal may lessen in the light of developments in personalized medicine, it is likely to continue to be needed in the practice of medicine.

References

Amundson R (2000) Against normal function. Stud Hist Philos Biol Biomed Sci 31(1):33–53
Cooper R (2012) Being ill and getting better: recovery and accounts of disorder. In: Rudnick A (ed) Recovery of people with mental illness: philosophical and related perspectives. OUP, Oxford
Daniels N (1985) Just health care. Cambridge University Press, Cambridge, UK
Davies R (1994) A cunning man. Penguin USA, New York
Davis LJ (1995) Enforcing normalcy. Verso, London
Dupré J (1998) Normal people. Soc Res 65(2):221–248
Foucault M (1976) The history of sexuality. Penguin, Harmondsworth
Hacking I (1990) The taming of chance. University Press, Cambridge
Hacking, I (1995) Rewriting the soul: multiple personality and the sciences of memory Princeton: University Press
King CD (1945) The meaning of normal. Yale J Biol Med 17:493–501
Rogers WA (1999) Menopause: is this a disease and should we treat it? In: Donchin A, Purdy L (eds) Embodying bioethics: recent feminist advances. Rowman and Littlefield, Lanham, pp 203–217
Slatman J, Widdershoven G (2010) Hand transplants and bodily integrity. Body Soc 16(3):69–92
Taylor D (2009) Normativity and normalization. Foucault Stud 7:45–63
Vácha J (1978) Biology and the problem of normality. Scientia 113:723–846
Wachbroit R (1994) Normality as a biological concept. Philos Sci 61(4):579–591
Wiesing U (2009) The history of medical enhancement: from *restitution ad integrum* to *transformation ad optimum*? In: Gordijn B, Chadwick R (eds) Medical enhancement and posthumanity. Springer, Dordrecht, pp 9–24

On Concepts of Positive Health

3

Lennart Nordenfelt

Contents

Abstract

This chapter presents some main interpretations of the concept of health from antiquity until today. There is an emphasis on ideas of health as a positive notion, i.e., as something over and above the absence of disease. After giving a summary of some classic intuitions about health, the chapter concentrates on contemporary attempts to analyze health in terms of well-being and ability. Starting with the famous WHO definition of health, where health is understood as complete physical, mental, social, and spiritual well-being, the chapter turns to a

L. Nordenfelt (✉)
Ersta Sköndal University College, Stockholm, Sweden
e-mail: lennart.nordenfelt@esh.se

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_2

presentation of some scholarly analyses of health, where health is mainly analyzed in ability terms. Examples are taken mainly from the philosophical literature but also from other disciplines. It is noted that almost all definitions in the nursing and feminist literature understand health in positive terms in contrast to such naturalist definitions as present health as merely the absence of disease. The chapter further contains a discussion of health as a culture-dependent notion and makes a comparison between the concepts of human and animal health.

Introduction

It is often maintained that health is one of the major goals of medicine or even *the* goal of medicine. This idea has been eloquently formulated by the American philosophers of medicine Edmund Pellegrino and David Thomasma in their book *Philosophy as the Basis of Medicine* (1981, p. 26):

> Medicine is an activity whose essence lies in the clinical event, which demands that scientific and other knowledge be particularized in the lived reality of a particular human for the purpose of attaining health or curing illness through the direct manipulation of the body and in a value-laden decision matrix.

Although certain other goals of medicine exist, such as the basic goal of saving lives and the recently developed goal of quality of life, health is, indeed, the foremost goal of medicine and public health (Pellegrino and Thomasma 1981; Callahan and Hanson 1999). Health also has a prominent position in many life contexts and is a crucial condition for maintaining and executing a profession, for enjoying leisure activities, and indeed for living a good life in general. It is a formal prerequisite for performing certain tasks or taking up certain occupations, such as that of soldier, police officer, or firefighter. More compelling is the place of mental health as a condition for moral and criminal culpability.

However, the formidable task of interpreting the nature of health remains before us. What exactly *is* health? To what more precise goal shall we direct our efforts in medicine and health care?

These questions are not simply academic. They are of great practical and thereby ethical concern. The consequences for health care diverge considerably, not least in economic but also in social and educational terms, depending on whether health is understood as people's happiness, as their fitness and ability to work, or as just the absence of obvious pathology in their bodies and minds. There are adherents of all these interpretations in the modern theoretical discussion on health.

Etymologically, health is connected with the idea of wholeness. This is evident in the verb "heal," with the sense of regaining wholeness. The healthy person is a person who is whole in the sense of having all the properties that should pertain to a human being. Health has thus traditionally been viewed as an ideal notion, a notion of perfection that very few people, if any, can completely attain. Today health also sometimes functions as an ideal notion. This is indeed the case with the formulation of health by the World Health

Organization in its initial declaration, published in 1948: "Health is a state of complete physical, mental and social well-being and not only the absence of disease or injury."

The notion of health is the object of scientific study from several points of view and within several disciplines. Besides research in medicine, public health, nursing, and associated disciplines, there is health-related research in anthropology, psychology, sociology, and philosophy. In some of these disciplines the focus is on a particular aspect of the notion: in psychology, the experience of health and illness; in anthropology and sociology, health and illness as factors of social importance. Philosophical analyses of health have often involved an attempt to formulate global definitions of the idea, and in the following most references will be to philosophical theories of health.

The Varieties of Health

Health, thus, is a notion primarily applicable to the human being as a whole. On the other hand there are more specific derivative notions. Ever since antiquity, and reinforced by the Cartesian distinction between body and mind, it has been natural to separate somatic health from mental health. The interpretations of mental health have varied over time. The ancient notion of mental health was closely connected to morality, whereby the mentally healthy person was a person who lived a virtuous life, but this notion has lost most, though not all, of its significance today. The idea of spiritual health is also current in the health sciences and is evident in the current formulation of the health concept of WHO (see below) although it is not systematically recognized. Bernhard Häring (1987) is a leading spokesman for a notion of health including a spiritual dimension: "A comprehensive understanding of human health includes the greatest possible harmony of all of man's forces and energies, the greatest possible spiritualization of man's bodily aspect and the finest embodiment of the spiritual" (154).

The various categories of health have connections to each other. Sometimes bodily health has been given priority in the sense that it has been viewed as a prerequisite for mental health. Galen (ca. 129–216/7) in some of his writings attempted to explain mental properties of the person in terms of specific mixtures of the bodily parts (Galen 1997). Consider also the ancient proverb *mens sana in corpore sano* (a healthy mind in a healthy body). In the modern discussion about mental illness, one position, favored in particular by doctors, is that all mental illness has a somatic background, i.e., all mental illnesses – if they exist at all – are basically somatic diseases (Szasz 1974). The customary view, however, also in Western medicine, is that a person can at the same time be somatically healthy and mentally ill, or vice versa.

The Latitude of Health

Since antiquity theorists of health have emphasized that the health–ill-health dichotomy is not represented by two opposite states. There is instead a dimension or latitude of health from optimal health to maximal ill-health. According to this idea a person can be in a state that is far from optimal health but still be healthy. Likewise, people's states

of ill-health can vary between mildly ill and seriously ill. Galen is perhaps the philosopher who has contributed most to the analysis of the latitude of health. He made more distinctions in this respect than are customary today. He not only distinguished between health and ill-health but also acknowledged a state in between these called the neutral state. This means that, according to Galen, a person can be neither healthy nor ill but instead in a neutral state along the health–ill-health dimension. This idea concerning latitude of health was developed in several sophisticated ways in the medieval medical discussion. (For a thorough analysis, see Ottosson 1982.)

Some Classic Theories of Health

Health as Absence of Disease; the Idea of a Natural Function

Although health is often described in nonmedical terms and with reference to nonmedical contexts, it has its primary place and function as a medical concept. Health in the medical arena is contrasted in particular with disease but also with injury, defect, and disability. Culver and Gert (1982) have adopted the term "malady" to cover the negative antipodes of health. In many medical contexts (Hesslow 1993) and in several philosophical reconstructions of the notion of health (Boorse 1977, 1997), health has been defined as the absence of diseases or the absence of maladies. Thus the perfectly healthy person, according to this analysis, is the person who does not have any diseases or maladies. This is the highly influential negative definition of health which is in focus in another chapter of this Handbook. The present chapter, in contrast to this, is focused on concepts of positive health.

In many contributions to the theory of health a distinction is made between the concepts of disease and illness (Boorse 1975; Twaddle 1993; Fulford 1989). The general idea behind this distinction – although the distinction has been drawn in different ways by different authors – is that a disease is a deranged process in the person's body whereas an illness is the person's negative experiences, for instance, pain or anguish, resulting from the disease. In addition, some theories include disability in illness (see below). The distinction between disease and illness has proved useful in several contexts, including the clinical one (Hellström 1993), for separating the disease as a pathological phenomenon from its impact on the person as a whole.

Health as Balance

An extremely powerful idea in the history of medicine is that positive health is constituted by bodily and mental balance. The healthy person is a person in balance, normally meaning that different parts and different functions of the human body and mind interlock harmoniously and keep each other in check. The Hippocratic and Galenic schools (Hippocrates 460–380 BD and Galen 129–216/7 AD) were the first Western schools to develop this idea in a systematic way. A healthy body was seen as

one where the primary properties (wet, dry, cold, hot) of the body balance each other. In the medieval schools, following Galen, this idea was popularized and formulated in terms of a balance between the four bodily humors: blood, phlegm, yellow bile, and black bile (Galen 1997).

The idea of balance is strong in several non-Western medical traditions. The Ayurveda tradition in India, for instance, declares that there are three humors acting in the body, the breath (*vata*), the bile (*pitta*), and the phlegm (*kapha*). The proportions of the three humors vary from person to person, and their actions vary according to the season, the environment, and the person's lifestyle and diet. In good health the humors are in equilibrium. Disease is the result of their imbalance (Singhal and Patterson 1993).

Balance is a powerful idea also in modern Western thought, in particular within physiology. The idea is then often to be recognized under the label of *homeostasis* (the Greek word for balance). Walter Cannon's (1871–1945) classical work on homeostasis (1932) describes in detail how the various physiological functions of the body control each other and interact in feedback loops in order to prevent major disturbances.

The idea of balance or *equilibrium* (the Latin word for balance) has a rather different interpretation in the writings of Ingmar Pörn (1993). Here balance is a concept pertaining to the relationship between a person's abilities and his or her goals. The healthy person, according to Pörn, is the person who can realize his or her goals and thus retain a balance between abilities and goals (cf. section "Health as Ability," below).

The Ideas of Health as Well-Being and Ability

In addition to the interpretation of health as balance between primary qualities Galen gave a holistic interpretation of health on the level of the person: *Health is a state in which we neither suffer from evil nor are prevented from the functions of daily life* (Galen: *De sanitate tuenda*, I, 5; cited in Temkin 1963, p. 637). In this formulation Galen includes two thoughts which have had a central position in the philosophy of health right up to our own day: health is well-being in some sense, and health entails that the person has a basic ability to perform actions which are crucial to the person, in particular in his or her daily life.

Well-being and ability as realized in a person can also be seen as criteria for the existence of bodily and mental balance in this person.

Health as Well-Being

The WHO definition of health can be said to represent a strong tradition in the philosophy of health where well-being is the central concept. Positive health is identified with well-being or happiness; illness is understood as suffering or pain. The first formulation, from 1948, is the following:

> Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.

In the conference on health promotion (WHO 1996) this definition was enlarged to include spiritual health:

> Health is a dynamic state of complete physical, mental, social and spiritual well-being, and not merely the absence of disease or infirmity.

It should be noted that this extremely positive characterization of health still has a conceptual connection to the notions of disease and infirmity. A necessary condition for complete health, according to this definition, still is that no disease or infirmity is present.

Although the WHO definition has been much criticized and has had rather little clinical significance, its impact on the rhetoric of health policy has been enormous and is frequently cited in various official health contexts.

The idea of identifying health with complete well-being is slightly modified in the WHO's Ottawa Charter for Health Promotion (WHO 1986). In this characterization there is indeed a reference to the basic WHO definition, but there is an interesting addition pointing in the direction of an ability-oriented definition of health (see below):

> Health is, therefore, seen as a resource for everyday life, not the objective of living. Health is a positive concept emphasizing social and personal resources, as well as physical capacities. Therefore, health promotion is not just the responsibility of the health sector, but goes beyond healthy lifestyles to well-being.

Problems Concerning the Identification of Health with Well-Being

Critical points can be raised with regard to such characterizations of health as are solely in terms of a person's well-being. It is clear that well-being (in some sense) is compatible with disease. A person can feel well and have a disease, even a serious disease in its initial stage. This is sometimes cited as a counterintuitive proposition. The general idea that health is related to well-being can, however, be modified to cover this case too. The individual with a serious disease, it may be argued, will sooner or later have negative experiences such as pain, fatigue, or anguish. Thus the ultimate criterion of a person's health could be seen as his or her present *and future* well-being. (For a different approach suggesting that complete health is compatible with the existence of disease, see Nordenfelt 1995, 2001.)

The WHO definition, which keeps a connection to the absence of diseases and infirmities, has been criticized for other reasons. One set of criticisms is related to the word "complete." Several writers have noted that this euphoric definition is readily falsified. It has been shown that in a 14-day period the average adult experiences about four symptoms. Viewed in this light we are all ill! (Larson 1996).

It must be admitted, though, that, however health is ultimately characterized, a set of feelings are normally related to it. The person who is healthy normally has a

number of positive experiences associated with being so, the person who is unhealthy a number of negative ones. The difficult question to settle is whether these experiences *constitute* the state of health in question or whether they are just normally associated with it. WHO in the basic documents obviously settles for the constitutional idea. So do various phenomenological accounts (see below).

It is a difficult task to characterize the well-being purported to constitute health. If one includes too much in the concept, as in the WHO context, there is a risk of identifying health with happiness. As many critics have said, health cannot reasonably be identical with complete physical, mental, and social well-being. The absurd conclusion from this conception could be that all people who are not completely successful in life are to be deemed unhealthy. A way of solving the problem concerning what kind of well-being constitutes health would be to say that it only concerns what may be described as *inherent* well-being (which is to say, well-being directly connected with the person's body and mind) and not such well-being as is directly dependent on particular circumstances. In this case, a person's happiness about, for instance, having had a baby or earned a fortune would not be a part of the person's health.

Some authors (Gadamer 1993; Leder 1990) have pointed out that phenomenological health (or health as experienced) tends to remain as a forgotten background. Health is in daily life hardly recognized at all by its subjects. People are reminded of their previous health only when it is being disrupted, when they experience the pain, nausea, or anguish of illness. Health is "felt" only under special circumstances, the major instance being after periods of illness when the person experiences relief in contrast to the previous suffering.

Feelings and the Notion of Subjective Health

It may therefore be argued that it is not a plausible strategy to *identify* health (in general) with a set of feelings. On the other hand, it is more than plausible to identify one aspect of health, viz. its subjective aspect, with a set of feelings. Some theorists make the distinction between objective health and subjective health, where the latter refers to the subject's beliefs and feelings. A person who feels well or believes that he or she is well is subjectively healthy, according to this line of thought. The converse holds for subjective ill-health or subjective illness.

Subjective ill-health is obviously holistic in nature. Here it must be the feeling person as a whole who is ill. Some theorists (for instance, Canguilhem 1978; Marinker 1975; Twaddle 1993; Young 1982; Hofmann 2013) strongly emphasize this element of experience and wish to make it a prerequisite for using the illness label. Also in the phenomenological literature there is an understanding of health and illness in terms of the person's *experiences* or at least in terms of his or her subjectivity. As Fredrik Svenaeus 2012, p. 103, puts it, "To be ill means not to be at home in one's being in the world, to find oneself in a pattern of disorientedness, resistance, helplessness and perhaps even despair, instead of in the homelike transparency of healthy life."

However, although well-being or absence of ill-being is an important trait in health, most modern positive characterizations of health have focused on other traits. One such trait is health as a condition for action, i.e., ability.

Health as Ability

A number of authors in modern philosophy of health have emphasized the place of health as a foundation for achievement (Parsons 1978; Whitbeck 1981; Seedhouse 1986; Nordenfelt 1995, 2001; Fulford 1989). In fact they argue, in partly different ways, that the dimension ability/disability is the core dimension determining whether health or ill-health is the case. A healthy person has the ability to do what he or she needs to do, and the unhealthy person is prevented from performing one or more of these actions. There is a connection between this conception and the conception that illness entails suffering. Disability is often the result of feelings such as pain, fatigue, or nausea. Conversely, disability is often the cause of some suffering.

The formidable task for these theorists is to characterize the set of actions that a healthy person should be able to perform. Parsons (1972) and Whitbeck (1981) refer to the subject's wants (i.e., the healthy person's being able to do what he or she wants), Seedhouse (1986) to the person's conscious choices, and Fulford (1989) to such actions as could be classified as "ordinary doings." Fulford, who pays most of his attention to the negative notion of illness, declares that "the patients who are ill are unable to do everyday things that people ordinarily just get on and do, moving their arms and legs, remembering things, finding their ways about familiar places, and so on" (1989, p. 149). Nordenfelt settles for what he calls the subject's vital goals. These goals need not be consciously chosen (also babies and people with dementia have vital goals) – their status as vital goals derives from the fact that they are states of being which are necessary conditions for the person's minimal happiness in the long run. Health in Nordenfelt's theory is thus conceptually related to, but indeed not identical with, happiness.

Although it is evident that health, as ordinarily understood, is connected with ability, and ill-health with disability, one may still doubt whether the dimension ability/disability can remain the sole criterion of health/ill-health. An important argument concerns those disabled people who are not ill, according to the common understanding, and who do not consider themselves to be ill. These people are to be classified as unhealthy according to the ability theories of health.

One answer to this question (Nordenfelt 2001) is that disabled people need not be unhealthy if their disability is established merely according to conventional measurements. There are several standardized instruments for the measurement of disability today (see, for instance, the so-called DALY instrument, Reidpath et al. 2003). People are unhealthy, says Nordenfelt, only if their disability is established in relation to their individual vital goals. Moreover, a disabled person need not, of course, have any disease. Both disability and ill-health in Nordenfelt's system are compatible with the absence of diseases. Health can be reduced by other

causes than maladies. Another answer, proposed by Fredrik Svenaeus, 2001, is that there is a phenomenological difference between the disabled unhealthy person and the disabled healthy person. The unhealthy person has a feeling of not being "at home" with regard to his or her present state of body or mind. This feeling is not present in the case of the disabled in general.

Observe also that the notion of disease (or malady) will have a slightly different connotation given a concept of health centered on ability or well-being than it has according to the naturalistic account. For the naturalist Boorse, who defines health in terms of the absence of diseases, a disease is a dysfunction in relation to the survival of the individual or the species, while for the theorist who defines health in terms of health or well-being a disease is a dysfunction in relation to the individual's ability or well-being.

Nursing and Feminist Characterizations of Health

It is striking that almost *all* characterizations of health in the nursing and feminist literature are of a holistic kind, referring to a person's well-being and/or abilities. Some simply adopt the WHO definition.

The following examples are attributable to the nursing theorists Hildegard Peplau and Imogene King, respectively:

> Health is a word symbol that implies forward movement of personality and other ongoing human processes in the direction of creative, constructive, productive, personal, and community living.
>
> Health implies continuous adaptation to stress in the internal and external environment through optimum use of one's resources to achieve maximum potential for daily living. (See Marriner-Tomey 1994, p. 310 and 329)

These examples illustrate the fact that the intuitions about health which are to be found in the nursing literature and at least in part of the feminist literature go in a positive or holistic direction. It may also be noted that in this literature no conditions as to the absence of disease or infirmity are raised.

Some theorists contend that the way we define and in general look upon health and health care is dependent on our gender (Oakley 1993). This difference is well reflected in the traditional health professions. The traditional doctor is a man who is basically concerned with the physical condition of his patients. He sees his primary task as being to cure the diseases of the patient by use of well-established treatments, often in the form of surgery and drugs. To the traditional doctor a functional analysis of health and disease is the natural choice. The traditional nurse is a woman who is basically concerned with the general well-being of the patient. She sees her primary task as being to care for the person as a whole. Caring, for her, means above all "relating to the ill person as a whole person whose psyche is equally involved with her or his soma in the illness in question" (Oakley 1993, p. 40). Thus one may say that a holistic analysis of health is the most natural one to the nurse.

Several feminist writers also argue in the direction of a positive definition of health:

> There is also good reason to reject the view that health is nothing but the absence of such infirmity, in favor of the claim that it requires positive well-being. Discovering what positive well-being might mean for women in a world without sexism is a project to which we all must turn. (Purdy 1996, p. 177)

There is an interesting notion often encountered in feminist literature and labeled as "reproductive health." In using this term authors often refer to circumstantial factors improving or sustaining people's general health. Reproductive health is promoted if people are given the opportunity to enjoy a secure and positive sexual life, through legal abortion, efficient contraceptive devices, and protection against sexual violence. However, there is also a formal definition of reproductive health in holistic terms:

> The purposes of sexual health care should be the enhancement of life and personal relationships, and not merely counselling and care related to procreation or sexually transmitted infections. Reproductive health implies that people are able to have a responsible, satisfying and safe sex life and that they have the capability to have children and the freedom to decide if, when and how often to do so.

This definition of reproductive health was officially adopted by the WHO member countries and confirmed at the fourth International Conference on Population and Development in Cairo in (WHO 1994).

There are several other gender-dependent issues in the philosophy of medicine and health. Most of them concern circumstantial and causal factors behind health and ill-health. One concerns men's violence against women and the consequences for women's health. The same holds for issues such as the discrimination against women in various sectors of society. For one thing women are often prevented from taking leading positions and having high salaries. Women often find themselves in a situation where they have double occupations: paid work plus a substantial responsibility for the family household. As a consequence there are also notable differences in the disease panorama. Women's paid work may contain different kinds of stress which have an effect particularly on mental rather than physical well-being. Two other factors are the discrimination against women and sexual harassment. Likewise, poverty among women may cause ill-health and restricts opportunities to improve health.

Health and Cultural Relativism

If health is a value-laden concept, then, some would argue, there are differences in the interpretation of health between cultures both historically and geographically. It is important to note that these differences can be quite profound.

The concepts of health can vary from culture to culture because there are fundamental differences in the basic philosophy of health and health care, as between Western medicine and traditional Chinese medicine or the traditional

Indian *Ayurveda* medicine. Western medicine, which is to a great extent based on a naturalistic philosophy of man, arrives easily at a naturalistic understanding of health, whereas oriental schools with a holistic understanding of man in a religious context derive a notion of health which incorporates forces and developments that are partly supernatural.

The ways of and reasons for ascribing health to people may, however, vary even if there is a basic common theory of health and disease. Consider a particular physiological state, the state of lactase deficiency, which has the status of disease in a Western country but not in most North African countries. Lactase deficiency causes, in combination with ordinary consumption of milk, diarrhea and abdominal pain. Thus in Western countries, where people ordinarily drink milk, lactase deficiency will typically lead to illness. Therefore this state ought to be included in a list of diseases in these countries. In North Africa, however, people rarely drink milk. Therefore lactase deficiency seldom leads to illness. Consequently it would be misleading to consider lactase deficiency a disease in this part of the world.

What makes the difference between the Western and the African cultures in this example is not necessarily different concepts of disease. It could be a question of different lifestyles and different environments judged from the point of view of a single concept of disease.

Given an ability-based concept of health there may be further cultural dependencies, since a particular ability is always ability in relation to a certain background. A background may enable a person to perform a certain action but may also make this action difficult, or even prevent the person from performing it. It is more difficult to build a house on the Himalayas than on the plain further south in India, so the claim that a person has the ability to build a house amounts to very different things in these two regions. In addition to the physical environment we have the societal environment. A society can enable one to pursue a particular course of action, or render it impossible. What, then, are the consequences of this for the concept of health?

Assume that G is a crucial goal common to most people in both the societies S1 and S2 and that G is more easily fulfilled in S1 than in S2. The psychophysical resources needed for realizing G in S2 are much greater than in S1. Being able to achieve G in S1 is thereby significantly different from being able to achieve G in S2. And, therefore, being healthy in S1 (at least as far as this variable is concerned) means something different from being healthy in S2. (For a discussion about health as a culture relative concept, see Khushf (2001).)

Health of Human Beings Versus Health of Animals and Plants

Health, disease, and the other central medical concepts are not used only in the human context. We ordinarily ascribe health and disease also to animals and plants. Do we then apply the same concept of health?

In this case the answers differ. The naturalists, who relate health solely to survival and reproduction, can easily transpose their concept to the world of animals and

plants. (For such a view see, for instance, Broom 1993.) The same could hold for balance theorists. It is more problematic to use the idea of health as ability or, even more, the idea of health as well-being all over the world of animals and plants. This can serve as an argument in favor of the naturalistic account. On the other hand it can be argued that there is an enormous difference between the human context and the context of other living entities. Human beings live in complex societies with complex demands and with a system of health care that is supposed to serve these demands. Health is important for a human being because it enables him or her to engage in crucial activities, such as work, political activities, and leisure activities – and, not least, to engage in close human relations such as friendship and love. Thus the health concept that is of interest to most people is a holistic one embracing all such relevant abilities. Therefore, it is no wonder that the concept of human health has evolved in directions quite different from those taken by the concepts of health concerning animals and plants.

Definitions of Key Terms

Health	Health is a central, but at the same time controversial, concept in medicine, health care, and health promotion. In modern philosophy of medicine there are two main lines of thought. One is of the naturalist kind, where health is seen as the absence of disease. Disease in its turn is defined as a reduced biological function. The other line of thought is normally called normativist or holistic. According to this idea health is a value-laden concept referring to the well-being or ability of the human being as a whole.
Subjective health	It is common to distinguish between health in general (or objective health) and subjective health. The latter notion is connected to the subject's beliefs and feelings. A person who feels well or believes that he or she is well is subjectively healthy, according to this line of thought.
Balance	According to a powerful idea in the history of medicine the healthy person is one who is in bodily and mental balance. This idea has been interpreted in several ways. The ancient writer Galen, for instance, viewed health as the balance between the primary properties of the body. In modern physiology the favored term is *homeostasis*, entailing that the various physiological functions of the body control each other and interact in feedback loops. A completely different idea is that health is constituted of a balance (often called *equilibrium*) between a person's abilities and his or her goals.
Well-being	The term "well-being" can be used to refer to the whole spectrum of positive feelings, from sensual pleasure to spiritual happiness. When the notion of well-being is used for the

	purpose of defining health, one normally has in mind a kind of well-being that is directly connected to the person's body and mind and not merely the result of immediate positive external influence.
Ability	A person's ability constitutes, together with his or her opportunity, the person's practical possibility to perform a set of actions. The ability part consists of such conditions internal to the person's body and mind as are required to reach an essential goal. In the philosophy of health there exist a number of attempts to specify the goals of such abilities as are required for a person's being healthy.
Culture relativism	The term "culture relativism" in relation to health is understood in at least three different senses in this article: (1) health as dependent on culture-relative philosophies of man, (2) health as dependent on culture-relative habits, and (3) health as dependent on culture-relative platforms for action.

Summary Points

- Health is the major goal of medicine.
- Health is primarily applicable to the human being as a whole.
- There is a dimension or latitude of health from optimal health to maximal ill-health.
- According to a classical idea health is constituted by bodily and mental balance.
- In contemporary philosophy of health there are two main lines of thought: naturalism and normativism (holism). According to the former health is the absence of diseases. According to the latter health is constituted by the subject's well-being and ability.
- The WHO concept of health refers to complete well-being.
- The concept of subjective health refers to the subject's feelings and beliefs.
- Some influential contemporary philosophers of health have developed theories where the subject's ability to reach essential goals is in focus.
- Nursing and feminist characterizations of health all go in the normativist (holistic) direction.
- There are cultural differences in the interpretation of health. These can be dependent both on different ideologies and different life opportunities.
- Health is a concept applicable also to animals and plants.

References

Boorse C (1975) On the distinction between disease and illness. Philos Public Aff 5:49–68
Boorse C (1977) Health as a theoretical concept. Philos Sci 44:542–573

Boorse C (1997) A rebuttal on health. In: Humber J, Almeder R (eds) What is disease? Biomedical ethics reviews. Humana Press, Totowa
Broom DM (1993) A usable definition of animal welfare. J Agric Environ Ethics 6(Suppl 2):15–25
Callahan D, Hanson MJ (1999) The goals of medicine: the forgotten issues. Georgetown University Press, Washington, DC
Canguilhem G (1978) On the normal and the pathological. D. Reidel Publishing Company, Dordrecht
Cannon WB (1932) The wisdom of the body. Norton, New York
Culver CM, Gert E (1982) Philosophy in medicine: conceptual and ethical issues in medicine and psychiatry. Oxford University Press, Oxford
Fulford KWM (1989) Moral theory and medical practice. Cambridge University Press, Cambridge
Gadamer H-G (1993) Über die Verborgenheit der Gesundheit. Suhrkamp Verlag, Frankfurt am Main
Galen (1997) Selected works, translated with an introduction and notes by P.N. Singer. Oxford University Press, Oxford
Häring B (1987) Medical ethics. St Paul, Middlegreen/Slough, tricky
Hellström O (1993) The importance of a holistic concept of health for health care: examples from the clinic. Theor Med 14:325–342
Hesslow G (1993) Do we need a concept of disease? Theor Med 14:1–14
Hofmann B (2013) On the downplay of suffering in Nordenfelt's theory of illness. Health Care Anal 21:283–297
Khushf G (2001) What is at issue in the debate about concepts of health and disease? Framing the problem of demarcation for a post-positivist era of medicine. In: Nordenfelt L (ed) Health, science, and ordinary language. Rodopi, Amsterdam, pp 123–169
Larson JS (1996) The World Health Organization's definition of health: social versus spiritual health. Soc Indic Res 38:181–192
Leder D (1990) Clinical interpretation: the hermeneutics of medicine. Theor Med 11:9–24
Marinker M (1975) Why make people patients? J Med Ethics 1:81–84
Marriner-Tomey A (ed) (1994) Nursing theorists and their work, 3rd edn. Mosby, London
Nordenfelt L (1995) On the nature of health: an action-theoretic approach, 2nd edn. Kluwer, Dordrecht
Nordenfelt L (2001) Health, science and ordinary language. Rodopi, Amsterdam
Oakley A (1993) Essays on women, medicine and health. Edinburgh University Press, Edinburgh
Ottosson P-G (1982) Scholastic medicine and philosophy. Bibliopolis, Naples
Parsons T (1972) Definitions of health and illness in the light of American values and social structure. In: Jaco EG (ed) Patients, physicians, and illness. The Free Press, New York, pp 107–127
Pellegrino ED, Thomasma DC (1981) A philosophical basis of medical practice. Oxford University Press, Oxford
Pörn I (1993) Health and adaptedness. Theor Med 14:295–304
Purdy LM (1996) A feminist view of health. In: Wolf S (ed) Feminism and bioethics: beyond reproduction. Oxford University Press, Oxford, pp 163–183
Reidpath DD, Allotey PA, Kouame A, Cummins RA (2003) Measuring health in a vacuum: examining the disability weight of the DALY. Health Policy Plan 18:351–356
Seedhouse D (1986) Health: foundations of achievement. Wiley, Chichester
Singhal GD, Patterson TJS (1993) Synopsis of Ayurveda, based on a translation of the Treatise of Susruta. Oxford University Press, Delhi/Oxford
Svenaeus F (2012) What is phenomenology of medicine? Embodiment, illness, and being in the world. In: Carel H, Cooper R (eds) Health, illness & disease: philosophical essays. Acumen, Durham, pp 97–112
Szasz T (1974) The Myth of Mental Illness: Foundations of a Theory of Personal Conduct. Harper & Row, New York.
Temkin O (1963) The scientific approach to disease: specific entity and individual sickness. In: Crombie AC (ed) Scientific change: historical studies in the intellectual, social and technical

conditions for scientific discovery and technical invention from antiquity to the present. Basic Books, New York, pp 629–647
Twaddle AR (1993) Disease, illness and sickness revisited. In: Twaddle AR, Nordenfelt L (eds) Disease, illness and sickness: three central concepts in the theory of health, vol 18, Studies on Health and Society. Linköping University, Linköping, pp 3–18
Whitbeck C (1981) A theory of health. In: Caplan AL, Engelhardt HT Jr, McCartney JJ (eds) Concepts of health and disease: interdisciplinary perspectives. Addison-Wesley, Reading
WHO (1948) Official records of the World Health Organization, vol 2. WHO, Geneva, p 100
WHO (1986) The Ottawa Charter for health promotion, first international conference on health promotion, Ottawa, 21 Nov
WHO (1994) Report on the international conference on population and development in Cairo 5–13 Sept 1994. A/CONF.171/13/Rev.1
Young A (1982) The anthropologies of illness and sickness. Annu Rev Anthropol 11:257–285

Disease as Scientific and as Value-Laden Concept

4

Elselijn Kingma

Contents

Abstract

Health and disease are central concepts in medical practice. Defining them may assist in determining the scope of medicine; legitimizing medicine and psychiatry; and determining or even justly distributing medical care. This chapter reviews the philosophical literature on health and disease. It discusses naturalism (the view that disease is a value-free concept), normativism (the view that the concept

E. Kingma (✉)
University of Southampton, Southampton, UK
e-mail: e.m.kingma@soton.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_75

of disease is, essentially, value laden), and reasons why the polarizing opposition between naturalism and normativism may have to be rejected or, at least, nuanced.

Introduction

Is obesity a disease? How would one answer such a question? Obesity may shorten one's life-span: but so does poverty or working as a surgeon, and some chronic diseases do not affect life-span. Obesity is influenced by the environment as well as by genetic constitution, but so are infectious diseases and personality. Obesity causes or is associated with other illnesses, but so, once again, is poverty – whereas some diseases occur in isolation. Finally, obesity may be bad for one, but that is another feature it shares both with many diseases and many non-diseases, such as insecure employment and an unhappy marriage.

To answer the question of whether obesity is a disease, it is necessary to know what makes something a disease. And that is a question that has exercised philosophers of medicine for at least the last 40 years. Their main concern has been whether disease is a value-free or scientific concept – which suggests that scientific inquiry alone can answer the question of whether obesity is a disease – or whether it is a value-laden concept, in which case obesity's being a disease would depend primarily on how obesity is evaluated. This chapter will offer an introduction to and overview of these debates.

Section "Background" provides some background and history that will explain why this question was foregrounded, and why it continues to be an important one. Section "Conceptual Analysis" gives a brief sketch of what sort of project one engages in when one poses the question "what is a disease ?" Section "Naturalism" focuses on the first main position: naturalism, which claims that disease is a value-free, empirical, and/or scientific concept. Section "Normativism" considers normativism, the view that disease is, by definition, a value-laden concept. Section "Beyond Naturalism and Normativism" reviews some of the reasons given in the more recent literature for rejecting or moving beyond a simple binary opposition between naturalism and normativism.

Background

There are many reasons beyond mere intellectual curiosity for being concerned with what disease is. One historically important reason originates in the so-called "anti-psychiatry" movement in the 1960s. Anti-psychiatrists contended that psychiatry was neither a science nor concerned with legitimate diseases. Instead, anti-psychiatrists claimed that psychiatry medicalized ordinary problems in living and acted as a tool of social control (Szasz 1960). Examples of supposed misuse or even abuse of psychiatric diagnoses include considering masturbation a disease; hysteria

and drapetomania – i.e., the "disease" of slaves who ran away from their masters; and the practice of locking up of political dissidents in psychiatric institutions in the Soviet Union (Caplan et al. 1981).

Attempts to respond to the anti-psychiatrist challenge produced a substantial literature on the nature of disease. This literature was primarily concerned with the question of whether diseases, including mental disorders, were real, natural, or scientific entities – which would provide a defense against anti-psychiatric criticisms – or whether diseases were social or evaluative categories, a position whose relation to anti-psychiatry is more murky. On the one hand, such a position might be seen to legitimize anti-psychiatry, because it confirms that diseases are social categories. On the other hand – and this was undoubtedly meant to be the main thrust of the argumentation – if it were true that *all* disorders are social categories, then anti-psychiatrists would not have identified a problem that sets psychiatry apart from the rest of medicine; if this is a problem, it is a problem for medicine as a whole.

This debate became practically relevant when homosexuality became a point of contention in the construction of the third edition of the Diagnostic Statistical Manual (DSM), an influential list of mental disorders issues by the American Psychiatric Association (Bayer 1987). Homosexuality was listed as a mental disorder in the second edition of the DSM, but by the early 1970s, an active and increasingly powerful gay lobby campaigned to take it out. Psychiatrists were divided on the issue, and in the subsequent debates, the definition of disease took on a crucial importance. If diseases were natural entities – e.g., biological dysfunctions – it was believed that homosexuality would be a disease, because it was thought not to promote the biological function of reproduction. If diseases were conditions that were disvalued by patients, then homosexuality would not be a disease other than in those who disvalued it themselves.

The gay lobby won this debate, and an account of disease as condition that patients disvalue and are willing to present in a clinical context persists in the DSM to the present day. But the debate about this issue has not ceased, nor the underlying questions become irrelevant; quite the opposite. Psychiatry remains under particular pressure for medicalizing ordinary problems such as grief/depression (Horwitz and Wakefield 2007); enforcing social control, such as in cases of misbehaving children and diagnoses of ADHD (Hawthorne 2007); and creating spurious diagnoses, such as, perhaps, social anxiety disorder. Meanwhile disease concepts such as obesity; risk factors such as elevated blood pressure; and vague or contested diagnoses such as fibromyalgia show that these questions are not restricted to psychiatry: they are far from straightforward in somatic medicine. Examples such as these continue to raise questions about the scope of medical practice and about the nature of disease.

These questions have also become significant in two further contemporary contexts. First, since the 1970s, there has been an ever-increasing concern with containing the rising costs of healthcare and the distribution of healthcare resources. These debates often start by delineating the conditions that fall within or outside the scope of the healthcare system (e.g., Daniels 1985). Second, the last two decades have seen a significant interest in philosophy in debates about medical

enhancements. In these debates – when focusing on the contrast between treatment and enhancements – assumptions about what are and aren't diseases play an important role. In all these contexts, then, the question what disease is, and how to distinguish it from health, seems of crucial importance.

Conceptual Analysis

Before looking at dominant views on what disease is, it is important to consider what one is doing when asking the question "what is disease?" One possible answer is that one is trying to give an exhaustive account or exact description of all the ways in which people use the term "disease" and all the things that they are, or think they are, referring to with that term. But such a project is neither desirable nor philosophical. It is not philosophical because the best way to find this out would be to run an empirical piece of research mapping people's actual usage of the term "disease." It is not desirable because the very reason for posing the question is that there are certain controversial or borderline cases, where there is disagreement or uncertainty about the answer. Asking people what they think would, presumably, just replicate that controversy; it would not uncover a way of resolving it.

A second way of conceiving the project is to take it as an attempt to discover what the real-world category is that our term "disease" picks out. Thus, for example, in chemistry it is discovered that water is H_2O or that gold is really Au. Or, to give a disease example, it is discovered that the constellation of symptoms known as "consumption" is really a disease entity identified by a particular cause: pulmonary tuberculosis.

Such a project, however, makes substantial presuppositions. It presupposes that whatever category is salient to us must also exist as a category in the mind-independent world. It also presupposes that categories in the natural sciences take precedence over other ways of categorization. But both assumptions can be questioned. The category "young black men," for example, is both salient and very important in discussions of racial profiling; inequalities in incarceration; and other forms of social injustice. But it is not so clear that this is a category that exists "in the world" – or outside our social tendency to recognize it as such (Michael 2016). The second assumption can be questioned by considering that the biologically marginal distinction between onion and garlic could never reduce its great culinary significance (Dupré 1993).

This second way of conceiving the project, then, may presuppose the very question at hand: whether disease is a real, natural, or scientific concept.

A third way of conceiving the project lies between the two. In some ways it attempts to provide a more polished version of what the first project is after: how people employ the term disease. But this third way of conceiving the project assumes that in our – often messy and frequently inconsistent – use of the concept disease, we are latching onto particular underlying general ideas or features that unite our common usage. The philosophical task lies in uncovering these underlying themes,

criteria, or ideas that guide and determine actual usage. Although this project, like the first version, relies on actual usage, it is – unlike the first version – able to revise and refine it; by doing so one might uncover underlying themes that can then be consistently and usefully extended to resolve controversies or usefully disentangle a confusion. For example, by specifying and then distinguishing between murder and killing – or pornography and erotica – one can focus on whether there are cases of killing that are not bad, or what – if anything – makes pornography harmful.

This third version of the project is what is usually known as *conceptual analysis* – and this is the project that most of the relevant literature appears to be engaged in. That also explains why the literature on disease – almost without exception – is not concerned with what laypeople or even doctors call disease, but with a much broader set of conditions, including disability and trauma. The philosophical literature on disease is thus using "disease" as a term of art, covering any condition that is a departure from health. This chapter shall conform to that broad usage.

Thus far, the third version of the project has been distinguished from the first, but what distinguishes it from the second version? The difference between versions two and three is that, although the outcome of conceptual analysis might well be that the concept of disease picks out a natural category – in which case the third and second versions of the project generate the same result – the third version of the project does not presuppose such an answer; it leaves open that "disease" may pick out some other kind of category, if a unified category at all. And it is this that explains why the question "what is disease?" is a question for philosophy and not for science: because the question *whether it is* a question for science is still up for debate. And indeed science does not presently have a straightforward answer to the question of what disease is. Science can attempt to discover how obesity is caused, what consequences it has, how to affect obesity rates and/or an individual's body weight, and so on. But science cannot discover whether obesity *is* a disease, until the meaning of the term "disease" is pinned down.

And so it is to that question that the discussion now turns. From the introduction above, it will be recalled that there are two main approaches to disease: naturalism, which maintains that disease is an objective, empirical, and value-free concept that scientifically picks out a real-world concept, and normativism, which maintains that disease is primarily value laden.

Naturalism

Naturalists (e.g., Ananth 2008; Boorse 1977, 1997, 2014; Garson and Piccinini 2014; Hausman 2012; Kass 1975; Kendell 1975; Scadding 1988; Schramme 2007; Szasz 1960) contend that "disease" is an empirical, value-free concept: a scientific concept that picks out a natural, real-world category. This category is generally proposed to be "biological dysfunction." This section shall both provide an overview of the arguments generally mounted in favor of and against naturalism, and summarize specific naturalist positions.

Arguments for and Against Naturalism

A main attraction of naturalism is that it appears to provide some assistance in answering the difficulties that motivate our search for an account of disease. For if disease is a scientific concept or category, then we could at least partially outsource difficult social questions – such as who the medical system should treat; who is and isn't criminally liable; and who deserves our special consideration – to science. We cannot outsource them fully because even the most ardent naturalist realizes that no naturalist concept can determine social policy or moral requirements; we would always need to add values before a naturalist concept can usefully be applied to social and moral questions (Kingma 2012; note 4). A scientific definition would help, however, for example by restricting the scope of enquiry by providing a default for further moral analysis. Normativists, by contrast, cannot rely on science in this way and hence have a difficult job justifying why one kind of evil – say, murderous intent – should be treated so differently from what on their view is just a different kind of evil: mental disorder.

Naturalism is also seen as attractive because it would legitimize medicine and psychiatry as engaging with real entities – diseases – and allow one to say that certain now-discredited historical diagnoses – such a hysteria or drapetomania – were factually wrong.

Finally, naturalism seems to map onto our pre-theoretical assumptions about disease. There is often something informative and reassuring about being told that one's suffering arises from a disease. This is based on the idea that the disease is some specific and real thing – often out of one's control – and not merely one of many other bad things that could cause or constitute suffering, or vices that one is guilty of. There is also a lot of prima facie plausibility to the idea that health is "bodies and minds functioning as they should" and that disease is "bodies and minds going wrong," i.e., biological dysfunction.

But despite those initial attractions of naturalism, the position faces some powerful objections. First, although, prima facie, it seems plausible that disorders are biological dysfunctions, something else that diseases seem to share is that they are generally *bad* and near-universally disliked. Accordingly, a frequent objection to naturalism is that it fails to capture the shared features of diseases that make them, and have historically made them, salient to us as a category: these are conditions that are bad for us and that are grouped together for that reason (e.g., Agich 1983; Cooper 2002; Engelhardt 1976; Goosens 1980; Martin 1985; Margolis 1976; Nordenfelt 1987, 1993). Perhaps even more powerfully, critics contend that naturalists strip diseases of the precise feature that make us care about what diseases are. The reasons for the interest in defining disease are social and moral: they are questions about *how* to treat people justly and fairly. If diseases have no moral value as a category – if they are not conditions that are bad or disliked – then this category lacks relevance for the role apportioned to it in these morally charged debates (e.g., Lewens 2015, Chap. 11). A main criticism of naturalism, then, is that naturalist accounts of disorder have started to float free of both why diseases matter and why it is important to determine what it is.

Naturalists tend to respond to this by, in effect, adding a moral condition: only *bad* or *harmful* dysfunctions are relevant to social debates. But one might think that

does not avoid the objection; if the main emphasis is on biological dysfunction and "harmful" is added as an afterthought, then this still fails to *explain* why diseases are or should be so salient as a category.

In addition to these principled objections, many arguments against naturalism focus on the details of proposed naturalist accounts. They contend that these accounts fail to be value-free or fail to generate the right result for particular examples. We will look at some of these arguments when the different accounts are discussed in the following paragraph. Note that the force of the latter type of argument is always difficult to establish in a project of the present kind; the defender of the account always has the option of arguing that such a mismatch is precisely one that calls for revision of our opinion and usage of a term, in light of the consequences of their conceptual analysis.

Two Naturalist Accounts of Disease

Naturalists tend to define disease as biological dysfunction (see e.g., Ananth 2008; Boorse 1977, 1997; Garson and Piccinini 2014; Kass 1975; Szasz 1960). But what is biological dysfunction? A broken leg appears not to perform its biological function, but how obesity should be understood in such terms is not immediately clear. The question of what biological function – and function more generally – is has generated a substantial literature in philosophy of biology (e.g., Ariew et al. 2002). Two main accounts of biological function are featured prominently – which, applied to the present debate, result in two main naturalist accounts of disease.

The first account of function is a so-called *backward-looking* or *etiological* account of function (Ariew et al. 2002, Chap. 4). This defines function by reference to the evolutionary history of a trait: the function of a trait is the effect for which a trait is selected. So, for example, the function of the human pelvis is (among other things) to support upright walking, because it is that which explains the evolution of the pelvis into its current human shape.

The second account of function is forward-looking: it defines the biological function of a trait as its contribution to survival and reproduction (Ariew et al. 2002, Chaps. 3, 6). Thus the human pelvis (among other things) has the function of supporting upright walking, because that is how it contributes to our survival and reproduction: walking upright frees up our hands which allows us to do many things that are important for survival and reproduction, such as tool making and cooking.

The difference between these two accounts may not be immediately obvious; surely the effects of a trait that explain why it was naturally selected *are* the contributions to survival and reproduction of that trait? But, although the differences between the accounts are indeed subtle, they do exist. Consider, for example, turtle's flippers. They were selected for swimming, but turtles also use them to dig nests and bury their eggs. According to the etiological account, only swimming is the function of turtle's flippers, because only swimming – not burying – is what these flippers were selected for. According to the forward-looking account, both digging and

burying are functions of turtle's flippers, because both make contributions to survival and reproduction.

This example immediately leads to one main criticism of the etiological account of dysfunction and disease: surely medicine is interested in what traits do here and now. In other words, when there is a discrepancy between a trait's etiological function – what it was selected for in, possibly, quite a distant ancestral past and/or different ancestral environment – and the contributions a trait makes to survival and reproduction in present circumstances, then surely medicine would be interested in the latter, and specifically in failures of the latter, even if they aren't failures of etiological function? More simply, surely turtle doctors would care about flippers that – for whatever reason – don't bury, but still swim (Kingma 2013; Murphy and Woolfolk 2000).

A second objection to etiological accounts is that there are many human traits that weren't selected for in the way required for them to gain etiological function. These could therefore never dysfunction or be diseased. An example is the ability to learn to read – which cannot be selected for but must arise from our exercising traits selected for other functions. Dyslexia therefore might not be a disease (Kingma 2013; but see Griffiths and Matthewson forthcoming, for a response).

Another main criticism of the etiological account of function and disease is that it is not of much help: the relevant evolutionary effects could never be accessed. Instead the account is epistemologically circular: it infers from our current judgment about what is disease to what the putative evolutionary underpinnings must be (Murphy and Woolfolk 2000).

Most of the focus in the literature on health and disease however – in contrast to the philosophy of biology, where the etiological account is more prominent – has not been on an etiological account of disease, as most prominently developed by Wakefield (1992; see also Wakefield et al. 1999), but on the forward-looking account. And the remainder of the discussion of naturalism shall focus on this.

Boorse's Biostatistical Account of Disease

Christopher Boorse (1977) developed the forward-looking account of biological function into what is the most influential and most widely discussed account of disease. It defines health as normal function and normal function as the statistically typical contribution to survival and reproduction in a reference class. Reference classes are age and sex and perhaps some other polymorphisms (such as white skin for the purpose of vitamin D absorption). Disease, finally, is a departure from normal function.

Consider an example. Pelvic health is normal function. The normal functions of the pelvis are all the statistically typical contributions to survival and reproduction that pelvises make in a reference class. That includes contributing to walking upright in adults, but not in newborns. Similarly, in the reference class of women, it is a statistically typical contribution to survival and reproduction of the pelvis to bear babies, but not in men. A pelvis that does not perform these functions, for example,

because it is broken and can't support walking, or – in women – because it is deformed and won't let through babies, is diseased.

Boorse's account has received the most attention of any account of disease and has been subjected to a staggering array of criticism, not all of which can be discussed here – instead see Boorse (1997, 2014). Broadly speaking, criticisms of Boorse's account fall into four kinds: first, technical objections; second, complaints that the account is value laden; third, charges that Boorse does not employ good biology; and fourth, a long list of specific counterexamples. Since most recent discussion has focused on the first two categories, they will be focused upon. The third objection will not be discussed because discussions about the correct account of biological function are best left to philosophers of biology. The fourth set of objections is very specific, and – as noted earlier – their force in the present kind of analytic project is always a matter of dispute. In any case, some of the most interesting counterexamples arise in the context of the first two criticisms.

One of the longest-running objections to Boorse's account is that it is value laden. Specifically, it has been argued that the BST is value laden because of its choice of goals: survival and reproduction (Agich 1983; Brown 1985; Engelhardt 1976). More recently it has also been argued that the BST is value laden through its choice of reference class, for which no biological justification can be given (Kingma 2007). For why is being a male white baby a reference class, but not a male baby with Down syndrome? On pain of circularity, the answer can't be that the male white baby is healthy. But without another answer available, this argument contends, it sounds like an awful lot of work is being done by an arbitrary and possibly value-laden choice of reference classes. Moreover, the argument claims, when a controversy is couched in terms of what reference classes should be adopted – e.g., should there be a different reference class for people who are gay or deaf? – then the BST does not seem able to provide a nonarbitrary answer.

Boorse responds that such objections, either about goals or about reference classes, do not make the concept of disease value laden. Although he grants that medicine could have chosen to engage with something other than disease, it is engaged with disease. And disease, he claims, is explicated by the BST. "[E]ven when a concept is precisified one way rather than another for evaluative reasons, the result can still be a value-free concept: cf. 'meter', 'degree C', or virtually any unit of measure. To think otherwise is the genetic fallacy" (Boorse 1997: 28). His critics are not persuaded: no one actually believes that there is anything other than a historically arbitrary choice behind our concept of a meter; it was created, not discovered, by science and does not – so to speak – "carve nature" at any particular joint. But surely the naturalist suggestion and appeal is that what disease is *is* a real or respectable category out there in the world – and *not* the result of a historically arbitrary cultural choice.

Other main criticisms of Boorse's account are more technical and focus on the details of his account. One such question concerns *where* Boorse locates the distinction between health and disease. Statistically typical function invariably indicates a range of variation, with fuzzy boundaries. So when is a given contribution to survival and reproduction *enough* of an adverse departure from normal to be a disease? Boorse argues that such a boundary will just have to be placed at a particular population

frequency, the bottom 5 % of function, for example. But no such simple method works (Schwartz 2007): a far larger proportion of the population of 70-year-old men than of 16- year-old women will have what we think of as a heart disease. Schwartz proposes a solution to this problem, which appeals to the consequences of such a departure: when these consequences become significantly negative, the departure becomes a disease. Whether this solution is value-free is up for debate (Kingma 2014).

Another technical criticism of Boorse's account concerns whether the BST can accommodate environmental variation and in particular harmful environments in which diseases are the statistically normal result (Kingma 2010; Nordenfelt 1993). The BST will want to designate the movement or "give" in a woman's pelvis during labor functional, because it contributes to her survival and reproduction. But being in labor is not statistically typical for a woman: it may be a biologically normal – even essential – state, but one that makes up a tiny proportion of her lifetime. To capture such normal functions occurring in infrequent states or circumstances, the BST must define normal functions as what is statistically typical *given* a specific environment or situation. What, then, should the BST say about the organism's statistically typical contribution to survival and reproduction *given* that it is poisoned or infected with cholera? The BST cannot account for both situations (Kingma 2010). This argument has sparked a lively debate (Garson and Piccinini 2014; Hausman 2011; Kingma 2015).

A final persistent sticking point regarding the BST is the example of so-called ubiquitous diseases (Boorse 1997; Guerrero 2010). The BST defines health as that which is statistically normal. So – the objection goes – what would happen if a nuclear attack made the entire world blind, for the rest of human future? According to the BST, after an initial transition period (because statistically normal function needs to be determined with respect to a "reasonable time slice of the species,") blindness would be statistically normal and hence not a disease. That, it is argued, is unacceptable. But this argument is a fine example of the kind of argument whose weight is difficult to determine in a debate about conceptual analysis. Boorse and others argue that the BST gets this right: biology changes, and were all humans to become blind, this would be the new norm – just as the lack of a tail and the disappearance of skin pigmentation in populations of European heritage is normal (Boorse 1997).

Normativism

The main rival of naturalism is normativism (e.g., Agich 1983; Cooper 2002; Clouser et al. 1981; Engelhardt 1976; Goosens 1980; Martin 1985; Margolis 1976; Nordenfelt 1987, 1993; Reznek 1987; Whitbeck 1978) Normativism is the view that health and disease are primarily, or essentially, evaluative concepts. Cooper writes: "[b]y 'disease' we aim to pick out a variety of conditions that through being painful, disfiguring or disabling are of interest to us as people. No biological account of disease can be provided because this class of conditions is by its nature anthropocentric and corresponds to no natural class of conditions in the world" (Cooper 2002: 271).

In many ways, normativism is less easily characterized than naturalism because it is such a heterogeneous position (Simons 2007). Whereas naturalists are pretty united on how disease is to be analyzed naturalistically – as biological dysfunction, which admits of two main interpretations – normativists merely share the view that health and disease are, in some sense at least, primarily value-laden concepts. But there are many possible accounts of health and disease as value-laden notions and many substantive conceptions on what values are. Moreover, normativists are not always clear what interpretation of the claim "disease is value-laden" they have in mind.

It is not possible to discuss all actual or possible varieties of normativism here. Instead we shall first consider the general arguments in favor and against the view that health and disease are value laden. A brief overview of different specific normativist accounts will then be provided, which will give a clear flavor of the variety on offer. In comparison to the section on naturalism, the discussion of arguments supporting these normativist accounts will only be brief. This is because individual normativist accounts have not been subjected to as much critical analysis in the literature than their naturalist counterparts.

Arguments for and Against Normativism

Arguments for and against normativism mirror those against and in favour of naturalism. A main attraction of normativism over naturalism is that it does one main thing that naturalism failed to do: it promises to explain *why* we care about disease, and perhaps even why we care about disease in such a specific way that, among many misfortunes, they are of special concern to social justice. This argument has considerable prima facie plausibility: diseases on the whole are conditions that are of interest because they are bad for us. Naturalists may counter this argument by claiming that, although value-laden concerns may well motivate the interest in diseases – just as beauty and resistance to corrosion explain the interest in gold – the concept of disease, like the concept of gold, nonetheless picks out a natural category (Kingma 2012). But as mentioned, this reply and others fail to explain why the connection between disease and social concern remains so salient.

The main arguments against normativism are, first, that normativists make a conditions status as a disease wholly dependent on what societies think. Thus, or so the argument goes, normativists could never maintain that Victorians were wrong to think that masturbation and hysteria are diseases; that certain locally desired and culturally promoted deformities – such as bound feet in China – are a disease/disability; or that a society could be wrong about whether homosexuality is a disease.

Normativists can respond in one of two ways to this argument. First, they can stand their ground: when a practice is culturally desired, it is not a disease; after all piercings, tattoos, ritual scarring, or circumcised penises are not considered diseases. Second – and this is perhaps less widely noted – it could be argued that a commitment to value ladenness need not entail a commitment to relativism or conventional

accounts of moral value. Philosophy offers plenty resources to resist the entailment; for example, one could argue that whether something is right or wrong is still a matter of (moral) fact and not determined by "whatever any society happens to thinks about it."

A second main argument against normativism points toward animal and plant health: it seems that illness can meaningfully be attributed to a mouse or a tree. But it is not clear that a mouse – let alone a tree – can place a value on its condition. Nor is it clear that such judgments are driven by human negative evaluation of the condition; a gardener could gleefully report that the dandelions in the garden seem to have fallen ill – or an environmental activist that she has succeeded in releasing a pest onto some hated GMO crop. The fact that these speakers value these illnesses does not stop them from labeling them illnesses. Thus one main argument against normativism and in favor of naturalism stresses that if disease can meaningfully be applied to the natural world in a value-free way, it can meaningfully be applied in that way to humans; there is no reason to think the term should suddenly change to being different and value laden.

Finally, normativist accounts are often criticized for failing to distinguish between diseases and other misfortunes. This is an important criticism because one of the reasons to be clear about disease is that the concept of disease is used to make relevant moral distinctions between different kinds of misfortunes: for example, in court, the difference between being mentally ill (and hence excused) and simply "evil," or not raised well – and hence culpable – matters a great deal (Cooper 2002). Think also of how differently people who are judged "genuinely" ill are treated compared to those who are merely judged lazy or suffering from being unhappily married.

Normativist Accounts of Disease

As mentioned, normativist accounts of disease are less easily characterized than naturalist accounts. First, because they are far more varied (Simons 2007 gives a good overview). Second, because many normativists haven't clearly pledged allegiance to one account or the other, defending instead a more general claim that disease is a value-laden concept. Third, normativists nearly always defend their accounts in contrast to Boorse's, rather than in contrast to each other, which means that there is comparatively little work done on their comparative advantages. Therefore the three most important ones will be discussed here – and their comparative merits will be left up to the reader to decide. Detailed normativist accounts that are not reviewed here in detail and that have received even less attention in the literature are offered by Clouser, Culver and Gert, and Whitbeck. Clouser et al. (1981) define "disease" as a condition, other than a rational belief or desire, that incurs or significantly increases the risk of incurring a harm or evil. Whitbeck (1978) defines diseases as conditions that people wish to be able to prevent because they interfere with the bearer's capacity to do things people commonly wish and expect to be able to do.

Cooper and Reznek on Disease

Rachel Cooper (2002) offers a normativist account of disease in which the idea that diseases are bad or harmful conditions takes centre stage. According to Cooper's account, a condition is a disease if the following three jointly necessary and sufficient conditions are met:

1. The condition is bad for the sufferer.
2. The condition is abnormal or unlucky.
3. The condition is potentially medically treatable.

In interpretation of the first condition, Cooper explicitly states that the condition has to be bad for the sufferer herself, not merely bad for society. To use her example, pedophilia is not a disease, regardless of its effects on others, if it is not bad for the sufferer herself. Cooper defers the further task of cashing out what values are and of what it is to be "bad for the sufferer" to the literature in ethics and metaethics (which of course provides plenty of sources for saying both that pedophilia is and is not bad for the sufferer).

Cooper's second criterion is meant to delineate unwelcome, potentially medical treatable conditions such as facial hair growth or menstruation or feeling tired from actual diseases. The former, it is claimed, are normal; one is not unlucky to suffer them.

The third condition functions to distinguish diseases from other kinds of misfortunes. Cooper explicates "potentially medically treatable" as "deemed by society as the kind of thing that medicine treats," whereby she proposes a sociological account of medical treatment. Thus, say, a psychosis is a mental disorder, but mere evil intent and being in an unhappy marriage are not diseases according to Cooper's account. Although all these conditions may be unlucky and bad for the sufferer, the latter two are not diseases because they are not conditions for which we deem medical treatment appropriate. This interpretation of the third condition distinguishes Cooper's account from an earlier, almost identical account by Reznek (1987), who lists the same three conditions, but interprets "medical treatment" to consist of particular kinds of interventions (e.g., administering drugs).

An example may help us get an overall grasp of Cooper's account. A broken leg is a disease according to Cooper because, first, it is bad for the sufferer; second, it is unlucky/abnormal (most humans don't have broken legs); and, third, it is the kind of thing that medicine should treat. But having a harmless mole isn't a disease because it isn't bad for the sufferer, menstruation isn't a disease because it isn't abnormal, and being poor isn't a disease because it is not the thing that medicine should treat.

One interesting outcome of Cooper's account is that unwanted pregnancy can be a disease on her account. For, first – on a plausible construction of badness – a pregnancy is bad for a woman who has a good reason not to want become pregnant; second, becoming pregnant when trying to avoid it is unlucky/abnormal, and, third, it has become appropriate for medicine to treat unwanted pregnancy in many

countries, at least in its early stages. Cooper argues that this outcome of her account is correct, and that, given time – as widespread reliable contraception and safe abortion are a relatively recent occurrence – we will come to regard it as such. But many may feel that this can't be: surely a healthy pregnancy is a core example of a body's successful execution of some of its most basic normal functions? How could it possibly be a disease? The naturalist will be convinced by this reaction. The normativist, however, is likely to regard it as simply, and wrongly, presupposing that disease is biological dysfunction.

Nordenfelt on Second-Order Inabilities

A second, but much more revisionary, normativist account of disease is offered by Lennart Nordenfelt. Nordenfelt (1987) defines disease as a second-order inability to reach one's vital goals. One's vital goals are the goals that are jointly necessary and sufficient for minimal happiness. A second-order inability is the inability to gain an ability. Thus, for example, one may not have the ability to speak Spanish but possess the second-order ability to *learn* to speak Spanish. How does this account work out in practice? When, for example, one has a broken leg, one lacks the ability to do very many things that are important to achieve minimal happiness, such as walking around. Therefore having a broken leg is a departure from health, on Nordenfelt's account.

One main objection to Nordenfelt's account concerns people with very specific or ambitious goals: suppose that one is in such a deep personal crisis that it becomes necessary for one's minimal happiness to undergo plastic surgery that will make one look like Kim Kardashian. Does that mean the person's failure to look like Kim Kardashian is a disease? Most would think that it is not. Nordenfelt's response partially concurs. He claims that a person probably would be judged to be ill if looking like Kim Kardashian was one of their vital goals. But Nordenfelt suggests that such a person could be helped by getting them to adjust their goals – which would get rid of the illness – rather than providing them with plastic surgery.

Neo-Aristotelian Accounts of Disease

The third normativist account is more of a family of accounts: neo-Aristotelian accounts (Foot 2001; Megone 1998). These accounts go back to an ancient tradition of conceptualizing philosophy as "medicine for the soul," and these accounts consider both mental and physical health as primarily being about the subject's flourishing.

These accounts have been located where the naturalist no doubt would put them: with the normativists. This is because they posit that an understanding of health requires an understanding of what is a good life for the subject. But it is not clear that neo-Aristotelians themselves would classify their position as normativist. The relationship between neo-Aristotelian accounts, naturalism, and normativism is

complicated because neo-Aristotelians are committed to a very specific Aristotelian account of well-being or flourishing. This is itself supposed to be rooted in what one might think of as a natural norm: in the kind of thing that something is. Thus, to use an example, trees grow and catch the light. This means that for the Aristotelian a "good" tree, a "happy" tree, and a "healthy tree" – all of which on this account effectively amount to the same thing, as indeed is a "well-functioning tree" –are trees that grow large and catch lots of light. Aristotelian accounts, then, deny a distinction between biological functions and values; to them, these are all the same notion. To be a good entity of some sort is to do the kind of things that that sort does.

Neo-Aristotelian accounts of disease face their own particular criticisms. One is that they cannot distinguish between diseases and vices (Cooper 2007). And, indeed, to the Aristotelian, both of these are departures from a healthy and/or virtuous soul and body. But from a contemporary perspective, which is keen to maintain a distinction between mental disorder and vice – think of criminal liability – this may be an insurmountable problem. Another criticism of Aristotelian accounts is that they follow an outdated and indeed false picture of biology, which cannot be upheld in the face of modern understanding (Lewens 2015, Chap. 10).

Beyond Naturalism and Normativism

The brief discussion of Neo-Aristotelian accounts and the way in which neo-Aristotelians might reject a strict opposition between naturalism and normativism – or fact and value – is an excellent introduction to this final section. Most of this chapter has been focused on naturalist accounts, normativist accounts, and the opposition between them. And indeed, those are the familiar terms in which the debate on health and disease is set up. But there are several reasons for thinking that the opposition is both too stark and simplistic – and indeed, possibly, unhelpful. The chapter shall finish with mentioning three of these reasons.

First, as the example of Aristotelian accounts already indicated, there are accounts that do not subscribe to the naturalist-normativist opposition. As discussed before, Aristotelians think that biological functions are normative and that values – or what constitutes well-being – are facts grounded in natural norms. Indeed if attention is turned to the literature on metaethics, which is the literature on what values *are*, then there are more positions that propose to give a *naturalized* account of values – where it is the values themselves that can be naturalized. On the other hand, there are people who argue that what may appear as a fact is in fact a value – but that only in areas where there is disagreement about the values, do the values light up as values; values which are agreed are disguised as facts (Fulford 1989). Metaethics is not, however, an area that the literature on health has paid special attention to.

Second, the opposition between naturalism and normativism is too simplistic. One very recent, metaethically informed proposal, for example, argues that normativists and naturalists do not differ on one but on two dimensions (Broadbent, manuscript under review). One dimension is value/value-free, but the other dimension is *subjective/objective.* At present, normativism tends to be characterized as

"value-laden + subjective" and naturalism as "value-free + objective." If the two dimensions are placed in a matrix, however, there are two other positions that can be held. First one could hold that a concept is value laden, but nonetheless objective. Such a position may correspond to views in metaethics which hold that moral facts although not reducible to nonmoral facts are facts nonetheless: whether murder is right or wrong, on this view, is not a matter of cultural convention, attitude, or opinion, nor is it reducible to a nonmoral fact: it is a matter of (moral) fact. Such a view is not implausible: indeed most of us are (or once were) of the view that there is a moral truth about the wrongness of murder. Translated to health, it would be possible to hold a view where health is value laden, but the values are objective, not subjective. Given that one of the main criticisms of normativism is that it would lead to some form of cultural relativism, it is surprising that this is a position that is not explored more. Second, it would be possible to have a position where health and disease are value-free concepts, but nonetheless subjective. Like causation or color, they are something that *we recognize* – and do so as being out there in the world – but that are nonetheless only brought into existence through perceptual engagement with the world. This is the position Broadbent favors.

Third, the opposition is too stark. There are at least in principle several possible ways in which naturalism and normativism could be combined, beyond defining disease as harmful dysfunction. Kingma (2012, 2014) proposes one way of combining the insights and benefits of both normativism and naturalism. She argues that naturalists are correct to think that disease is dysfunction, but that values play a role in precisifying and specifying the concept of dysfunction.

To conclude, while the idea that disease is either a value-laden or a scientific concept has had an important role in shaping the debate and remains, to some extent, a useful starting place to think about the question of what disease is, it may not be the best way to continue that analysis. The way forward is almost certainly not to polarize further by emphasizing the contrast between naturalism and normativism, but to adopt a more nuanced perspective either by crafting a position in the middle, or by recognizing multiple dimensions of opposition, or by finding other ways of combining the insights from both camps, or, indeed, by doing all of the above.

Definitions of Key Terms

Etiological account of function	An account of biological function that appeals to the evolutionary history of a trait and usually defines the function of a trait as the effect for which that trait was selected.
Anti-psychiatry	A social and intellectual movement in the 1960s that criticized psychiatry. It claimed that psychiatry lacked legitimacy as a medical or scientific discipline; that it was engaged in mere social classification, misidentifying social problems as disorders; and, at worst, that it was best characterized as a tool of social control in support of existing power structures.

BST	Biostatistical Theory: an account of health and disease developed by Christopher Boorse that defines health as normal function, disease as an adverse departure from normal function, and the normal function of a trait as the statistically typical contribution by that trait to survival and reproduction in a reference class.
Conceptual analysis	The philosophical project of uncovering underlying criteria and (in)consistencies in our concepts, with an eye to providing clear definitions or criteria for application, disentangling or even resolving controversies, and/or isolating specific philosophical questions.
DSM	The Diagnostic Statistical Manual: an authoritative list of mental disorders and their symptoms updated and published roughly every two decades by the American Psychological Association.
Enhancement	The improvement of human (or animal/plant) characteristics beyond what is considered "normal" or healthy (usually contrasted with "treatment").
Forward-looking account of function	An account of function that looks at the disposition of a trait to contribute to survival and reproduction.
Metaethics	The branch of philosophy that considers the nature of ethical concepts, attitudes, judgments, and properties.
Naturalism	The view that disease is an empirical, scientific, and/or value-free concept.
Normativism	The view that disease is, essentially, a value-laden concept.
Reference class	Technical term in Boorse's BST: an age group of a sex (of a race) of a species.

Summary Points

- Understanding the concept "disease" is important for determining the scope of medical practice, for determining or distributing entitlements to medical care, and for understanding the distinction between treatment and enhancement.
- The question "what is disease" is philosophical question, often answered by conceptual analysis.
- Two positions on disease are usually contrasted: naturalism – which defines disease in value-free terms, as biological dysfunction – and normativism, which provides a value-laden analysis of disease.
- Naturalism promises objective assistance to answering social and moral questions, but fails to explain why diseases should be especially relevant to these domains.

- Normativists preserve a close connection between diseases and social and moral domains, but struggle to explain how we apply the concept disease to nonhuman organisms and face charges of historical and cultural relativism.
- Naturalists usually define diseases as dysfunctions, but their proposals are subject to many technical objections.
- Many normativist accounts can be criticized for not adequately distinguishing between diseases and other disvaluable conditions.
- Progress in this debate must come from moving beyond a simple naturalist-normativist opposition, through combining the two positions and/or disentangling the multiple contrasts between them that are presently conflated.

References

Agich GJ (1983) Disease and value: a rejection of the value-neutrality thesis. Theor Med 4:27–41
Ananth M (2008) In defense of an evolutionary concept of health: nature, norms and human biology. Ashgate, Aldershot
Ariew A, Cummins R, Perlman M (2002) Functions. Oxford University Press, Oxford
Bayer R (1987) Homosexuality and American psychiatry: the politics of diagnosis. Princeton University Press, Princeton
Boorse C (1977) Health as a theoretical concept. Philos Sci 44:542–573
Boorse C (1997) A rebuttal on health. In: Humber JM, Almeder RF (eds) What is disease? Humana Press, Totowa
Boorse C (2014) Replies to my critics. J Med Philos 39:683–724
Broadbent A (under review) Health as a secondary property
Brown WM (1985) On defining "disease". J Med Philos 10:311–328
Caplan AL, Engelhardt HT Jr, McCartney JJ (1981) Concepts of health and disease: interdisciplinary perspectives. Addison-Wesley Publishing, Reading
Clouser KD, Culver CM, Gert B (1981) Malady: a new treatment of disease. Hastings Cent Rep 11:29–37
Cooper R (2002) Disease. Stud Hist Phil Biol Biomed Sci 33:263–282
Cooper R (2007) Aristotelian accounts of disease – what are they good for? Philos Pap 36:427–442
Daniels N (1985) Just health care. Cambridge University Press, New York
Dupré J (1993) The disorder of things: metaphysical foundations of the disunity of science. Harvard University Press, Cambridge
Engelhardt HT (1976) Ideology and etiology. J Med Philos 1:256–268
Foot P (2001) Natural goodness. Clarendon, Oxford
Fulford KWM (1989) Moral theory and medical practice. Cambridge University Press, Cambridge
Garson J, Piccinini G (2014) Functions must be performed at appropriate rates in appropriate situations. Br J Philos Sci 65:1–20
Goosens W (1980) Values, health and medicine. Philos Sci 47:100–115
Griffiths PE, Matthewson J (forthcoming) Evolution, dysfunction and disease: a reappraisal. Br J Philos Sci. Accepted 21 Sept 2015
Guerrero JD (2010) On a naturalist theory of health: a critique. Studies in the History and Philosophy of Biological and Biomedical Sciences 41:272–278
Hausman D (2011) Is an overdose of paracetamol bad for one's health? Br J Philos Sci 62:657–668
Hausman D (2012) Health, naturalism and functional efficiency. Philos Sci 79:519–541
Hawthorne S (2007) ADHD drugs: values that drive the debates and decisions. Med Health Care Philos 10:129–140
Horwitz AV, Wakefield JC (2007) The loss of sadness: how psychiatry transformed normal sorrow into depressive disorder. Oxford University Press, Oxford

Kass LR (1975) Regarding the end of medicine and the pursuit of health. Public Interest 40:11–42
Kendell R (1975) The concept of disease and its implications for psychiatry. Br J Psychiatry 127:305–315
Kingma E (2007) What is it to be healthy? Analysis 67:128–133
Kingma E (2010) Paracetamol, poison and polio: why Boorse's account of function fails to distinguish health and disease. Br J Philos Sci 61:241–264
Kingma E (2012) Health and disease: social constructivism as a combination of naturalism and normativism. In: Carel H, Cooper R (eds) Health, illness and disease: philosophical essays. Acumen Publishing, Durham
Kingma E (2013) Naturalist accounts of disorder. In: Fulford KWM, Davies M, Graham G, Sadler J, Stanghellini G, Thornton T (eds) Oxford handbook of philosophy and psychiatry. Oxford University Press, Oxford
Kingma E (2014) Naturalism for health and disease: adding nuance for progress. J Med Philos 39:590–608
Kingma E (2015) Situational disease and dispositional function. Br J Philos Sci. Advanced access doi:10.1093/bjps/axu041
Lewens T (2015) The biological foundation of bioethics. Oxford University Press, Oxford
Margolis J (1976) The concept of disease. J Med Philos 1:238–255
Martin M (1985) Malady and menopause. J Med Philos 10:329–337
Megone C (1998) Aristotle's function argument and the concept of mental illness. Philos Psychiatry Psychol 5:187–201
Michael J (2016) Race. The Stanford encyclopedia of philosophy (Spring 2016 Edition), Zalta EN (ed) URL http://plato.stanford.edu/archives/spr2016/entries/race/
Murphy D, Woolfolk RL (2000) Conceptual analysis versus scientific understanding: an assessment of Wakefield's folk psychiatry. Philos Psychiatry Psychol 7:271–293
Nordenfelt L (1987) On the nature of health: an action-theoretic approach. Reidel, Dordrecht
Nordenfelt L (1993) Quality of life, health and happiness. Ashgate Publishing, Aldershot
Reznek L (1987) The nature of disease. Routledge, London
Scadding JG (1988) Health and disease: what can medicine do for philosophy? J Med Ethics 14:118–124
Schramme T (2007) A qualified defence of a naturalist theory of health. Med Health Care Philos 10:11–17
Schwartz P (2007) Defining dysfunction: natural selection, design, and drawing a line. Philos Sci 74:364–385
Simons J (2007) Beyond naturalism and normativism: reconceiving the "disease" debate. Philos Pap 36:343–370
Szasz TS (1960) The myth of mental illness. Am Psychol 15:113–118
Wakefield JC (1992) Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder. Psychol Rev 99:232–247
Wakefield JC et al (1999) Special section: the concept of disorder: evolutionary analysis and critique. J Abnorm Psychol 108:371–472
Whitbeck C (1978) Four basic concepts of medical science. PSA Proc Biennial Meet Philos Sci Assoc 1:210–222

Mental Disorders as Genuine Medical Conditions

5

Jerome C. Wakefield

Contents

Abstract

Psychiatry and the other mental health professions treat problematic psychological conditions that are claimed to be "mental disorders" that qualify as genuine medical disorders. Yet the question of whether psychological conditions such as feelings, thoughts, and actions can be medical disorders remains a matter of intense controversy. To resolve whether psychological conditions can be genuine medical disorders requires an analysis of the meaning of "medical disorder." Several standard analyses, such as that medical disorders always involve physical lesions, or medical disorders are simply undesirable bodily and mental conditions, do not explain our nuanced judgments about disorder versus nondisorder. The common argument that all mental disorders occur in the brain, therefore all mental

J.C. Wakefield (✉)
Silver School of Social Work and Department of Psychiatry, School of Medicine, New York University, New York, NY, USA
e-mail: wakefield@nyu.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_1

disorders are brain diseases is not valid and does not explain how we recognize psychological conditions as disorders in the absence of knowledge of the hypothesized brain lesions. The best analysis appears to be the "harmful dysfunction" analysis that combines a value criterion with a scientific criterion: a medical disorder is the harmful failure of some internal mechanism to perform one of its biologically designed functions. Because psychological functions, like physical functions, have been biologically designed by natural selection, this account explains how mental disorders can be genuine medical disorders in exactly the same sense of "disorder" as other medical disorders, namely, they are both harmful failures of biologically designed functioning. To the degree that current psychiatric diagnostic categories are plausible attempts to identify such harmful psychological dysfunctions, they are plausibly genuine medical disorders.

Introduction: Is Psychiatry Part of Medicine?

Psychiatry, clinical psychology, clinical social work, psychiatric nursing, and the other mental health professions (collectively referred to here as "psychiatry") claim to treat mental disorders, understood as psychological conditions that are disorders in the literal medical sense that they represent diminished health. These professions of course do many other things as well. For example, they help normal individuals to adapt to the stressful demands of our modern society when normal-range human reactions, such as anxiety over public speaking or difficulty adjusting to a night work schedule, do not allow individuals to satisfy those demands (Wakefield 2015b). However, their focus on mental disorders is the rationale for locating them within the "health" professions and providing them with the exceptional privileges and responsibilities associated with medicine, including reimbursement for treatment by medical insurance. The idea that there are genuine mental disorders in the medical sense is integrated into our laws and social practices in many other ways as well. For example, in forensic contexts, there is the "insanity plea" and other special provisions for the mentally ill, as well as potential involuntary institutionalization of the mentally disordered, and our social support systems provide disability benefits and school and work accommodations to those suffering from mental disorders. In ordinary life, too, we have come to rely on mental disorder as an explanation for actions that otherwise seem inexplicable, such as why someone commits a mass shooting or why a friend who seemed to have everything would commit suicide.

Nonetheless, psychiatry's status as a medical discipline that treats medical disorders of the mind has been vigorously challenged in controversies that have raged across half a dozen academic disciplines over the past half century. The fundamental assumption that psychological conditions consisting of thoughts, emotions, and actions can be medical disorders has been disputed by a variety of groups, ranging from scientologists and political libertarians to behavioral psychologists and some psychiatrists themselves (e.g., Szasz 1974). The questions raised are, first, whether the concept of mental disorder – that is, the concept of a psychological condition that is a medical disorder – makes sense at all and, second, whether some of the

psychological conditions currently recognized as mental disorders are in fact justifiably considered medical disorders. These questions and proposed answers will be explored in this article.

Among medical specialties, the conceptual challenge to the legitimacy of psychiatry as part of medicine is unique. The root cause lies in psychiatry's broad social implications and uniquely sensitive social position. Psychiatry has moved from asylum- and hospital-based inpatient treatment of severe conditions to community- and outpatient-based intervention with milder conditions where the difference between normal variation and disorder is harder to discern, raising complex conceptual and bioethical issues about the disorder/nondisorder distinction and potential medical control over what are in fact normal variations in human experience (Wakefield 2010a). After all, intense anxiety, intense sadness, lack of ability to read, distraction and lack of interest in schoolwork, and delinquent behavior are widespread in the community, and they can all be undesirable but normal parts of life, yet under some conditions they are diagnosed as anxiety disorders, depressive disorders, reading disorders, attention-deficit/hyperactivity disorders (ADHD), or conduct disorders, so what is the difference between the normal-range problems and the superficially similar mental disorders? Moreover, when treating mental disorders, psychiatry manipulates emotions, thoughts, and behaviors that are of the same general kind involved in all major areas of human endeavor and that in modern pluralistic democracies are considered to be inappropriate targets for overt social control, thus raising a question of human freedom (Wakefield 2010b). In fact psychiatry repeatedly has been used for oppressive social control purposes, from classifying runaway slaves as disordered in the antebellum American south and classifying Victorian women as sexually disordered if they experienced orgasms to classifying Soviet dissidents as psychotic. So, psychiatry's mislabeling of conditions as disorders for social control purposes is no mere theoretical problem.

Conceptual questions about the disorder/nondisorder distinction are especially of concern in today's psychiatric environment in which the mental health field has become a major industry and a conduit for the pharmaceutical industry's products. Moreover, the dominant view at present is that all mental disorders are brain disorders, which inclines many clinicians toward thinking that the most appropriate intervention is medication, with its potential side effects, rather than psychotherapy. Critiques of psychiatric diagnosis thus often focus on claims concerning the influence of pharmaceutical manufacturers on psychiatric practice and diagnosis. Certainly the use of psychotropic medication has become extraordinarily widespread among both children and adults. For example, recent surveys show that about a quarter of all women in their 40s and 50s in the USA are taking antidepressants (Pratt et al. 2011). Similarly, over the course of childhood about one in five US boys is diagnosed with ADHD at some point, and about three-quarters of those diagnosed receive medication (Visser et al. 2014).

However, pharmaceutical companies and prescribing physicians are merely exploiting what is warranted by psychiatric diagnostic criteria. Once diagnostic criteria for a mental disorder are accepted, individuals satisfying the criteria are presumed to suffer from an internal dysfunction, which biases intervention toward psychiatric

medication as an appropriate treatment. Whatever the benefits and costs of nondisordered individuals taking psychotropic medication, presumably misdiagnosing normal individuals as disordered thus biasing treatment decisions toward prescribing medication is unacceptable and, if done knowingly, perhps unethical. Consequently, the issue of whether various psychological conditions really are mental disorders isat the heart of many of these disputes about excessive prescribing.

Conceptual Analysis of "Medical Disorder" and Its Goals

To address the concerns raised by the critics of psychiatry, philosophers of medicine have tried to analyze the concept of medical disorder and to answer the question of which, if any, psychological conditions legitimately fall under that concept. This is essentially a conceptual question: What is our implicit concept of a medical disorder (i.e., what do we mean by the phrase "medical disorder"), and is it a coherent concept? Once that is answered, the next question is whether in principle psychological conditions can be genuine medical disorders. Third, following from that general conceptual analysis, there is the more applied and detailed question of whether the conditions currently considered mental disorders plausibly fall under the concept of medical disorder.

Although judgments about disorder and nondisorder obviously vary, the degree of professional and lay consensus about such classificatory judgments suggests the possibility of an analysis that identifies the concept of disorder that people tend to share. Ideally, such an account would not only illuminate the kinds of conditions that are clear cases of disorder and nondisorder but also explain why some conditions are borderline or fuzzy cases. Moreover, it could illuminate disagreements by indicating how people who have different specific beliefs or theories about a psychological condition may be led to opposite answers as to whether the concept of disorder applies.

Any analysis of "mental disorder" should at a minimum shed light on psychiatry's two central issues, legitimacy and scope. First, by clarifying how a mental disorder can be a disorder in the medical sense, the analysis should explain the legitimacy of psychiatry as a genuine medical field. Second, it should offer guidance regarding the limits of mental disorder as medical disorder and thus indicate why psychiatry cannot legitimately declare any socially deviant or disapproved feeling or behavior to be a mental disorder subject to potentially oppressive medical intervention and social control.

"Disorder" as a Generic Term for Medical Conditions That Diminish Health

In considering whether psychological conditions can be genuine medical disorders despite their differences from standard physical diseases, the notions of "disease" and "illness" may be too specific to cover some mental disorders, so we need some generic term to cover all possible types of medical deviations from health.

"Disorder" is used here as a default generic term for all deviations from health, including diseases, illnesses, injuries, traumas, and others. Because "disorder" is used in psychiatric diagnostic manuals, it is often mistakenly claimed that "disorder" was introduced recently as a fuzzy term for psychiatric illness or disease. However, medical conditions that involve decrements in health are of many types, and the usual terms like "disease" and "illness," although they can sometimes be used as synonyms for "disorder," have nuances that generally limit the kinds of medical conditions to which they refer. For example, a broken arm due to a trauma is not comfortably described as a disease or as an illness, yet it certainly is a medical condition involving a health problem. "Disorder" has long been a recognized term of art for referring to all of the diverse forms of deviations from health (e.g., diseases, traumatic injuries, congenital defects) diagnosed in psychiatry and medicine more generally starting as far back as Samuel Johnson's 1755 *Dictionary of the English Language* and subsequent revised editions (e.g., "Disorder: … Breach in that regularity in the animal economy which causes health"; "Derangement: Disorder, discomposure of mind"; "Illness: Sickness, disorder").

No Cartesianism Need Be Presupposed by "Mental Disorder"

Some argue that to talk about mental disorders as opposed to physical disorders is to already implicitly make an error, namely, to accept a Cartesian split between mind and body. No such assumption is necessary; the concept of mental disorder can be explored independent of Cartesian versus mind/brain identity or other philosophical accounts of the mind. Mental disorders are here assumed to be a distinctive set of processes, mechanisms, and functions that, although raising profound metaphysical issues, are from one perspective just another biologically shaped feature of the organism. They are distinguished by the fact that psychological processes are representational in that they have content that is aimed at certain features of the world (e.g., beliefs, emotions, and desires are about something) and are sometimes consciously experienced. Based on this characterization, certain capacities and features – such as belief, thought, perception, emotion, language, intentional action, and desire – are psychological and part of mental functioning. There is no intended Cartesian implication about any special ontological status of the mental; it is just an identified set of functions and processes distinguished by representational content or conscious awareness.

Must All Mental Disorders Be Brain Disorders?

There is a common and currently quite influential line of argument that psychological conditions can genuine medical disorders. The argument starts from a classical view of medical disorder as physically describable anatomical lesion in the body, implying that medical disorder is ultimately physical disorder. Psychological conditions could then be medical disorders if they are the symptoms of a lesion or other anatomically identifiable pathology in the brain, the seat of psychological functioning.

This general idea goes back to ancient Greek and Roman medicine, in which it was assumed that psychological conditions that are disorders must be due to bodily conditions that are themselves medical disorders. For example, depression was labeled "melancholia" (literally, "black bile disorder") by Greek and Roman physicians based on the theory that it was due to the body's excess of black bile. Hysteria meant "wandering uterus," indicating a physical cause underlying this partly behavioral disorder in women. In the modern era, German physician Wilhelm Griesinger (1882), for example, asserted in his mid-nineteenth-century psychiatry textbook that mental disorders are always symptoms of brain diseases. More generally, psychological symptoms have historically often been analogized to psychological changes that accompany physical disorders; for example, psychotic depression was long characterized as "delirium without a fever" by analogy to the delusions people have during high fevers accompanying physical diseases (Horwitz and Wakefield 2007).

Does a psychological disorder always require a brain pathology of some kind? Late nineteenth century successes in identifying such pathologies in a few cases, such as establishing that the disorder of "general paresis" is in fact due to neurosyphilitic infection and establishing that Alzheimer's disease involves pathology of brain tissue, nourished a faith among prominent psychiatric theoreticians that all mental disorders must be due to such brain pathologies. This hope was central to the initial aspiration of the task force that recently revised the American Psychiatric Association's diagnostic manual, the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association 2013), that brain etiologies and biomarkers could be incorporated into psychiatric diagnosis, thus making psychiatry more clearly medical. This goal was eventually abandoned for DSM-5, and no such additions were made to the manual, for the simple reason that such discoveries of brain-based physiological/anatomical causes have eluded researchers and continue to elude researchers to this day. It has become a scientific faith that such causes will eventually be identified in all cases of mental disorder, but no such causes are known at this time.

Lacking any such concrete scientific proof that the conditions labeled mental disorders are manifestations of brain disorders, those who argue along these lines are forced to use a more abstract argument. There is in fact an extremely influential argument that mental disorders must be manifestations of brain disorders, which goes like this:

> Mental processes take place in the brain.
>
> Therefore, if something goes wrong with mental functioning, something must be going wrong with brain functioning.
>
> Therefore, all mental disorders are brain disorders describable in anatomical/physiological terms.

For example, psychiatrist and neuroscientist Eric Kandel, a Nobel laureate, states: "All mental processes are brain processes, and therefore all disorders of mental functioning are biological diseases. . ..The brain is the organ of the mind. Where else could [mental illness] be if not in the brain?" (Weir 2012, p. 30). So, the argument is

that because mental disorders take place in the brain, whenever there is a mental disorder, there must be a brain disorder underlying it in which the pathology is describable nonpsychologically in sheerly anatomical or physiological terms.

This argument may seem persuasive at first glance, but many have found it wanting. An alternative perspective is that at best one must remain agnostic because whether brain diseases are always present when there are mental disorders is a scientific question to be evaluated by scientific discoveries, not a conceptual question to be resolved by logical analysis. This perspective allows that it is at least a coherent idea that there could be mental disorders that are not brain disorders.

The most powerful counterargument to the "mental disorders take place in the brain, therefore they must be brain diseases" argument consists of an analogy. The analogy comes from cognitive science. Cognitive scientists often liken psychological processes like thinking to computer programs running in brain tissue. For example, sequences of thoughts are like the symbol manipulation from line to line in a computer program. Brain tissue in which mental processing occurs is analogous to the silicon chips that comprise the computer hardware in which programming runs, and indeed cognitive scientists often portray the brain as a kind of computer.

Now, it is common for things to go wrong with a computer's software even when there is nothing whatever wrong with the computer's hardware. In fact, most computer problems are of this sort, in which there are programming malfunctions for reasons occurring at the programming level of description, but the hardware is perfectly sound. In such cases, one would search in vain for some hardware damage to resolve the problem with the software, and even switching to a new computer may not alter the problem. This is because the software is itself designed to function in a certain way at the level of the manipulation of symbols, and this can go wrong independently of anything being wrong with the specific hardware in which it is run. Inappropriate inputs may be introduced to the program, two processing tasks may clash, memory may be exhausted, and all sorts of other programming malfunctions may occur even when the hardware is perfectly sound.

The cognitive science analogy between the thinking/brain and software/hardware distinctions provides an elegant counterexample to the logic of Kandel's argument. To see this, one simply has to substitute software/hardware for mental/brain in his argument:

> Software processes take place in hardware.
> Therefore, all software malfunctions are hardware malfunctions.

This argument is plainly invalid. The premise only implies that all software malfunctions, being software processes, take place in hardware. It does not imply that the hardware process that comprises the software malfunction is itself also a hardware malfunction. This counterexample reveals that the original "all mental disorders are brain disorders" argument is also invalid.

A simple example of a psychological disorder that is not a brain disorder is misdirected imprinting in geese. The gosling imprints on whatever creature it sees first and faithfully follows it around. That creature, which is of course almost

invariably its mother. If a gosling accidentally imprints on a passing fox at birth) and follows the fox around leading to its death, the gosling's condition is a plausible candidate for a mental disorder. Yet there is nothing wrong with the gosling's brain at any level, which has worked exactly as it was designed to work. The problem is that the function of the brain mechanisms involved in imprinting is to get the gosling imprinted on its mother by internalizing an image that represents the mother has gone wrong. The failure is of meaning, not physiology. If as it happened the image derived from the fact the gosling's mother looked like a fox due to a cosmetic disorder, the very same image and thus the very same brain state would be perfectly healthy. It is only at the level of meaning and reference and psychological functioning, not at the brain level per se, that anything can be said to have gone wrong.

The claim that mental disorders are medical disorders because they are brain disorders has a further problem. Psychiatric science has not identified any brain etiologies, so it remains mysterious why so many conditions are judged to be mental disorders if brain disorder is the only basis for such an attribution. It seems instead that the grounds for a judgment of disorder lie in the nature of psychological processes themselves, from which it is then inferred that there might be a brain disorder. The "brain disorder" approach provides no independent criterion for when a psychological process is likely pathological. Yet, throughout medical history, mental disorders have been identified based strictly on the nature of their psychological symptoms. For example, Hippocrates defined melancholia as sadness or fear that goes on for too long a time. From Aristotle and Galen onward, physicians elaborated on how long is too long, specifying that disorder likely exists when the sadness or fear is out of proportion to real circumstances of loss or threat. Even today, the criteria for mental disorders are psychological and behavioral, as listed in the symptom criteria in psychiatric diagnostic manuals.

Indeed, this line of analysis could be used with equal force to argue to the contrary that we are not justified in considering any of the psychological conditions listed in psychiatric manuals as mental disorders that are medical disorders. After all, not one of these categories has been shown to be a manifestation of brain disease despite decades of sophisticated research. This is precisely the argument deployed by Thomas Szasz (1974) as well as other anti-psychiatric critics to prove that mental disorder is a myth (see below).

Interestingly, the currently much-discussed "Research Domain Criteria" (RDoC; Garvey et al. 2010) initiative in the USA is based on the idea that mental disorder often consists of brain circuitry that is pathologically overactivated or underactivated, so that the functioning of brain circuitry should be the focus of mental disorder research. Obviously, one might then look for underlying anatomical pathology. However, there is nothing inherently or generally pathological about high or low levels of neuronal circuitry activity; presumably, low activity of many circuits is common during sleep, as is high activity of specific circuits during sexual orgasm or acute fear in the face of imminent threat. To label circuit activity as pathological, one has to go beyond pure brain physiology and understand the psychological context of the activity and what the circuit is for, that is, what it is biologically designed to do (where "biological" is used in the evolutionary sense, not the sense of

physiology). Psychological evidence is often necessary before one can recognize whether a brain circuit's firing pattern is pathological, rather than the other way around (Wakefield 2014b).

For example, if a certain set of neurons in the male rat's brain is activated, it attacks an intruding male; if another set of neurons is activated, it attempts to mount an available female; and if yet another set of neurons is activated, it attacks an available female as if it is an intruding male (Anderson 2012; Lee at al. 2014). Presumably the latter case represents something like pathology, but there is nothing about the levels or locations of activation taken by themselves that would tell one this. One has to place the activation in the context of the presumed evolution of the psychological/behavioral functioning of the rat for mating and territorial defense. One does not first recognize a pathology in the brain and reason from that to the abnormality of the result of attacking receptive females; rather, one observes plainly pathological behavior and reasons to the likely pathology in the underlying pattern of brain activity.

Consequently, there must be something about psychological processes as such that can be at least plausibly recognized as normal versus pathological, independent of knowing the status of the underlying brain states. These considerations lead to the conclusion that to understand why some psychological conditions such as those listed in psychiatric manuals are judged to be disorders, one needs a criterion independent of brain functioning that concerns the psychological processes themselves.

The Values Account of Medical Disorder

There is an alternative strategy for locating mental disorder within medical disorder that abandons the idea of linking it to brain pathology and that defangs the anti-psychiatrists' conceptual objections in the process. However, it also undermines any scientific objectivity to the concept of disorder. This strategy is to claim that medical disorder itself is a value concept applying to undesirable bodily conditions. Given that the main concern about the concept of mental disorder is that it expresses value judgments about psychological functioning rather than labeling medical conditions, if medical disorder itself is a value concept, then there is no longer an argument; psychiatry can be about controlling socially undesirable psychological conditions and still be a legitimate medical discipline, too.

The value account surely has one part of the truth. It is not enough that something is objectively "wrong" with your body or mind for that to be a medical disorder in which health is diminished. The problem must also be harmful to the individual to be considered a disorder.

Most people have what physicians call "benign anomalies," that is, minor malformations that are the result of genetic or developmental errors but that cause no significant problem, and such anomalies are not considered disorders. For example, benign angiomas are small blood vessels whose growth has gone awry, leading them to connect to the skin, where they appear as small red dots. However,

because they are not harmful, they are not considered disorders and certainly are not thought to reduce health. To take another kind of example, inability to learn to read due to a dysfunction in the corpus callosum (assuming that this theory of some forms of dyslexia is correct) is harmful in literate societies, but not harmful or relevant in preliterate societies, where reading is not a skill that is taught or valued, and thus is not a disorder in those societies. One might imagine such conditions occurring in early human populations long before reading was invented, so no possible harm could come from such inabilities, and surely those individuals were not medically disordered.More generally, there are endless minor malfunctions, mutations, and other failures of normal functioning that are entirely harmless, and no one considers to reduce health. Every time one goes out in the sun, one's skin DNA suffers thousands of mutations, but these are not considered disorders – unless they accumulate in a way that triggers cancer or some other harmful skin condition. Another case in point is Typhoid Mary, who was a carrier of typhoid fever and so definitely had something pathological going on inside her body, yet developed no disease whatever and was universally considered to be healthy because the infection did her no harm (Wakefield 2014a). So, harm is essential to disorder.

However, harm is not enough. Any notion that mental disorder just is harmful psychological functioning can be dismissed because there are unfortunately so many forms of harm that human beings can suffer that are not disorders, such as ignorance, ugliness, poor judgment, lack of talent, lack of skill, moral weakness, illiteracy, bad manners, and sheer foolishness. So, granting that disorders are prima facie harmful conditions, they are just one category of the many negative mental conditions that can afflict a person.

What then determines that a negative condition is a disorder, beyond a value judgment? One might try to defend the value account by arguing that it is just the *kind* of values involved that determine whether a category of undesirable psychological states is considered a disorder. However, this cannot work because disorders can manifest in problems that are quite similar to problems one finds in other domains. Illiteracy due to lack of education and illiteracy due to dyslexia in which a neurological disorder makes learning to read extremely difficult can yield similar negative outcomes in our reading-oriented society, but one is a disorder and the other is not. Normal-range adolescent delinquency and conduct disorder can yield similar issues with the criminal justice system (indeed, DSM-5 itself indicates that a normal response to a threatening environment can yield many of the same problematic behaviors as conduct disorder), yet one is a disorder and the other is not. Lack of sexual arousal may indicate a disorder in some contexts, but, as DSM-5 indicates, if there is lack of adequate stimulation or a relationship that is abusive, it may be perfectly normal. Anxiety may indicate a disorder or it may not, depending on whether it is a response to a credible imminent threat. Grief is seen as normal, whereas similarly intense sadness "out of the blue" is seen as depressive disorder, yet major depression and normal grief can be painful in pretty much the same ways. Indeed, DSM-5 specifies in a note to the depression criteria that the clinician should use judgment in diagnosing depression because very same symptoms can occur in normal reactions to loss and in pathological depression. In sum, disorder need not

involve a distinctive type of harm, so it must be something other than the nature of the harm that distinguishes disorder from other negative conditions.

A hint of what that additional requirement is emerges from examining one of the most famous of the value theorist's arguments. Peter Sedgwick (1982) claimed: "All sickness is essentially deviancy [from] some alternative state of affairs which is considered more desirable. . . The attribution of illness always proceeds from the computation of a gap between presented behavior (or feeling) and some social norm" (pp. 32–34). This is true enough, but the fact that all disorders are undesirable and harmfully deviate from socially valued conditions shows only that values are *part* of the concept of disorder, not that disorder is composed only of values.

Sedgwick attempted to demonstrate that values are all that matter in disorder judgments through vivid examples showing that there is nothing objective or scientific that distinguishes conditions considered disorders from other processes in nature, leaving the value element as the only identifying characteristic:

> There are no illnesses or diseases in nature. . .. The fracture of a septuagenarian's femur has, within the world of nature, no more significance than the snapping of an autumn leaf from its twig; and the invasion of a human organism by cholera-germs carries with it no more the stamp of "illness" than does the souring of milk by other forms of bacteria. . .. Out of his anthropocentric self-interest, man has chosen to consider as "illnesses" or "diseases" those natural circumstances which precipitate . . . death (or the failure to function according to certain values). (1982, p. 30)

However, there is a relevant difference between snappings of femurs and leaves. Leaves are biologically designed to fall off from certain trees at certain times of year, and the tree is not designed to require the leaf for its continued functioning, whereas the possession of an intact femur is part of the way a person, even an old person, is biologically designed to function, and there is no natural selection for broken femurs. Similarly, once extracted from the cow, milk has no natural function, so the bacteria that invade and sour it are not causing a dysfunction, whereas the person infected with bacteria is suffering (or in danger of suffering) loss of functional integrity. Thus, there is a scientifically definable non-value difference between Sedgwick's examples of natural processes that are disorders and those that are not; the disorders disrupt a natural biologically designed function, whereas the nondisorders do not.

Recognizing that most undesirable states are not considered medical disorders (e.g., poverty, oppression, being sexually rejected), Sedgwick (1982) tried to save the value account by adding one factual requirement – that the cause of the undesirable condition could not lie entirely in external circumstances but must be inside the individual's body or mind. This eliminates the above counterexamples, but it does not explain why many other undesirable conditions that are internal, such as ignorance, lack of talent, the pain of teething, or unwanted pregnancy are also not considered disorders. The biological dysfunction criterion explains why these latter conditions are not disorders: although internal, they do not involve a breakdown in the biologically designed functioning of an internal mechanism.

In sum, no distinction based sheerly on kinds of negative outcomes can explain our concept of disorder versus nondisorder. The concept of disorder must include an

additional conceptual component beyond the value component that distinguishes those negative conditions that are disorders from the vast array of nondisordered negative conditions. The analysis of Sedgwick's examples suggests that such a distinction might be based on the concepts of biological function and dysfunction in the evolutionary sense, referring to what a feature is biologically designed to do. However, for disorder, neither harm (because there are many nondisordered harmful conditions) nor biological dysfunction (because there are many benign anomalies that do not diminish health) is sufficient. Instead, both are necessary. "Disorder" must be understood as a hybrid fact/value concept – harmful dysfunction – that identifies conditions in which a failure of biologically designed functioning of some internal mechanism causes harm to the individual (Wakefield 1992).

From Anatomical Structure to Evolutionary Function

It is time to circle back and reconsider the anti-psychiatry argument in order to gain further clarity about mental disorder. Szasz (1974), the leading anti-psychiatric critic of the notion of mental disorder, argued as follows. Physical disorder is a legitimate concept based on a clear foundation, namely, the presence of a lesion that is a recognizable deviation in anatomical structure. Mental disorder, to be a legitimate concept, must be identical to this original concept. However, mental disorder is used to label behavior that deviates from social norms and is typically not accompanied by any identifiable lesion of the brain or of any other part of the body. Thus, Szasz argued that the lesion concept of disorder that is applicable to physical conditions is not applicable to mental conditions, and mental disorders are not literally disorders. Mental disorder – a "lesion of the mind" – is at best a metaphor that has been mistaken for the literal truth.

One obvious response is that we just have not discovered such lesions yet but will in the future. Szasz could reply that biological psychiatrists often talk as if such discoveries are around the corner, but there is no evidence for this or any sense of which conditions will be found to correspond to lesions. Consequently, for now the only real grounds for classifying and treating people is the social undesirability of their behavior. Moreover, if a lesion was discovered, the condition would be considered a physical disorder. Szasz concluded that "there is no such thing as 'mental illness'" (1974, p. 1).

A different way of responding to Szasz emerges from our analysis above, if we ask: what makes a lesion a medical disorder? The weakness in Szasz's argument lies in the inadequacy of the lesion account of physical disorder, which rests on two theses: (a) a lesion (or abnormal bodily structure) is a statistical deviation from typical anatomical structure and (b) a physical disorder is simply a lesion. However, first, the idea that a lesion in a sense relevant to disorder identification can be directly recognized by its deviant anatomical structure is incorrect. Bodily structures normally vary from person to person, and many normal variations are as unusual as any lesion. Moreover, some lesions are statistically nondeviant in a culture, such as atherosclerosis, minor lung irritation, and gum recession in American culture and

hookworm and malaria in some others. Therefore, recognition of a lesion is not simply a matter of observing anatomical deviance; something more is involved.

Second, even if one could recognize anatomical abnormalities as lesions, the existence of a lesion is neither necessary nor sufficient for disorder. There are physical disorders, such as trigeminal neuralgia and senile pruritis, for which there are no known anatomical lesions and for all we know there may be no such lesions, just as for mental disorders (Kendell 1975). Moreover, a lesion can be a harmless abnormality that is not a disorder that diminishes health, ranging from benign anomalies to the heart being in reversed position on the right side of the body but retaining functional integrity, or a virus invading some cells but being held in check by the immune system and causing no symptoms or contagion. Thus, the lesion account of physical disorder fails, and with it goes the classic skeptical argument that the concept of disorder cannot literally apply to mental conditions not caused by lesions.

How, then, do we recognize lesions (i.e., anatomical abnormalities) that are disorders? Roughly, we recognize a variation in anatomical structure in a specific mechanism as a lesion rather than as a normal variation if the variation is caused by some failure in the mechanisms that generated it to perform their functions, so the lesion is an outcome of dysfunction in lower level or other related mechanisms. Moreover, the lesion itself is recognized as a dysfunction if it impairs the ability of the mechanisms of which it is a part to accomplish the functions that they were biologically designed to perform (i.e., for which they were naturally selected). Finally, we recognize a dysfunction as a disorder if the dysfunction affects the well-being of the overall organism in a negative way. A wayward malfunctioning cell that has no impact on the overall organism's well-being is not a diminution of health and not a medical disorder. So, the idea of a lesion was never fundamental to medical disorder; it was only via the failures of biological function that lesions indicate that lesions are linked to medical disorder.

Dysfunction as Failure of Biologically Designed Function

There are many concepts of function and dysfunction. The challenge is to identify the concept of dysfunction that is medically relevant. Presumably a dysfunction implies an unfulfilled function, that is, a failure of some mechanism in the organism to perform its function. However, not all uses of "function" and "dysfunction" are relevant to judgments of disorder. Clearly, the medically relevant sense of "dysfunction" is not the common colloquial sense in which the term refers to failure of an individual to perform well in a social role or in a given environment, as when people say things like "I'm in a dysfunctional relationship" or "I've been tired and dysfunctional at work." These kinds of problems need not be individual medical disorders. A disorder is different from a failure to function in a socially or personally preferred manner precisely because a dysfunction exists only when something has gone wrong with internal functioning, so that a mechanism cannot perform as it is naturally (i.e., independently of human intentions) supposed to perform.

The professional literature too sometimes confusingly uses "dysfunction" to mean impairment of social role performance, which is not the sense relevant to attributing disorder because people have challenges in fulfilling social roles for all sorts of nonmedical reasons.

A plausible alternative, suggested by the analysis of Sedgwick's examples, is that the functions that are relevant to disorder are "natural" or "biological" functions, meaning those functions for which a mechanism was naturally selected and thus the function that explains why we have the mechanism in the first place (Wakefield 1999, 2011, 2016). Such functions are frequently attributed to inferred mental mechanisms that may remain to be identified, and their (inferred) failures are labeled "dysfunctions." For example, a natural function of the perceptual apparatus is to convey roughly accurate information about the immediate environment, so gross hallucinations indicate dysfunction. Some cognitive mechanisms have the function of providing the person with the capacity for a degree of rationality as expressed in deductive, inductive, and means-end reasoning, so it is a dysfunction when the capacity for such reasoning breaks down, as in severe psychotic states.

If "dysfunction" refers to failure of naturally selected features, then, although vague, the notion of a dysfunction in principle is a factual scientific idea. Admittedly, however, at this point we remain in great ignorance of evolved mechanisms, functions, and dysfunctions, which is what makes psychiatric classification so challenging and is why we depend so much on inference from circumstantial evidence. Discovering what in fact is natural or dysfunctional (and thus what is disordered if harmful) may be difficult and may be subject to scientific controversy, especially with respect to mental mechanisms, yielding confusion and controversy about mental disorders. However, functional explanations still can be plausible and useful even when little is known about the actual nature of a mechanism or even about the nature of a function. For example, we know little about the mechanisms underlying sleep, and little about the functions of sleep, but circumstantial evidence persuades us that sleep is a normal, biologically designed phenomenon (despite incapacitating us for roughly one-third of our lives), and the circumstantial evidence enables us to distinguish some disordered sleep conditions from normal sleep despite our ignorance.

Medical Disorder as Harmful Dysfunction

In sum, a medical disorder exists when there is a *harmful dysfunction* – that is, there is a failure of some part of the organism to perform its function such that the impairment causes harm to the organism. The harmful dysfunction analysis of the concept of disorder explains what the value theorist's account cannot explain, namely, which negatively evaluated conditions are medical disorders and which fall under other categories of harm or misfortune. It also explains what the skeptics' lesion account cannot explain, namely, which anatomical deviations are lesions (those produced by biological dysfunctions) and which lesions are disorders (those that are harmful).

If disorder is essentially a functional concept, then mental conditions and physical conditions can literally be medical disorders for the very same reason, namely, their functional implications. Considering that psychological processes play important species-typical roles in human survival and reproduction and seem to be supported by specific neurocircuitry modules with differentiated function, there is no reason to doubt that mental processes were naturally selected and have natural functions, as Darwin himself often emphasized. Because of our evolutionary heritage, we possess physical mechanisms such as livers and hearts and their submechanisms; that same heritage gave us mental mechanisms such as various cognitive, motivational, affective, personological, hedonic, linguistic, and behavioral dispositions and structures.

Mental Disorders as Harmful Psychological Dysfunctions

The chapters in the DSM indicate the various domains in which psychiatrists feel confident that human beings are biologically designed to function in certain ways so that disordered deviations when things go wrong can be plausibly identified.

For example, very roughly:

Psychotic disorders involve failures of biologically designed thought processes to work as designed to provide rationally justified beliefs and perceptual processes to provide roughly accurate information about the environment.
Anxiety disorders involve failures of anxiety- and fear-generating mechanisms to work as designed.
Depressive disorders involve failures of biologically designed sadness and loss-response regulating mechanisms.
Disruptive behavior disorders of children involve failures of biologically designed socialization processes and processes underlying conscience and social cooperation.
Sleep disorders involve failure of sleep processes to function properly.
Sexual dysfunctions and paraphilias involve failures of various biologically designed mechanisms involved in sexual motivation and response.
Eating disorders involve failures of biologically designed appetitive mechanisms.
And so on.

There is a certain amount of nonsense in psychiatric classification, and the diagnostic criteria are often overly inclusive, creating false-positive diagnoses that mislabel normal experiences as disorders (Wakefield 2015a), as in the inflated categories for depressive and anxiety disorders (Horwitz and Wakefield 2007, 2012, Wakefield 2013). Some categories are likely entirely composed of nondisorders. For example, in DSM-5, disorders such as circadian rhythm sleep-wake disorder, shift work type (in which an individual has difficulty adjusting to shift work and sleeping at odd hours); social anxiety disorder (social phobia), performance only type (in which the only problem is anxiety about public performance); and psychological factors affecting other medical conditions (which involves various personality

features or actions that are not disorders but interfere with medical treatment or exacerbate a medical condition) are arguably vacuous disorder categories because the target conditions that the category is constructed to encompass do not involve dysfunctions. However, the vast majority of categories in the standard diagnostic manuals are inspired by conditions that both a professional and a lay person would plausibly identify as a harmful failure of biologically designed psychological functioning and therefore a mental disorder in the medical sense of "disorder."

Conclusion: Mental Disorders as Harmful Psychological Dysfunctions

The analysis of "medical disorder" thus reveals that psychological conditions can be "mental disorders" that are genuine medical disorders. Mental disorders in the medical sense are harmful failures of the biologically designed functioning of psychological mechanisms. Such disorders can often be plausibly identified from circumstantial evidence and thus identified from the psychological evidence alone, without any reference to whether or not there is an underlying brain lesion, and thus many current psychiatric categories of disorder are likely "conceptually valid" (Wakefield 1992), meaning that they do identify genuine disorders. However, other categories do not plausibly satisfy the harmful dysfunction criteria and thus are likely "false positives" in which individuals with normal distress or other normal variants of psychological conditions are being mistakenly labeled as disordered, and such errors need to be corrected.

The harmful dysfunction analysis thus provides the two crucial elements of a conceptual analysis of psychiatry as a part of medicine. It explains psychiatry's in-principle legitimacy as a medical discipline addressing genuine medical disorder in the same sense of "disorder" used in physical medicine. And, it defines the limits of legitimate psychiatric labeling and thus provides some in-principle protection against the oppressive use of psychiatry for social control purposes.

Summary Points

- For a psychological condition to be a genuine medical disorder, it has to be a disorder in the same sense of "disorder" as physical disorders.
- A medical disorder is a harmful dysfunction, that is, a harmful condition that is caused by the failure of some biologically designed mechanism in the individual to perform its function.
- Value judgments are not sufficient to define mental disorder because most negative or harmful psychological conditions, ranging from lack of talent and ignorance to criminality and abrasiveness, are not medical disorders.
- Something going wrong with biologically designed psychological functioning is not sufficient to define medical disorder because many psychological anomalies are harmless.

- Human beings have been biologically designed with psychological functions just as they have been biologically designed with physical organs.
- To be a genuine medical disorder, a psychological condition must be a "harmful psychological dysfunction"; that is, it must be harmful, and the harm must be caused by the failure of some biologically designed psychological mechanism to perform its function.
- Psychiatry and the other mental health professions are a legitimate part of medicine.
- Because most negative or socially disapproved psychological conditions are not genuine medical disorders, there are strict limits to the legitimate use of medical power to control socially disapproved psychological functioning.
- Although all mental disorders may be brain disorders, it is also possible that some mental disorders involve failure of biologically designed cognitive or emotional processing rules even though there is no physiological brain dysfunction, just as software can malfunction when running in perfectly sound hardware.

References

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Association, Arlington

Anderson DJ (2012) Optogenetics, sex and violence in the brain: implications for psychiatry. Biol Psychiatry 71(12):1081–1089

Garvey M, Heinssein R, Pine DS et al (2010) Research domain criteria (RDoc); toward a new classification framework for research on mental disorders. Am J Psychiatry 167:748–751

Griesinger W (1882) Mental pathology and therapeutics (trans: Robertson C, Rutherford J). William Wood %26 Company, New York

Horwitz AV, Wakefield JC (2007) The loss of sadness: how psychiatry transformed normal sorrow into depressive disorder. Oxford University Press, New York

Horwitz AV, Wakefield JC (2012) All we have to fear: psychiatry's transformation of natural anxieties into mental disorders. Oxford University Press, New York

Kendell RE (1975) The concept of disease and its implications for psychiatry. Br J Psychiatry 127:305–315

Lee H, Kim DW, Remedios R et al (2014) Scalable control of mounting and attack by ESR1+ neurons in the ventromedial hypothalamus. Nature 509(7502):627–632

Pratt LA, Brody DJ, Gu Q (2011) Antidepressant use in persons aged 12 and over: United States, 2005–2008 (NCHS data brief, no 76). National Center for Health Statistics, Hyattsville

Sedgwick P (1982) Psycho politics. New York: Harper & Row

Szasz TS (1974) The myth of mental illness: foundations of a theory of personal conduct. Harper % 26 Row, New York

Visser SN, Danielson ML, Bitsko RH et al (2014) Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. J Am Acad Child Adolesc Psychiatry 53:34–46

Wakefield JC (1992) The concept of mental disorder: on the boundary between biological facts and social values. Am Psychol 47:373–388

Wakefield JC (1999) Evolutionary versus prototype analyses of the concept of disorder. J Abnorm Psychol 108:374–399

Wakefield JC (2010a) Misdiagnosing normality: psychiatry's failure to address the problem of false positive diagnoses of mental disorder in a changing professional environment. J Ment Health 19(4):337–351

Wakefield JC (2010b) False positives in psychiatric diagnosis: implications for human freedom. Theor Med Bioeth 31(1):5–17
Wakefield JC (2011) Darwin, functional explanation, and the philosophy of psychiatry. In Pieter R. Adriaens and Andreas De Block, Maladapting minds: Philosophy, psychiatry, and evolutionary theory (pp. 143–172). Oxford: Oxford University Press
Wakefield JC (2013) The DSM-5 debate over the bereavement exclusion: psychiatric diagnosis and the future of empirically supported treatment. Clin Psychol Rev 33(7):825–845
Wakefield JC (2014a) The biostatistical theory versus the harmful dysfunction analysis, part 1: is part-dysfunction sufficient for medical disorder? J Med Philos 39(6):648–682
Wakefield JC (2014b) Wittgenstein's nightmare: why the RDoC grid needs a conceptual dimension. World Psychiatry 13(1):38–40.
Wakefield JC (2015a) DSM-5, psychiatric epidemiology, and the false positives problem. Epidemiol Psychiatr Sci 24(3):188–196
Wakefield JC (2015b) Psychological justice: DSM-5, false positive diagnosis, and fair equality of opportunity. Publ Aff Q 29(1):32–75
Wakefield JC (2016) The concepts of biological function and dysfunction: toward a conceptual foundation for evolutionary psychopathology. In D. Buss (Ed.), Handbook of evolutionary psychology, 2nd edn (vol. 2) (pp. 988–1006). New York: Oxford Press
Weir K (2012) The roots of mental illness: how much of mental illness can the biology of the brain explain? APA Monit 43(6):30

Curing and Healing: Two Goals of Medicine

6

Dorota Szawarska

Contents

Abstract

This chapter will begin by signaling potential problems with the dichotomy between curing and healing. It continues to explore three different ways of thinking about the dichotomy: rational vs. irrational, the meaning of curing and healing as experienced by patients and practitioners, and the relationship between curing and healing in the practice of medical care. It is argued that while the distinction between curing and healing is not a universal one, as it is based on a Western distinction between disease and illness, both curing and healing require taking responsibility for the well-being of the vulnerable patient.

D. Szawarska (✉)
Warsaw, Poland
e-mail: d.szawarska@gmail.com

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_59

Introduction

Curing and healing are two categories that appear central to the practice of medicine. And yet given the variety of healthcare systems, practices, and beliefs present in the world and the universal problem of illness and vulnerability, there is some doubt as to the validity of the distinction between the terms. This is the central problem that organizes the remainder of this chapter. In it the problem of the dichotomy between healing and curing will be explored, as well as the uses and meaning attached to these categories.

The Problem

Being ill and in need of medical assistance is a universal human experience. At times, we all need specialized help in order to make us better. What is different is the way a health-related problem is approached in various cultural contexts and the explanatory models of illness that are present. Once attempts at self-curing or perhaps self-healing are exhausted, one will presumably contact a specialist. In a provincial small British town, one probably will pay a visit to the general practitioner, who will prescribe some medicine and a course of action fitting with the Western medical tradition. In an isolated indigenous community in the Amazon, one will probably turn to a traditional healer or a shaman, who will provide the sick with herbal medicine and perhaps perform a healing ritual. In a multicultural city such as Hong Kong, one will face a choice between Western medicine and traditional Chinese medicine and depending on circumstances and personal beliefs will choose accordingly. Assuming the specialists approached are successful in their endeavors to help the patient and the patient feels markedly better or better still – recovers completely – ask yourself this: is it fair to say that the GP cured his patient and the shaman healed his? Now, try an experiment and say: The GP healed his patient, the shaman cured his. Are you completely happy with the swap of the terms? Assuming you live in the West and work for a Western company, would you be comfortable saying to your employer – "I am fit to work, my doctor healed me completely." Maybe not. And it is interesting why not. And would you trust a healer as much as you would trust a doctor? But now ask yourself this – is there an essential difference between the accomplishment of the GP and that of the shaman? Both were successful; both patients were made better; the problem was solved. Perhaps the source of some of the discomfort one might feel in swapping the terms, or using the word "heal" in a formal Western context, lies not so much in the difference between "healing" and "curing" but in what we consider to be proper medicine. Perhaps the GP was successful because she employed the principles of scientific, evidence-based medicine, whereas the shaman was just lucky, and his actions, despite existing system of beliefs in the given part of the Amazon, were groundless, especially from a Western point of view, "healing" being some fuzzy concept, free from common sense, and an efficacious practice. But if we so lightly discount non-Western medical systems and traditional healers as medicine proper,

then why the problem of "healing and curing as two goals of medicine"? Is not "curing" enough? And if we do not disregard non-Western medical traditions and accept that indeed traditional healers have the basis to help their patients, cannot we just use the terms curing and healing interchangeably in the context of non-Western medical practice? Perhaps the processes of curing and healing are two sides of the same coin – namely, making the patient better (see also Hutchinson et al. 2009).

Three Approaches

There are three broad ways to approach the differences between curing and healing. One is to focus on the difference between rational Western medicine as opposed to nonrational healing lying outside of medicine and the problem of defining "medicine" as a scientific practice. The second is to consider what the terms might mean to the people experiencing curing and healing as patients or as practitioners of medicine broadly understood. The third possible approach is to explore the terms not so much by contrasting them but by analyzing the relationship between curing and healing in the context of the relationship between patients and practitioners. This last approach will be briefly considered in the section "Curing and Healing as Two Aspects of Medical Care."

Let us first turn to the first approach.

1. The distinction based on curing grounded in rational Western medicine and irrational healing may be presented in the following manner. Bear in mind this does not take into account ethnographic evidence relating the process of becoming a healer and non-Western explanatory medical models, and which is why in presenting the argument the focus is on "faith healing":

   Contemporary medicine as a rational system of knowledge and practice is governed by laws of science and is evidence based. The results of actions based on that system are predictable and repeatable and work regardless of cultural context. Curing comes about as a result of applying appropriate medical knowledge and practice to the medical problem at hand. Medical practitioners, thanks to an organized and thorough system of education, know and can usually explain and provide evidence for why certain actions bring about a cure and why they occasionally fail to do so.

   In turn faith healing can be said to lie outside of the rational. Healing comes about as a result of applying through prayer and laying on hands of a supernatural, mysterious power that cannot be explained, taught, or rationally acquired. The healer heals but does not know why he can heal, how his gift works. In a case of miraculous faith healing, one focuses on the persona of a healer, where as in the case of curing, one focuses on a system of knowledge and practice. The results of thus understood healing are unpredictable, are unrepeatable, and from the point of view of Western medical practice are accidental. There is no causal connection, unless one takes into the account the placebo effect, between the actions of a healer and the state of health of his patient. From the point of view of

Western medicine, faith healing is a potential source of harm and should be discouraged.

There are of course problems with this argument. Most of these center on recognizing Western medical system as the only valid scientific system there is and, by definition, better than any alternative. This Eurocentric point of view might make us blind to what non-Western medical systems have to offer both in terms of healing and curing, especially that healing cannot be reduced to "faith healing." Moreover, non-Western medical systems do not necessarily have the distinction between illness and disease on which it will be argued the dichotomy between curing and healing is largely based. But if the difference between the terms does not boil down to the difference between the rational and the nonrational and Western and non-Western, where does it lie? Let us turn to the second approach to the difference between the terms and focus on what it means in terms of the experience of being cured or healed.

2. It is sometimes said that one cures a disease and heals an illness, where simply speaking illness is the personal experience of being unwell, shaped in part by one's culture, place in society, and personal circumstances and the disease is the underlying organic, physical cause of being unwell (Cassell 2004, 2012; Lerner 1994), and together, disease and illness describe a sickness. From the differences between illness and disease follows another important dichotomy, namely, the difference between pain associated with a disease and suffering associated with an illness (Lerner 1994). All those elements have an impact on the place and role of the medical practitioner (or practitioners) and the patient in the process of getting better.

   Imagine for a moment that the practitioner is a firm believer in the biomedical model, with its focus on curing the disease. This approach is said to limit the involvement of the patient in the process of getting better: the patient is interviewed, various tests are performed, diagnosis is given by the physician, and a course of action is prescribed – it might consist of further tests, or taking some form of medication, or some more advanced treatment performed by the physician or a whole medical team on the patient, with limited contact between the patient and at least some of the individual members of the team (e.g., the patient will probably not see or talk to the radiologist analyzing complex USG or tomography images or the person analyzing blood samples). In short, things are done to the patient (Milstein 2005), and the patient is expected to comply with the action prescribed. Can this approach be valid and successful? Certainly, provided one is dealing with relatively simple matters that are easy to resolve: a straightforward case of appendicitis or some simple infection easily treated with a series of antibiotics, easy to diagnose, and easy to treat. It is worth remembering during this thought experiment that while many good physicians also take into the account the needs of their patients as human beings with specific social circumstances, worries, and resources, healthcare systems in developed countries tend to focus on the underlying organic causes of medical problems, simply because it is easier to put a price tag and a time frame on the treatment required.

> And yet to focus entirely on curing a disease might not be sufficient to make a patient completely better, simply because patients, apart from having a disease, are also part of a wider sociocultural fabric, which makes them react to being unwell in a specific manner and which also makes them attach a particular meaning to the episode of being unwell. Patients not only feel pain, they also suffer. And while a pain killer might be sufficient to deal with physical pain, it might not be sufficient to deal with suffering. This is where healing comes in. As argued by Egnew: "Healing is the personal experience of the transcendence of suffering" (Egnew 2005: 258).

Given the definitions of illness and disease, it is debatable whether humans ever experience disease as such. Being aware of being unwell is already a part of the cognitive, emotional, cultural, or even spiritual experience of being ill. And therefore a physician's focus on curing a disease might not be sufficient to deal with the problem, especially if we are dealing with a chronic or incurable condition. Healing on the other hand is said to take into the account the human condition and experience of being unwell, including social, cultural, historical, and economic factors (Crandon Malamud 1991; Finkler 1994; Waldram 2000). But what exactly is healing and how is it achieved? It is said that healing is a process that promotes health and restoration of balance between mind and body (McGlone 1990: 77–84). There is no agreement among academics as to what exactly the process entails, but the following elements appear in various accounts and definitions of healing (e.g., Glaister 2001; Hutchinson et al. 2009; Egnew 2009):

(a) Healing actively engages the patient.
(b) Healing is multidimensional.
(c) Healing is creative and meaning making.
(d) It leads to restoration of balance and the acceptance of status quo.
(e) Healing process can involve a whole group of people. The problem does not have to be an individual one. Neither the healed nor the healer needs to be an individual.

Let us briefly explore these elements of healing also in relation to the concept of curing.

A practitioner who works on healing an illness ought to engage with the ill person, in order to assist him in regaining the feeling of being in the right place and the right time as to his body and mind. However, it is said that healing is not something done to the patient but something that takes place within the individual with the help of his active participation, through the patient's commitment to doing what is required to heal (Glaister 2001: 64; Levine 1987; Mulloney and Wells-Federman 1996). Five steps in healing have been identified that are signs of active participation: awareness, appraisal, choosing, alignment, and acceptance (Scandrett-Hibdon and Freel 1989). The engaging aspect of healing can be contrasted with curing, which is seen as primarily doing something to the patient

(Milstein 2005: 566; Samuel 1990: 88). Yet this may be an overstatement. Under normal circumstances patients undergoing a cure in the context of Western medicine are not passive, as the phrase would suggest. For the most part they are actively engaged in the whole process, starting with the decision to visit the family doctor, complying or not with the doctor's advice (this is especially true in relation to lifestyle changes recommended or taking prescribed medicine), and finally the decision to terminate further treatment. Perhaps what matters in the understanding of the difference between curing and healing is not so much the factual engagement and participation of the patient in the process of getting better in the context of curing and healing but how that difference is constructed and perceived by patients and practitioners alike. In the case of healing as an element of alternative therapies in the West, patients tend to perceive healing rituals and activities as ones that engage them, while they see the doctor-patient relationship in the Western medical tradition as one riddled with power inequality and requiring passive compliance on their part (McGuire and Kantor 1998: 201).

Healing can be seen as multidimensional, especially if one sees healing as achieving a balance between various dimensions of the people undergoing the healing process, namely, the physical, emotional, mental, social, or spiritual (Glaister 2001: 64). In that respect healing is markedly different from a definite cure leading to an absence of a disease. The multidimensional aspect of healing means that the absence of disease is neither sufficient nor necessary for healing to occur. What matters is the acceptance of the status quo, coping with and integrating the demands of one's illness or disease (Coward and Reed 1996) and its aftermath. Consider the case of a woman undergoing treatment for breast cancer. First, she needs to adjust to the situation of being seriously ill and deal with the chemotherapy, its side effects, and their consequences in day to day life, as after all apart from being ill, she remains a daughter, mother, partner, and a woman. Even if following mastectomy and chemotherapy she is declared to be free of cancer, it will take more than that before she feels whole again. She, for example, needs to learn to accept herself without a breast or with reconstructed breasts. Also, a brush with a serious life-threatening disease might demand psychological and social adjustments and reevaluation of one's life (see also Dobkin 2009). Those elements are important parts of the healing process and lead us onto the next point, namely, that healing is said to be a creative and meaning-making process – in order to make sense of one's illness or even approaching death, one needs to give it meaning (Good 1994). The meaning might be created by the person undergoing healing, or it might be developed with the help of the healer (Egnew 2009).

The meaning-making aspect of healing points to another major difference associated with the dichotomy of curing and healing, namely, that the aim of curing is restorative, while healing is transformative (Hutchinson et al. 2009: 845). By curing one eradicates a disease or corrects a problem. One "removes" the changes caused by the sickness, and brings back the patient, as far as possible, to the ideal, healthy starting point. By healing one brings about a change in the patient, whether by changing his attitude to illness, by creating a new meaning in his life, or by

giving him or her greater sense of integrity and place in the world following an illness or while facing approaching death.

It is worth remembering that the healing process can stretch beyond the individual healer and the individual in need of assistance. At first glance this is not such a great difference from curing in the context of Western medicine. On the one hand the curing process can involve a whole medical team, and on the other it does not need to focus on a single individual. Such is the case of treating STDs or other venereal infections, where relevant practitioners prescribe medicine for both partners, or large-scale medical emergencies, where a whole population is the focus of medical activities and surveillance, such as in the case of the Ebola epidemy in Africa. However, in the case of the healing process, ethnographic evidence suggests that what is at stake is not so much the well-being of a collection of individuals but the well-being of the whole community bound by specific social relationships (Katz 1982; Vermeylen and van der Horst 2007: 179) or a family or kinship group (Turner 1967). Arguably, what is being healed and strengthened are the relationships between people. Whether healing is focused on individuals, the whole community, or specific relationships within, it depends on the social and cultural construction of self (Scheper-Hughes and Lock 1987).

Healing Without Curing

Healing is sometimes said not to necessitate a cure in the biomedical sense (e.g., Glaister 2001: 64). It is argued that getting better in terms of a patient's self-assessment can be achieved by better coping with sickness and a restoration of balance, both achieved through the process of healing. This process sometimes requires that the point of balance is shifted and that what is restored is not so much the previous status quo but a balance resting on a new understanding and acceptance of self in the world. This is particularly the case of people coping with chronic diseases and those nearing the end of their lives. The easing of suffering is achieved through gaining acceptance of the situation, giving it meaning, and adapting.

Curing Without Healing

Hypothetically, curing, in the biomedical sense, can also be achieved without healing. This is especially so in cases where the patient is not aware of being sick and of having a disease. In such cases the problem might be diagnosed by some routine testing during, say, a health check and easily treated, without giving the patient the time to consider herself unwell. Perhaps a good example of this is a case of mild vitamin D deficiency. Before diagnosis, symptoms associated with it, if at all noticed, might be blamed on the time of year, overwork, etc. but might not be connected to one's health. Another type of situation in which one might be dealing with a kind of curing without healing is one in which from the biomedical point of

view the problem is sorted or managed as well as possible according to current medical knowledge, with any physical symptoms being well taken care of, without the patient regaining their sense of well-being and balance. This might be, for example, the case of a woman recovering from stab wounds inflicted by her partner during domestic abuse incident. Her physical wounds might be cured, but she might, as argued by Erickson (2007: 10), never feel truly healed. That is, in spite of a successful cure, her quality of life continues to suffer (see also Eisenberg 1977).

Curing and Healing as Two Aspects of Medical Care

For all the importance attached to the distinction between curing and healing in Western medical practice and thought and philosophical and anthropological work on both Western and non-Western medical systems, it is worth remembering that each and every medical system is a cultural system (Rhodes 1996) and each involves elements of both curing and healing. Indeed, it may be argued that the distinction between curing and healing is overstated as is the dichotomy between illness and disease. Cassell (1976) points out that the very notion of an organic disease as a cause of a sickness is the central concept in the Western medical model. The notion that a malfunctioning body is what makes a person feel ill lends itself to the formulation of the distinction between curing and healing, where we cure the disease and heal the person. But what happens in contexts where there is no concept, or only a limited concept of disease as a cause of a sickness, and where the explanatory model of illness is completely different from the Western one? In such a context, the distinction between curing and healing is unlikely to be valid. If the sickness is believed to be caused by an invasion of evil spirits, or witchcraft, or upset ancestors there, the medical practitioner needs to take culturally appropriate action to deal with the problem, and that is not identical to dealing with a disease. And even within the Western context, Waldram (2000: 606) argues that healing an illness and curing a disease are not separate, unrelated aspects of the treatment of sickness. As argued by Lown (1999: 313): "Whereas the medical transaction is largely concerned with curing a disease, the patient craves to be healed. The object of the patient's art is to have the doctor incorporate healing in the process of curing." This is apparent, for example, in the effect the interaction between patient and physician has on how one judges the efficacy of treatment. Consider how in the biomedical system the patient's self-assessment of how he or she is feeling following treatment is taken into the account in order to judge the effectiveness of the curative treatment. Similarly, a physician's positive proclamation on the effectiveness of treatment may lead to an improvement in the patient's subjective well-being (Waldram 2000: 607). This suggests that despite doubt as to the validity of the dichotomy between curing and healing, it is worthwhile to explore the relationship between the two processes as they are understood in the Western context.

Conclusion

Caring for a patient, whether we focus on curing or healing, involves many different aspects: from defining and accepting the person as a patient in need of treatment, diagnosing them, and treatment proper. Also, even in the Western biomedical context, the patient is not necessarily an individual: it can be a group of related persons or a group of people with a similar condition involved in group therapy. Bearing in mind that the line between curing and healing may be blurred, it is important to remember what the two concepts have in common, especially when translated into practice: improving the well-being of the person (or even persons) in need, noticing and defining their problem, and taking care of them – in short taking some responsibility for the patient's well-being. Without this, one cannot speak of either curing or healing or indeed of medicine.

Summary Points

- The dichotomy of curing and healing relies on the dichotomy between disease and illness.
- It can be argued that the dichotomy of curing and healing is not universally valid as the dichotomy between illness and disease is not universally recognized.
- Presenting curing as rational against irrational healing is a mistake, as medical systems throughout the world rely on varying explanatory models, making different actions rational in different cultural settings.
- The curing and healing dichotomy can be seen as diametrically different in terms of their aims. The first being restorative, while the second transformative.
- The patient is seen as passive in the process of curing and actively engaged in the process of healing.
- The dichotomy of curing and healing is useful at the level of analyzing the relationship between the medical practitioner and the patient, particularly in the Western context.

References

Cassell EJ (1976) The healer's art: a new approach to the doctor-patient relationship. Penguin, New York

Cassell EJ (2004) The nature of suffering and the goals of medicine, 2nd edn. Oxford University Press, New York/Oxford

Cassell EJ (2012) The nature of healing: the modern practice of medicine. Oxford University Press, New York/Oxford

Coward DD, Reed PG (1996) Self-transcendence: a resource for healing at the end of life. Issues Ment Health Nurs 17:275–288

Crandon-Malamud L (1991) From the fat of our souls: social change, political process, and medical pluralism in Bolivia. University of California Press, Berkeley
Dobkin PL (2009) Fostering healing through mindfulness in the context of medical practice. Curr Oncol 16(2):4–6
Egnew TR (2005) The meaning of healing: transcending suffering. Ann Fam Med 3:255–262
Egnew TR (2009) Suffering, meaning, and healing: challenges of contemporary medicine. Ann Fam Med 7(2):170–175
Eisenberg L (1977) Disease and illness: distinctions between professional and popular ideas of sickness. Cult Med Psychiatry 1(1):9–23
Erickson PI (2007) Ethnomedicine. Waveland Press, Long Grove
Finkler K (1994) Sacred healing and biomedicine compared. Med Anthropol Q 8:178–197
Glaister JA (2001) Healing: analysis of the concept. Int J Nurs Pract 7:63–68
Good BJ (1994) Medicine, rationality and experience: an anthropological perspective. Cambridge University Press, Cambridge
Hutchinson TA, Hutchinson N, Arnaert A (2009) Whole person care: encompassing the two faces of medicine. CMAJ Can Med Assoc J 180(8):845–846. PMC. Web. 26 June 2015
Katz R (1982) Boiling energy: community healing among the Kalahari Kung. Harvard University Press, Cambridge
Lerner M (1994) Healing and curing: the starting point for informed choice. In: Choices in healing. MIT Press, Cambridge, MA, pp 13–25
Levine S (1987) Healing into life and death. Anchor, Garden City
Lown B (1999) The lost art of healing. Practicing compassion in medicine. Ballantine Books, New York
McGlone ME (1990) Healing the spirit. Holist Nurs Pract 4(4):77–84
McGuire MB, Kantor D (1998) Ritual healing in suburban America. Rutgers University Press, New Brunswick
Milstein J (2005) A paradigm of integrative care: healing with curing throughout life, "being with" and "doing to". J Perinatol 25:563–568
Mulloney SS, Wells-Federman C (1996) Therapeutic touch: a healing modality. J Cardiovasc Nurs 10:27–49
Rhodes LA (1996) Studying biomedicine as a cultural system. In: Sargent C, Johnson T (eds) Medical anthropology: contemporary theory and method. Praeger, Westport, pp 165–180
Samuel G (1990) Mind, body and culture: anthropology and the biological interface. Cambridge University Press, Cambridge
Scandrett-Hibdon SL, Freel MI (1989) The endogenous healing process: conceptual analysis. J Holist Nurs 7:66–71
Scheper-Hughes N, Lock MM (1987) The mindful body: a prolegomenon to future work in medical anthropology. Med Anthropol Q N Ser 1(1):6–41
Turner V (1967) The forest of symbols. Cornell University Press, Ithaca/London
Vermeylen S, van der Horst D (2007) Beyond monetary valuation: understanding the social properties of property. In: Pertsova CC (ed) Ecological economics research trends. Nova, New York, pp 171–186
Waldram J (2000) The efficacy of traditional medicine: current theoretical and methodological issues. Med Anthropol Q 14(4):603–625

Illness and Its Experience: The Patient Perspective

7

Havi Carel

Contents

Abstract

This chapter offers a philosophical analysis of the illness experience. It uses a phenomenological approach to study the experience of illness and describe its salient features. Using a phenomenological framework, the chapter looks at the physical and social world of the ill person and at changes to self-identity, time, and death. The chapter opens with Toombs' definition of illness as a series of losses. It then turns to examine the experience of illness in terms of symptom experience, diagnosis, disease progression, and prognosis. I use Tolstoy's novella *The Death of Ivan Ilyich* to exemplify the experiential dimension and existential meaning of each stage. I then provide an analysis of the experience of illness by

H. Carel (✉)
Department of Philosophy, University of Bristol, Bristol, UK
e-mail: havi.carel@bristol.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_4

breaking it down into the geography of illness, the social dimension of illness, and the experience of illness as disability, in order to provide an analysis of the first-person experience of illness.

> The body when ill is a concert master not only of pain but of warmth and cold, bloating, pressures, fatigues, nausea, tinglings, itches. (Leder 1990, p. 82)

Introduction

This chapter offers a philosophical analysis of the illness experience. It uses a phenomenological approach to study the experience of illness and describe its salient features. Using a phenomenological framework, the chapter looks at the physical and social world of the ill person and at changes to self-identity, time, and death. Of course illness is a broad category, and its features vary greatly both between and within individuals. Therefore, the chapter also examines the differences between chronic and acute illness, somatic and mental disorder, and congenital and newly diagnosed illness, asking whether the descriptions and analyses offered here capture all, or some, of these subcategories of illness.

The Experience of Illness

Illness as a Series of Losses

A good starting point is looking at S. Kay Toombs' account of illness as a series of losses, which she suggests characterize any illness. She lists five characteristics of illness: the perception of loss of wholeness, loss of certainty, loss of control, loss of freedom to act, and loss of the familiar world. These losses represent the lived experience of illness in its qualitative immediacy and are ones that any patient, in whatever disease state, will experience. The losses cumulatively represent the impact of the illness on the patient's being in the world (1987). Toombs begins with the loss of wholeness. This loss arises from the perception of bodily impairment, which leads to a profound sense of loss of bodily integrity. The body can no longer be taken for granted and can no longer be seen as transparent or absent (cf. Leder 1990), as it assumes an opposing will of its own, which is beyond the control of the self.

The second kind of loss, the loss of certainty, ensues from the loss of wholeness. The patient "is forced to surrender his most cherished assumption, that of his personal indestructibility" (1987, pp. 230–231, 1992, pp. 92–94). This forces the individual to face her own vulnerability. It leads to the third loss – a further heightening of the sense of loss of control caused by the realization that the belief that medical science and technology protect us from the vagaries of ill health is nothing more than an illusion harbored by modern people. In addition, the ill person's ability to make rational choices is eroded because of her lack of medical

knowledge and limited ability to judge whether the health professional professing to heal can in fact do so (1987, pp. 232, 1992, pp. 95–96).

This leads to the fourth kind of loss, the loss of freedom to act. The ill person's ability to choose freely which course of action (which medical treatment) to pursue is restricted by her lack of knowledge of what the best course of action may be. Moreover, in deciding whether to accept medical advice, the patient often assumes that the physician understands and shares her personal value system and takes these values on board when recommending a certain course of action.

Finally, the fifth kind of loss, the loss of the everyday world, arises from the disharmony of illness and it being a distinct mode of being in the world (ibid.). The ill person can no longer continue with normal activities or to participate as before in the world of work and play.

We should start by asking if these losses characterize all types of illness: chronic and acute illness, somatic and mental disorder, and congenital or newly diagnosed illness. It seems that bar a few exceptions, the losses do indeed capture a fundamental experience of illness, in which the ill person feels that something is taken away from her, that at least in a loose sense falls under these five types of losses.

However, if we look outside modern Western culture, we find that illness experiences might be interpreted differently. For example, the twelfth-century nun Hildegard of Bingen suffered from migraines with visual disturbance, which she experienced as religious visions (Sweet 2006). Modern understanding of the oracle of Delphi's prophecies relates her divine inspiration to the inhaling of toxic fumes. And in other cultures, conceptions of the body and of illness include a deep spiritual element (Yoeli-Tlalim 2010). So, illness distinctly aligned with loss on Toombs' account – and more broadly in Western culture – might not be perceived as a loss in other cultures.

In addition, our culture privileges youth and health and perceives illness as a form of weakness or sometimes even personal failure (Ehrenreich 2010). But, other cultures have different values; many traditional cultures value old age and the wisdom and experience associated with it, so they may have a very different attitude to illness and frailty. Toombs' analysis is relevant to a phenomenology of illness but needs to be understood as more relevant to Western modern culture (as Toombs herself does in her later work; see Toombs 1992) rather than revealing universal features of experience.

From Symptom Appearance to Prognosis

How are these losses experienced in illness? Let us think about the course of illness, from appearance of symptoms to diagnosis, disease progression, and prognosis. These phases are not consecutive and may overlap or appear in cyclical form in the case of repeated exacerbations and recovery (e.g., in asthma, allergies, multiple sclerosis). The description that follows is therefore not intended as a series of phases that are independent of one another but as overlapping, often parallel, aspects of illness. Symptoms normally precede a visit to the doctor, sometimes by many months or years, although sometimes diagnosis is made prior to experienced symptoms.

In respiratory illness, for example, diagnosis is made often quite late, because patients are not immediately aware of their symptoms and delay seeing the doctor. By the time respiratory patients become aware of symptoms and consult a health professional, they may have lost 20–50 % of their lung function (GOLD 2010). One explanation for this is that many people do not exert themselves sufficiently in daily living to become aware of increased breathlessness; as a result, by the time a diagnosis is made, the damage to lung tissue and loss of function are usually moderate to severe (ibid.). Symptom appearance and the period before a diagnosis is made can be one of increasing anxiety and sense of abnormality as well as decreased ability. One may experience the symptoms as a loss of freedom and certainty.

Another common experience when new and strange symptoms appear is loss of control. Incontinence, fainting, and vomiting are uncontrollable and may cause extreme embarrassment and grief because they symbolize loss of control more generally. Less dramatic symptoms, like muscular weakness, fatigue, or mild pain, can also give rise to a sense of loss of control. "What is happening to me?" may be a common reaction to a new negative sensation such as pain, breathlessness, or fatigue. Familiar bodily sensations are replaced by alien, negative, bodily feelings experienced as loss of control. This loss is mirrored by the loss of familiarity and wholeness of one's body. Bodily integrity may be suspended or permanently lost when new symptoms appear.

It can be suggested that the loss of freedom – in the broadest sense – is the most prominent loss. More than anything, illness is the loss of opportunities, possibilities, and openness. It is the closure of a previously open future; future possibilities close down as illness progresses. It is also closure of the present: current daily activities lose their casual aspect and become demanding projects. What could once be done unthinkingly and with marginal effort is now an explicit task, requiring thought, attention, and a pronounced effort. The time of symptom appearance, prior to diagnosis but also after, is a time of great change and upsetting of previous life habits. Small things like running for the bus or taking stairs two at a time may become the stuff of fantasy for a respiratory patient. Although minor, they become things the ill person watches others do with awe. Even the envy disappears, which patients initially report they used to feel watching others do things they can no longer do, they are no longer live possibilities for the ill person.

Diagnosis signals a move towards turning symptoms into a less subjective entity. They are now organized in an explanatory pattern that excuses, explains, and predicts illness behavior. In this sense, diagnosis can be experienced as affirmation of subjectively experienced symptoms, making one not "just a complainer" but someone who has a medical condition justifying certain adjustments. For example, many women presenting with breathlessness due to the respiratory condition lymphangioleiomyomatosis (or LAM, which only affects women) are diagnosed as having some form of anxiety, panic attacks, or other psychological disturbance. When the correct diagnosis is made, a woman may feel vindicated that her complaints were not just figments of her imagination but a "real disease."

But the diagnosis also signals an appropriation of one's pain, one's "stomach as painful," by the other's point of view. Sartre writes:

> At this point a new layer of existence appears: we have surpassed the lived pain toward the suffered illness; now we surpass the illness toward the *Disease* [. . .] It is then objectively discernible *for Others*. Others have informed me of it, Others can diagnose it; it is present for Others even though I am not conscious of it. Its true nature is therefore pure and simple *being-for-others*. (2003, pp. 379–380)

So the time of diagnosis can be seen as the time in which the illness (the ill person's subjective experience of her ill body) becomes known by others and by the ill person *as disease* (the objective process causing the illness). It becomes objective (or objectified) and subjected to medical management, labeling, and so on. This movement from a private, subjective experience to an objectified disease, which continues to be experienced as illness by the ill person, is a significant transition.

The illness is no longer a private musing on the nature of bodily change, but an item in a medical vocabulary and ontology, to which shared meanings and knowledge are attached. One's hospital file, pushed around on a little trolley, exemplifies the appropriation of illness by disease. The file contains test results, letters to and from specialists, and requests for further tests, but nothing else. It is a file *about* the patient, but not *of* her. That file symbolizes the subsuming of breathlessness, pain, suffering, social awkwardness, sense of bodily failure, and fear of death, under a medical description. And under that aspect, the lived correlates of the medical information are often relegated to the "subjective-and-hence-secondary" pile.

The Death of Ivan Ilyich

Disease progression is probably the phase at which losses are experienced most acutely. The continuous denigration of freedom is experienced as diminished bodily capacities or increased reliance on medical aids but also as deepening erosion of one's freedom to plan and live. Disease progression is the most intense enactment of our finitude and of our realization not only of mortality but also of our bodily vulnerability and dependence (Carel 2013a; MacIntyre 1999). Here are the famous words of the ill Ivan Ilyich in Tolstoy's novella:

> That Caius – man in the abstract – was mortal, was perfectly correct; but he was not Caius, nor man in the abstract: he had always been a creature quite, quite different from all others [. . .] And Caius was certainly mortal, and it was right for him to die; but for me, little Vanya, Ivan Ilyich, with all my thoughts and emotions – it's a different matter altogether. It cannot be that I ought to die. That would be too terrible. (1995, p. 54)

Disease progression is frightening because it curtails possibilities and also because it is often part of dying. As Ilyich's illness progresses, his experience of pain, exhaustion, and helplessness takes prominence. Later on in the story, he is described thus: "He waited only until Gerassim had gone into the next room and then

restrained himself no longer but wept like a child. He wept at his own helplessness, at the cruelty of man, the cruelty of God, at the absence of God" (ibid., p. 76). It is not only the fact of his death that debilitates Ilyich. It is the realization that his having lived his life as an autonomous, self-sufficient, independent man is peeled away in his dying.

His surrender to his utter dependence on Gerrasim, his servant, and on his doctor to supply him with morphine to alleviate his pain, and his surrender to his own death, is Ilyich's ultimate transformation. He heeds MacIntyre's call for us to acknowledge not only our vulnerabilities and affliction but also our consequent dependence on others, advocating "the virtues of acknowledged dependency" (1999, p. 8). The illusion of autonomy and independence and the misunderstanding of adulthood as encompassing the whole of human life are two errors that lead to a moral view that is inadequate, argues MacIntyre. Ilyich's moral view is transformed through his illness, from someone who is solely interested in doing things *comme il faut*, to authentic conversion. Ilyich's self-understanding and his struggle to resist dependence are given up at the end of the story, replaced by acceptance.

> It occurred to him that what had appeared utterly impossible before – that he had not lived life as he should have done – might after all be true [. . .] And his professional duties, and his ordering of his life, and his family, and all his social and official interests might all have been false. He tried to defend it all to himself. And suddenly he realised the weakness of what he was defending. There was nothing to defend. (ibid., pp. 83–84)

What Ilyich learns through his gradual decline and movement towards death is dependence and humility. He also experiences the losses we started out with: loss of wholeness, certainty, control, freedom, and familiarity. By the end of his life, everything is lost. But Ilyich also learns that life is fragile and precious, that satisfying social expectations amounts to very little, and that he lacks real intimacy.

We can see from this exploration of the phases and losses of illness that illness affects one's entire way of being. Let us now turn to look at these changes in more detail: the analysis begins with changes to the physical world of illness. The chapter then turns to changes to the ill person's social world and psychological changes, which include changes to identity, self-perception, and emotional well-being. The chapter closes with a brief discussion of changes to our experience of time and our attitude towards death. Although the chapter is divided into sections, this by no means suggests that the different life domains are discreet or that changes in one domain do not also imply changes in other life domains, as will be demonstrated below.

The Geography of Illness

In illness, things grow heavier and farther away. A distance an ill person would once call "near" or "an hour walk" is now "far" or "impossible." Small tasks like carrying groceries home, lifting a child, or walking up a flight of stairs require preparation and

rests and may cause excessive fatigue. Everything is hard. Everything is far. Everything is strenuous. The ill person's world, and the world of those who are close to her, shrinks. For chronic patients, the trap is permanent. There is no release from it.

In respiratory illness, for example, the limitation is felt continuously. There is no respite from the exertion (likened to moving around in high altitude) and the breathlessness that accompany everything the respiratory patient does. Being perpetually breathless is, more than anything, uncomfortable. Movements are censored; activities are canceled or crossed out from the list of possible ones.

Spontaneity is lost, like the case of Schneider, the First World War soldier, who suffered a head injury causing him to lose the ability to plan, think abstractly, and fantasize. Merleau-Ponty (2012) describes Schneider's malaise as "existential": he has lost the capacity for spontaneity, for intellectual creativity, and playfulness. Schneider is "'bound' to the actual, and he 'lacks freedom', he lacks the concrete freedom that consists in the general power of placing oneself in a situation" (p. 137). Schneider cannot imagine and therefore cannot execute. Other patients cannot execute but can easily imagine. Creativity can be destroyed in one of two ways: either by removing the capacity to fantasize or by removing the capacity to execute.

The ill body is transformed in a number of ways: it is now experienced through the losses Toombs describes (1987) and as discussed above. In addition, the ill body is experienced explicitly, and often negatively, rather than transparently. The naturalness, if not transparency, that characterizes normal bodily commerce with the world is replaced with artificial and explicit attention to the body. This attention may be related to a medical assessment of the body: "has the cancer progressed?" It is also related to the everyday execution of routine tasks. Explicitness with respect to movement, effort, bodily functions, where toilets or resting places are, and so on is often a part of illness. For example, a diabetic must assess before a meal how much and what they intend to eat and drink. They then need to calculate how much insulin to inject. And they then need to stick to the calculated amount, so spontaneity is lost. The natural way in which we may sample a new kind of chocolate in a tasting stall or pour ourselves a glass of orange juice becomes an orchestrated affair.

The body is also experienced more frequently as an object of medicine and may be further objectified with cumulative exposure to medical examination and treatment. When looking at test results of kidney function or the images from a CT scan or x-ray, one sees one's body as never before. The invisible interior becomes visible, available for one's own scrutiny and an anxious anticipation of the medical pronouncement on one's kidney function or size of tumor.

The body is no longer transparent. Bodily breakdown becomes a common experience (see Carel 2013a). As Merleau-Ponty notes:

> [. . .] the procedures that [illness] employs in order to replace the normal functions that have been destroyed are themselves pathological phenomena. The normal cannot be deduced from the pathological, and deficiencies cannot be deduced from their substitutions, through a mere change of sign. The substitutions must be understood as substitutions, as allusions to a fundamental function that they attempt to replace, but of which they do not give us the direct image. (2012, p. 110)

These disturbances characterize all kinds of illness: chronic and acute, somatic and mental, and congenital or newly diagnosed illness – all disorders that fall under these categories give rise to a change in one's body and one's world. Even mental disorder, which may seem not to affect the body, when studied phenomenologically reveals substantial changes to one's sense of embodiment, bodily possibilities, and bodily feelings (Ratcliffe 2008, 2012; Stanghellini 2004). Even if they are not experienced distinctly as a loss, they still characterize illness in the broadest sense: illness is an alteration of one's bodily experience.

The Social Impact of Illness

How do you introduce yourself to people if you have a serious chronic health condition? What do you say? How can you assuage their discomfort, the sense that you are an alien being, with your wheelchair, insulin injections, oxygen tubes, and complex limitations? How can you carry on being socially "normal" when illness shapes everything you do? How do you handle chance encounters with old acquaintances? These are not, strictly speaking, medical issues, but they are part of what shapes the experience of illness. How the ill person is perceived by strangers, colleagues, and acquaintances will matter greatly to her experience of illness. Stigmatization can be incredibly costly for the stigmatized individual in terms of social relations, but also job prospects, income, and support networks. It is particularly acute in the case of mental disorder, even a common one such as depression (Blease 2012). We should also consider the role of friendship and the strains placed on it by illness. The experience of bodily betrayal and disappointment, the threat illness poses to intimacy, and fear of the diseased body all impact on our relationships.

The transformation is most visible and damaging in the ways it hampers the ill person's social participation, narrows the range of available activities, and makes interactions difficult. The ill person might be unable to participate in social events (e.g., inviting people to dinner if cooking is difficult), may feel awkward around the subject of illness or disability, may fall out of step with healthy people's activities and interests, and may experience a vicious circle of increasing isolation and depression. These alone could cause severe damage to a person's social world. But there are other problems: it is difficult to ask for or know when to offer help; people often experience unease around conversations about illness; and harshest of all, friends may stay away because they do not know what to say (Carel 2008).

Visible illness or disability often becomes the elephant in the room, unless the ill or disabled person, or their interlocutor, actively leads that elephant out. Illness is often seen as something that is not to be commented on or mentioned by polite people, who must not draw attention to it. But at the same time, the medical condition challenges normal interactions and makes "not talking about it" difficult, sometimes impossible. People often feel they ought to say something but are not sure what to say, or how, or when. They feel they should censor their expressions and self-reports, so as not to offend the ill person, but also feel curious, or disgusted, or admiring towards her. The

result is a general sense of discomfort, being ill at ease and unable to transcend the social barrier created by the illness (ibid.).

Illness and its visible signs may arouse fear, disgust, pity, anxiety, or curiosity in healthy onlookers or friends. These emotions may not be consciously experienced and cannot be addressed in a routine exchange. It is difficult to find the right time and words to express these feelings. Ill people frequently report attempts by people to offer encouragement and support, to express admiration and caring. The striking feature of these attempts was how difficult they seemed for the well-intentioned healthy person.

There are additional problems facing an ill or disabled person in their social interactions. There are practical problems, such as being unable to participate in social events such as walking, dancing, or drinking. Everyday activities have to be modified or sometimes given up if the condition does not enable the ill person to take part in them. The ill person can feel she is slowing the others down or hampering the natural flow of events merely by being present. This, in turn, leads her to give up attending some events, and a vicious circle leading to increased isolation may begin.

There are also novel social issues that arise from the illness. For example, the ill person may feel apprehensive about meeting new people because of the awkwardness created by the illness. She may feel the need to explain her condition and go into personal details, but also reluctance to do so. She could feel nervous about leaving the house and going to unknown territory, where the number of steps, wheelchair access, or the location of the nearest toilets is unknown. She might not have the energy to participate in some activities or fear that it would take too much effort for her or that she will embarrass herself by not being able to keep up.

This social architecture of illness mirrors the geography of illness discussed in the previous section. In the same way that distances increase, hills become impossible, and simple tasks become titanic, the freedom to go out into the social world and improvise, to act and interact, is compromised. A new world is created, a world without spontaneity and a world of limitation and fear: a slow, encumbered world to which the ill person must adapt. All people experience this loss of spontaneity through aging. In illness, this opaque and alien world can emerge overnight. This is a world of negotiation, of helplessness, and of avoidance. It is an encounter between a body limited by illness and an environment oblivious to such bodies.

Again, Schneider's existential malaise is mirrored here. Being able to stay up for a night out, to dance, or to just walk from one place to another, to converse and laugh, to have enough energy for talking to people, to stand up talking to people for long periods – all of these abilities might be gone or damaged. The spontaneity that enables social relations to develop is lost or damaged. In other words, the way that the physical and social environment is arranged makes it hard to negotiate while ill. Ill and disabled people invent a myriad of strategies and coping mechanisms to override the constraints inflicted on them by the environment and by the invisible background norms that govern our lives. The demand to be autonomous, independent, and self-sufficient is often met by failure in cases of illness or disability. And perhaps if there was not such a premium on autonomy and independence, the damage to ill people's social life and self-esteem would be lessened (cf. MacIntyre 1999).

These problems lead some ill people to become less sociable and to participate in fewer social events than previously. If we return to the transparency of health mentioned earlier, we can see that the transparency of health is also a social transparency. The transparent and natural way in which we engage in social interactions suddenly becomes cumbersome, weighed down by unspoken doubts and discomfort, and the effort required for genuine communication becomes greater. The social impact of illness is the loss of the transparency and immediacy of social interaction. This transparency of the body, of social ease, can be characterized more generally as a transparency of well-being.

Well-being is the invisible context enabling us to pursue possibilities and engage in projects. It enables us to follow through aims and goals, to act on our desires, and to become who we want to be. But the spatial and temporal possibilities that characterize health are altered in illness, as we saw in the previous section. This is not only the curtailment of spatial possibilities but the abrupt descent of limits onto a world previously larger, freer, and more open. These limits not only restrict physical movement but inflect existential possibilities. It is not only physical possibility that suffers in the hands of illness. It is ways of being and ways of being with that suffer, as discussed in the next section.

Illness as Disability

Illness, and in particular a poor prognosis, can have a deep psychological effect on the ill person and those around them. A distinct feature of the illness is a sense of helplessness, loss of control, and vulnerability. These stem from lost bodily capacities, and the disability and dependence that stem from this loss, but also from the inability to control or stop the disease process or symptoms from exacerbating or stopping the ill person from doing things. The ability to care for oneself, but also the autonomy to make one's way in the world, is seen as a fundamental feature of adult human life. Although this view has been criticized (MacIntyre 1999), we can use this as a starting point of this section's question, namely, what happens to one's life when one becomes unduly restricted by illness?

Heidegger (1962) characterizes human existence as "being able to be" [*Seinkönnen*]. For him, human existence is characterized by its openness, potential, and ability to become this or that thing. This underpins an existentialist picture of human life: one can become what one wants through taking the relevant actions. If one wants to be a polar explorer, one would have to train, build up strength, learn to navigate, and so on. Eventually, one would join a polar expedition and fulfill one's plan of becoming a polar explorer. Of course the obvious physical and temporal limitations would apply, and one would be restricted by his or her "thrownness" [*Geworfenheit*] – being born into a particular culture, historical period, etc. As Dreyfus notes, a contemporary person could not become a medieval samurai warrior, because that option is no longer available in the place and time they are thrown into (Dreyfus 1995).

So although this openness is restricted by common sense limitations, it still characterizes human existence as singular in its freedom, openness, and power of

self-reflection and self-determination. Our plans and aims connect present actions (e.g., studying navigation) to a future view of ourselves as being able to do a particular thing (navigate to the North Pole). Present actions have meaning in virtue of being part of a project that is forward looking and future oriented. I do something now in order to become something in the future. Importantly, I cannot become something in the future by merely thinking about it in the present or wishing for it. I must take relevant concrete actions in order to become what I want to be.

This definition of the human being is best understood by Heidegger's notion of projection. Projection means throwing oneself into a project, which connects the present with the future, and is also informed by the past (thrownness). Projection defines a human being's character and identity. If my project is being a teacher, I project myself accordingly by training to be a teacher, applying for teaching positions, and so on. This, Heidegger claims, is the essence of human existence: the ability to *be* this or another kind of person, to become something, even if this does not ensue from a conscious decision to be this or that kind of person engaged in this or that activity.

This view of the human being as becoming, as pursuing aims, and as constantly molding herself according to the project she pursues is appealing in many ways. It credits us with the freedom – and responsibility – to shape ourselves and our lives in a way we find fulfilling: to transcend our present self with a future self that is more developed, more able. This progressive view of the person sees it as constantly growing and developing, in line with the temporal structure of Dasein (literally "being there," Heidegger's term for a human being) (Heidegger 1962). Dasein "is temporal in the very essence of its Being" (Heidegger 1962, p. 428). And summarizing the temporal structure of Dasein (in its everyday existence), Heidegger defines Dasein as "Being-in-the-world which is falling and disclosed, thrown and projecting, and for which its ownmost ability-to-be [*Seinkönnen*] is an issue" (ibid., p. 225, translation modified, italics removed).

As Merleau-Ponty says, echoing Husserl and Heidegger, being in the world is not a matter of an "I think" but an "I can" (1962, p. 137). The active, goal-pursuing, able Dasein is Heidegger's model of a human being. It is important to note that although the paradigmatic cases of "being able to be" seem to be those of playing a certain social role (mother, head of a tribe, husband) or of pursuing a vocation (polar explorer, teacher), in fact, Heidegger intends to characterize the human way of self-interpretation that informs and orders our activities, rather than an explicit choice of goals and conscious life-planning activity (Dreyfus 1995, p. 95). As Heidegger says, "Dasein has assigned itself to an 'in-order-to', and it has done so in terms of an ability-to-be for the sake of which it itself is – one which it may have seized upon either explicitly or tacitly [. . .]" (1962, p. 119, translation modified).

But in illness, as well as in other situations of dependency, insufficiency, and incapacitation, understanding the human being as "ability to be" does not seem as useful or descriptive. In fact, one's first and final years are usually periods in which one's "ability to be" (in the Heideggerian sense) is restricted in certain ways and dependent on the facilitation of others (MacIntyre 1999). It does not feature the capacity for choice making in the broad, explicit sense (e.g., as in choosing what career to have or whom to marry).

So when thinking about Heidegger's characterization of the human being as "being able to be," we need to consider human life as a whole, including parts of life in which we are unable to do things and eventually completely unable to be (in death). We begin and end in insufficiency, lack of autonomy, and dependence. Heidegger's definition seems to only capture a limited part of human life in at least two senses. First, it only captures the middle part of the trajectory of a human life, excluding infancy, aspects of childhood, and old age.

Second, it only captures the paradigmatic cases of healthy, autonomous adulthood, in which the ability to be is not hampered by disability or illness. But inability to be is a prominent feature of human life. There are many ways in which we are unable to be, or are only partially able to be, or in which being requires extreme effort, that is, cannot be sustained long term. Let us take as an example the case of physical ability to be, say, that of being able to be an athlete. Someone may be (and hence is able to be) an athlete for many years. She exerts herself and suffers in training and sometimes pushes herself right to the edge, but she does not lose bodily control, pass out, harm herself, or experience physical failure.

So we can say that she is able to be an athlete for 20 years. But eventually her body declines, she cannot run as fast or jump as high, and a point comes when she is no longer able to be an athlete. What happens at this juncture to her ability to be? To her self-understanding as an athlete? To her sense of skilful performance, bodily control, and joy in her physicality? These are all radically transformed. However, if we consider the processes of *becoming* able to be an athlete and *no longer* being able to be an athlete, we can see that inability is implied by ability. Being able to be an athlete is always rooted in an organism that starts out and ends as unable to walk, let alone run. Heidegger's discussion excludes both ends of this natural trajectory and thus overlooks an important aspect of life, that of decline, inability, and failure to be.

When ill or aging, we become unable to do some things, perform particular roles, and engage in certain activities. This poses a problem for Heidegger's definition because it shows it excludes important parts of human life and common human situations. In some illnesses, especially mental and chronic illness, a person's ability to be, to exist, is radically changed and sometimes altogether curtailed. For example, in severe psychosis the possibility of having a project at all may become impossible. Similarly, in severe depression the possibility of any goal-oriented action, whether the goal is explicit or tacit, is lost. In less extreme examples of illness, certain projects and ways of being must be discarded, and sometimes a replacement for these is difficult to find.

This raises three questions. First, does Heidegger's account allow radically differing abilities to count as forms of human existence? Second, how flexible are human beings in modifying their projects and goals? And third, with respect to illness, how much can we adjust our projects and plans in the face of ill health, and how should we think of such adjustment?

Heidegger's definition is often understood too literally, but his characterization of existence as "being able to be" can be modified in two ways that would make it able to address cases of illness and other failures. Firstly, the notion of "being able to be"

can be broadened to include radically differing abilities. Secondly, "inability to be" needs to be recognized as a way of being that is other to death (which Heidegger defines as the complete inability to be, or "the possibility of the impossibility of any existence at all" (1962, p. 307, italics removed).

Heidegger's definition can be made more inclusive if we think about "being unable to be" as a form of existence that is worthwhile, challenging, and, most importantly, unavoidable. In order to achieve that, we should interpret the notion of "being able to be" as broadly as possible. It should include cases in which the smooth operation of the body, its compliance in carrying out plans and projects, is no longer there. It should also include cases of prognostic uncertainty or uncertainty about one's ability to pursue a certain goal. And it should also include cases of failure that arises from psychological and social barriers (this is not an exhaustive list).

As happens commonly in illness, current projects may have to be abandoned and new projects created, and these new projects must be thought of in light of new limitations imposed by illness. Such new projects therefore arise within a restricted horizon and are thus different to cases of simple "ability to be," where no unusual or unexpected restrictions limit it. "As long as it is, Dasein always has understood itself and always will understand itself in terms of possibilities," writes Heidegger (1962, p. 185).

But possibilities and their concrete availability to a particular individual are distinctly shaped by gender, race, political climate, and mental and physical disability, to give a few examples. It is naïve to think that most possibilities are not shaped and restricted to an extent by aspects of thrownness. This is said explicitly by Heidegger. But the step he does not take is reconfiguring the notion of "ability to be" in light of these restrictions. I suggest that radically differing abilities, including ones in which possibilities are curtailed, count as abilities to be, even if the freedom is experienced within a context of limitation.

Take a person in a wheelchair, someone with stage IV lung cancer, a person with learning disabilities, or a child with Down's syndrome – all of these are ways of being that differ in significant respects from the paradigm cases of "ability to be." But they are nonetheless human ways of being that contain elements of "ability to be" within a broader context of inability to be. Perhaps the outcome of applying Heidegger's notion of "being able to be" to cases of illness and disability is an acknowledgment of the diverse ways in which it is possible to be and the ways in which human beings differ in abilities and possibilities.

The opposite of being able to be is of course not being able to be, but this presupposes that the notions form a dichotomous system. We can replace this dichotomy with a spectrum of abilities to be. There are other modes of being able to be that are excluded by this dichotomy. Being partially able to be, learning to be able to be, and rehabilitating an ability to be are a few examples. The ability to be that characterizes human existence is territory to be experientially explored and developed, rather than delimited through this opposition.

We can easily find positive examples of this. Stephen Hawking may have wanted to be a footballer, but because of his illness, he was unable to pursue this project. Instead, he had another project, being a physicist, and has become extremely

successful at it. It is true that many projects that might have seemed attractive to him were closed off because of his illness. But even within a contracted horizon of possibilities, there is still an ability to be. We can also think of processes such as rehabilitation from drug use or after an accident; learning to be able to enjoy life after severe depression; being only partially able to walk, hear, see, or talk; and so on. None of these conform to Heidegger's definition in the strict sense, but if we understand ability to be more flexibly, we can include such cases within a broadened Heideggerian account.

Furthermore, in cases of aging, disease, or disability, we need to acknowledge inability to be as a way of being. One way of thinking about aging and illness is as processes of coming to terms with being unable to be; coming to think of one's existence as more reliant and less independent, more interlinked and less autonomous. The inability (or the limited ability) to be and do is the flipside of Heidegger's account. For some individuals, it is there throughout life, as in cases of chronic illness or disability. For all of us, it is present as experiences of inability and failure both in early and late stages in life, in childhood, and in aging. In fact, even the most "able to be" adult life is inevitably sandwiched between these two types of inability: before and after the prime of life. Inability and limitation are part of human life just as ability and freedom are. By introducing the notion of "being unable to be" as an integral part of human life, we can move from seeing ability as positive and desirable to seeing it as part of a broader, more varied flux of life.

Being unable to be is not an independent or context-free concept. It has to be seen in relation to being able to be. An inability to be is a modification of an ability to be that is lost. Being unable to fly, being unable to breathe under water, and so on are not examples of being unable to be. Otherwise, the concept would have nearly endless examples and we would be more unable than able to be if all such cases were taken into account. It is a *lost* ability or an ability that is never achieved, viewed against a background of a common capacity. Being unable to be is therefore intimately linked to an ability to be and vice versa. Being able to be is not unlimited. It is a way of existence that is granted temporarily and is never guaranteed. It is a fragile, transient gift. Considering inability to be is one way of expressing the broader context of being able to be: it is always rooted in inability. Even in cases of extreme physical disability, there is always the possibility of freedom of thought, imagination, emotion, and intellect. Freedom and imagination can enable even those who are unable to be in one way to be in a new way.

Acknowledging an inability and learning to see it as part of life's terrain are important lessons that illness can teach at any age. This knowledge enables the ill person to embrace the unable self as part and parcel of human existence. By having a more balanced view of life and its challenges as interplay of ability and inability, illness can become more accepted and less disruptive. This new understanding of human life as being both able and unable to be paves the way to understanding three important aspects of illness: first is the challenge of acceptance and of living (and living well) with illness. Second, the notion of "inability to be" in its extreme implies the closing down of all possibilities, namely, death, which marks the horizon of illness and all life. Third, the projection into future possibilities against the background of a

past thrownness takes place in time, giving us a uniquely temporal understanding of human life. These three fundamental themes – happiness, time, and death – and their relation to illness are explored in other work (see Carel 2008, 2013b).

Conclusion

This chapter examined the experience of illness as lived by patients. The chapter surveyed the key experiences associated with serious illness and characterized these as a set of losses, following Toombs. The chapter looked at the losses and changes to the lived world of the ill person in its geography, social dimension, and personal experience of loss and "inability to be."

It is important to remember that those who are close to the ill person, e.g., carers, family, friends, and so on, also have an intense and rich experience of the illness as viewed from the second-person perspective. This experience similarly deserves to be studied closely and examined phenomenologically. However, this chapter focused on the first-person experience of illness, in the hope that this will shed some light on an experience that is intense and often life changing. Viewing illness as a restriction and limitation is important as it gives us tools to understand what illness has taken away from the ill person. But it is also important to remember that illness can also be encountered as a challenge and an invitation to reexamine one's life, goals, and values. This chapter provided the foundation for the former, and I discussed the latter elsewhere (Carel 2008, 2014).

Definitions of Key Terms

Phenomenology	A philosophical method used to explore the ways in which embodied consciousness encounters the world and the related acts of consciousness which enable this encounter.
Disability	This term contrasts with Heidegger's definition of existence as "ability to be" (*Seinkönnen*). It delineates a domain in which one is able (one exists) but one is not fully able to be because one is impaired in some way.
Illness	The lived experience of undergoing a disease process. The subjective experience of disease.

Summary Points

- The experience of serious illness is life-changing and dramatic.
- Serious illness changes our global way of being.
- Illness is seen by Toombs as a series of losses.
- Illness changes the ill person's relationship to her environment.

- Illness changes the social world of the ill person.
- Illness can be seen as "disability" – a curtailing of the possibilities of existence.
- Illness changes the ill person's experience of space and time.
- The movement from symptom appearance to diagnosis and disease progression is not linear.
- Illness is analyzed here in deficit terms, but can also be an invitation for human growth and rediscovery of well-being.

References

Blease C (2012) Stigmatising depression: folk theorising and 'the Pollyanna backlash'. In: Carel H, Cooper R (eds) Health, illness and disease: philosophical essays. Acumen, Durham
Carel H (2008) Illness. Acumen, Stocksfield
Carel H (2013a) Bodily doubt. J Conscious Stud 20(7–8):178–197
Carel H (2013b) Ill, but well: a phenomenology of well-being in chronic illness. In: Bickenbach J, Felder F, Schmitz B (eds) Disability and the good human life. Cambridge University Press, Cambridge, pp 243–270
Carel H (2014) The philosophical role of illness. Metaphilosophy 45(1):20–40
Dreyfus H (1995) Being-in-the-world. A commentary on Heidegger's Being and time, division I. MIT Press, London/Cambridge, MA
Ehrenreich B (2010) Smile or die: how positive thinking fooled America and the world. Granta Books, London
GOLD (2010) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (Updated) http://www.goldcopd.org/
Heidegger M (1962 [1927]) Being and time. Blackwell, London
Leder D (1990) The absent body. University of Chicago Press, Chicago
MacIntyre A (1999) Dependent rational animals. Duckworth, London
Merleau-Ponty M (2012 [1945]) Phenomenology of perception. Routledge, New York
Ratcliffe M (2008) Feelings of being: phenomenology, psychiatry and the sense of reality. Oxford University Press, Oxford
Ratcliffe M (2012) Varieties of temporal experience in depression. J Med Philos 37(2):114–138
Sartre JP 2003 [1943]. Being and Nothingness. London & New York: Routledge
Stanghellini G (2004) Disembodied spirits and deanimated bodies. Oxford University Press, Oxford
Sweet V (2006) Rooted in the earth, rooted in the sky: Hildegard of Bingen and premodern medicine. Routledge, New York
Tolstoy L (1995 [1886]) The death of Ivan Ilyich. Penguin Books, London
Toombs SK (1987) The meaning of illness: a phenomenological approach to the patient–physician relationship. J Med Philos 12:219–240
Toombs SK (1992) The meaning of illness. Kluwer, Amsterdam
Yoeli-Tlalim R (2010) Tibetan 'wind' and 'wind' illnesses: towards a multicultural approach to health and illness. Stud Hist Philos Biol Biomed Sci 41(4):318–324

Nursing as Caring

8

Derek Sellman

Contents

Abstract

Nursing is generally considered to appeal to those who wish to care. It is not that doctors do not care but rather that the curative focus that comes with the practice of medicine can be interpreted as emphasizing cure at the expense of care in the provision of health-care services. In this chapter some background is provided regarding the relationship between nurses and doctors before an examination is undertaken of the popular misconception that nurses care while doctors cure. This purported distinction between caring nurses and curing doctors is exposed as relying on assumed gender distinctions and stereotypes regarding what it is that nurses and doctors do in their everyday work. Some discussion of the meaning and nature of care is offered before an outline of the way in which some nursing theorists have adopted the idea of caring for nursing is given. Some non-nursing influences regarding the nature of caring as a response to male-dominated assumptions about the value of caring are noted, and the chapter concludes with the suggestion that attempts to define nursing as caring have met with limited success.

D. Sellman (✉)
Faculty of Nursing, Edmonton Clinic Health Academy, University of Alberta, Edmonton, AB, Canada
e-mail: sellman@ualberta.ca

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_6

Introduction

The idea of nursing as caring suggests that it is the job of nurses to care, while it is the job of doctors to cure, indicating a clear distinction between that which falls within the practice of nurses and that which falls within the practice of doctors, respectively. Yet this idea refers to a largely simplistic portrayal of two complex occupations and of the relationship between them. Such a simplistic portrayal does a disservice to both by minimizing not only the variety of roles each plays in the provision of health care but also by suggesting that nurses have some kind of monopoly on caring while doctors need not – or, in some versions, do not – care. Nurses and doctors have important roles to play within health-care systems, yet it is nurses rather than doctors who are most readily associated with caregiving. This is reflected in everyday descriptive language that likely shapes or is shaped by general understandings of what health-care work involves and what each member of the health-care team actually does. Nurses are generally described by what they do in providing a service, while doctors tend to be described predominantly by the nature of their practice specialty. There is no doubt that nurses are expected to care in ways that doctors are not, but while caring is an important feature of nursing, the idea that nursing can be defined as or by caring is fraught with conceptual difficulties. While everyone thinks they know what it is that nurses do, a universally accepted definition has remained elusive, partly because of the wide range of activities in which nurses partake. What a nurse working in the adult intensive care unit does on a daily basis is so very different from what a nurse working in a mental health facility does that it can sometimes be hard to recognize that both are part of the same professional group. The idea of nursing as caring is attractive because there is a strong, if anecdotal, perception that individuals become nurses because they care and because they want to be of assistance to others. However, it is not just nurses who care and, as will be suggested in this chapter, any attempt to define nursing as caring must take account of the various ways in which the idea of caring is generally understood. One reason that the idea of nursing as caring has purchase is because, just like nursing, caring can be understood in a variety of ways and as such makes it possible for those with diverse perspectives to agree that nursing requires caring without any necessity to agree on what caring requires of nurses.

Background

The language of health care mixes the everyday and the technical in ways that enable differences to be minimized in order that the work of the various players in health-care services can proceed. (Unless otherwise specified, in this chapter "doctor" refers to doctor of medicine, that is, someone who is a registered medical practitioner, and "nurse" refers to someone who holds the legal title of registered nurse.) Without the compromises made possible by this mix of everyday and technical language, the differences between the many professional, semiprofessional, and nonprofessional groups involved in health-care delivery would make it very difficult to get anything done. For example, the tensions between, on the one hand, those who fund and

manage and, on the other, those who deliver health-care services have been a common feature of hospital and community services since their inception. Generally speaking while subscribing to the overall purposes of the provision of health care, the priorities of those who fund and manage health services are not necessarily shared by those who deliver those services. There is nothing new about this, in middle to late nineteenth-century England, the voluntary hospitals were reluctant to divert financial resources toward establishing schools of nursing especially given the hostility expressed by doctors toward the whole idea of professional nursing (Baly 1995). As scientific medicine developed, doctors in hospitals and elsewhere began to realize the value of a trained nursing workforce capable of following orders although it is safe to assume that the idea of an autonomous professional nurse would have been anathema to the vast majority of Victorian medical men – and at that time doctors were, of course, predominately men rather than women. Thus began a complicated relationship between doctors who largely required, or at least expected, nurses to follow doctors' orders and professional nurses who at various times and in various ways have staked claims for autonomous or semiautonomous practice. Remnants of this uneasy relationship persist and are reinforced by the recognition that on the whole it is doctors that diagnose ailments and prescribe treatments (what some might describe as essential acts of curing) while it is nurses who have a role in ensuring those prescribed treatments are given while tending to the physical and emotional needs of patients (what some might describe as essential acts of caring).

As noted, the language of health care is a mixture of the technical and the prosaic. Whether deliberate or not, this mix tends to obscure differences in aims and values as well as in the understandings of ideas, policy edicts, and operational processes of different occupational groupings purportedly working for the same aim: that is, for the benefit of the recipients of health service provision. Thus, while all professional groups can agree on the value of evidence-based practice, those same groups may very well disagree about what counts as evidence, and while the logic of interprofessional working seems unassailable, different professional groups may have quite different expectations of what it means to work interprofessionally: indeed, these terms are often understood in different ways by different individuals within the same professional group. Terms such as "evidence-based practice" and "interprofessional working" thus allow different professional groups with different values and aims to agree at the abstract level about the purposes of their work. And so it is with the term "care." Indeed, the term care permeates the very language of services for health as seen in the adoption of the broad umbrella term "health care" which has become ubiquitous in its application to all things related to the provision of health services, although in the UK the term has been extended by the addition of social, as in "health and social care." And within health-care services, doctors practice medicine, while nurses practice nursing. It may be here, in the language used to describe the work of nurses vis–àvis doctors, that the idea of nursing as caring has emerged. Nurses give nursing care and this caregiving has been commonly categorized by medical specialty – described, for example, as nursing care of the surgical patient, nursing care of the neurological patient, and so on – although some current textbooks for nurses seem to be deliberately avoiding the terminology of "nursing care" or even "care" (see, e.g., Day et al. 2010;

Potter et al. 2014), while others seem content to retain its use unapologetically (see, e.g., London et al. 2011). In contrast, the work of doctors tends to be described in terms of their practice specialty; doctors do not give surgical care, they are surgeons; they do not give neurological care, they are neurologists; and so on. This language of description lends itself to the idea that care is more associated with what nurses do, rather than cure which is usually associated with the work of doctors.

However, this purported cure versus care distinction between what it is that doctors and nurses do betrays a simplistic perception of the nature of professional work for health and of the way illness is understood and dealt with in modern Western society. It assumes the idea of a typical patient as a youngish, generally healthy, active, and productive individual with a condition that once cured will render that person fit to return to their regular healthy state and carry on much as before – just as might be the case for a broken washing machine or lawn mower taken in for repair and returned in full working order. While this way of thinking about health and illness works for a few, predominately acute, conditions (e.g., a limb broken in a skiing incident or a bad case of holiday food poisoning), it reflects only a small amount of what it is that work for health involves and thus work that only a relatively small number of doctors and nurses do in their everyday clinical environment. While work with acute conditions with its attendant requirement for typically high-tech, often expensive, and usually dramatic interventions made by doctors occupies the public imagination fuelled in large measure by television and film portrayals, by far the majority of doctors and nurses work with those whose health is affected by long-term rather than acute conditions or by the long-term effects of acute conditions. This less glamorous aspect of health-care provision requires not so much cure as containment together with a provision for long-term care (what might be described as the management of long-term conditions). Certainly it is true that on the whole the former (containment) will most likely require doctor-only prescriptions and treatments to be issued and administered, but the monitoring of such prescribed treatments is more accurately described as care insofar as it is part of providing for people with long-term (formerly, chronic) conditions. Thus, care, rather than cure, in this sense is a feature of the work of many doctors and belies the idea that cure rather than care is what doctors do. This is well illustrated in palliative medicine where doctors do not seek to cure in any traditional sense, but act so as to assist those who are dying to live well (RCP, nd). Many other specialties require that doctors think in terms of long-term care rather than merely in terms of cure; geriatrics, nephrology, neurology, and oncology are but a few of the specialties where curing, such as it is, invariably requires elements of long-term care. And while a few doctors can restrict their practice solely to the curative aspects of medicine and surgery, many others will be involved in the provision of long-term care.

Just as care is not the sole province of nurses, so too cure is not the sole province of doctors. Historically, nurses have tended to take on roles previously restricted to doctors. The measurement of blood pressure, venipuncture, the administration of intravenous medications, the ordering of diagnostic tests, and the prescribing of drugs have all, albeit at different times and in different places, been doctor-only features of health-care practice. Yet now, and depending on particular jurisdictions,

these things are either essential aspects of the general scope of nursing practice (that which determines what nurses are and are not permitted to do) or have become routine for a few, appropriately prepared, nurses. Indeed, nurse prescribing represents one feature of what hitherto has been, and still remains largely, understood as something that doctors, not nurses, do.

Tradition has it that doctors are men and nurses are women. This assumed "natural" gender divide is most predominant in Victorian times in Europe and in England and the legacy of this idea continues to color perceptions regarding the roles and gender appropriate to doctors and nurses with all the stereotypes regarding the relative authority of each and the power relationships between them. There are several versions of the history of nursing, but the general perception is that modern nursing began with Florence Nightingale who understood the role of nursing to be one suitable for women – most definitely not for men – for it is women who care: "Every woman or at least almost every woman has, at one time or another of her life, charge of the personal health of somebody, whether child or invalid – in other words, every woman is a nurse" (Nightingale 1980[1859], p. v) and who was of the opinion that women who wanted to become doctors were misguided:

> Instead of wishing to see more doctors made by women joining what there are, I wish to see as few doctors, either male or female, as possible. For, mark you, the women have made no improvement – they have only tried to be men and they have only succeeded in being third-rate men. (Nightingale 1860)

But nursing can trace a longer history with both men and women taking on the role of tending for the sick at different historical periods. Female matrons were available to the Romans, but it was male attendants who provided nursing services to the Crusaders and subsequently to soldiers in other military conflicts right up until the time of the arrival of female nurses during the Crimean War. Similarly, in New France on the North American continent in the seventeenth century, it was the missionary Jesuit priests who first provided nursing for settlers before the arrival of nuns who took over that role (MacPhail 1996). Indeed, it might be that it was the involvement of the women of religious orders in providing care for the sick that so fixed the idea that nursing work is women's work. At the time of writing, 44 % of those on the UK General Medical Council's list of registered medical practitioners are women (GMC 2014), while men account for approximately 10 % of those on the UK register of nurses (NMC nd), and, assuming these figures are reflected in other jurisdictions, this suggests that while medicine is no longer considered suitable only for men, nursing remains predominately a job for women.

The Nature of Care

The word care has many applications in both everyday and technical language. To say that a person takes care might indicate a number of things. It might be taken to mean, on the one hand, that she or he is meticulous and thorough or, on the other

hand, cautious and circumspect. Usually context will provide the necessary clue as to what the speaker intends: so, to say that a person takes care of a neighbor's cat would suggest some version of the former, while to say that a person takes care in crossing the road would normally indicate the latter. These two senses are not, of course, mutually exclusive nor do they necessarily entail that either person actually cares for or cares about anything in particular. Similarly, if it is said that person A provides care to person B, there is no necessary implication that A actually cares for or about B. Indeed, the general understanding regarding what it means to provide health care lacks any notion that one party necessarily cares for or about the other. That government might direct or regulate health-care activities does not imply that government cares about individuals either in general or in specific instances – and many would argue that it is not the place of government to care and the same can be said of nongovernmental providers of health care. Providing health care in this sense then implies no more than providing a service just like any other service – say utilities or highways.

However, when it is said that a nurse provides care, this is usually taken to imply more than the simple provision of a service. The person who takes care of a neighbor's cat would not be censured for merely looking after the animal, that is, making sure the pet is fed, that the litter tray is regularly cleaned, and that the feline does not escape and/or come to harm. There is no necessary requirement for the person taking care of the neighbor's cat to care beyond such minimal requirements, albeit that the neighbor might prefer something more. Yet when it is said that a nurse is to take care of a patient, or to provide care for a patient, or to provide nursing care for a patient, or simply to care for a patient, such a minimal formulation would be considered insufficient either from a professional or, at least it is assumed for most nurses, a personal perspective. All these "looking-after"-type things are indeed required, but the expectations of what a nurse should do when they provide care are of a higher order. And here is another set of ideas from which the idea of nursing as caring emerges. Nurses, it seems, are required both to care for (i.e., to give care in the sense of looking after) and to care about (i.e., to have a genuine interest in the well-being of the person in receipt of care). A distinction between caring for and caring about is commonly made yet not always in the same fashion. Noddings (2013 [1984]), for example, in her book length exposition on the nature of caring, reverses that distinction by drawing our attention to what she holds as common usage in that we "car[e] … about things and ideas" (Noddings 2013[1984], p. 21) but care for persons. Indeed, for Noddings caring for others, as well as for oneself, lies at the heart of what it means to be human and it should be no surprise that nursing, often considered an occupation most suited to women, has assumed a similar although not always well-articulated position. Nursing as caring would seem then to be a logical extension of the idea of nursing as women's work for women are expected to care in ways that men are not. Women are considered natural caregivers. It is women who give birth to and who care for their children, it is women who are expected to forgo career development in order to care for their families and to prioritize the needs of their husbands and children, and it is most often women who become carers for older relatives when those relatives can no longer care for themselves (Duxbury et al. 2009). This assumed natural caring nature of women follows women into the

predominately female profession of nursing. Indeed, Noddings' claim that our caring nature arises out of our experience of being cared for as infants reinforces the association between women and caring although she attempts to distance herself from any strong version of this idea by indicating that while caring is a feminine trait, it is not, nor should it be considered to be, manifested exclusively by women.

On this account, nursing as caring seems to many to follow the purported natural order of things. It is thus supposed that women do the caring while men do the curing, that women are subjective and emotional while men are objective and rational, and that the latter is somehow better than the former. This view, which might be described as fairly typical of the Victorian colonialists, continues to hold influence over the general opinion of not only what it is that nurses do but also the value of what they do. Caring is considered of less value and of a lower status than curing, while rationality is considered a higher form of psychological development than emotionality. Given the association of caring with nursing and given the emotional and physical labor that nurses invest into the role of the nurse, it should be no surprise to find that nurses tend toward a general dissatisfaction with accounts that emphasize cure and rationality at the expense of care and emotional engagement. It should be no surprise also then that nurses have been drawn toward accounts of nursing and beyond that seek to valorize caring.

In accounts of nursing, caring is often considered to be a central feature and, moreover, an important and valuable feature. Indeed, many accounts of nursing stress the importance of caring either as a response to the perceived lack of caring in systems of health-care provision dominated by rationality or as an attempt to reverse the general view that caring matters less than curing.

In the USA, the idea of nursing theory took hold in the 1950s and most theories of nursing were published between then and the 1990s since when refinement rather than new theory has been more common. In theories of nursing implicit and sometimes explicit notions of the nature of nursing, and thus the nature of nursing care, are largely determined by the assumptions made by the theorist about what it means to be a human being. (A word of warning: the view that nurses in the USA hold regarding the need for and value of theories of nursing is not universally shared as necessary or even desirable – nevertheless, the influence of American theoretical perspectives on nursing has to a large extent given impetus to the idea of nursing as caring.) Care, or sometimes caring, is sometimes more, sometimes less to the fore in these theories, and, generally speaking, there is a tendency for the theories to increasingly associate the act of nursing with the act of caring. Orem's theory of self-care and self-care deficits was perhaps one of the first to include care as a central feature. Originally developed as a conceptual framework for nursing in 1959 and subsequently refined over a period of 40 years (Coldwell Foster 2010), Orem argued that nursing care is required when a person is no longer able to provide care for herself or himself. For Orem then, nursing care seems to be task orientated insofar as it requires nurses to do for patients that which they cannot do for themselves. In this respect Orem would seem to require that nurses care for, rather than about, patients.

Caring for persons in the context of their culture forms the basis of Leininger's theoretical approach from which the idea of transcultural nursing has emerged

(George 2010). Once more, care has a central feature and for Leininger, caring requires the nurse to acknowledge the values and beliefs of a patient in relation to their cultural grouping and then to act in ways that respect the norms and expectations of that patient's culture (Leininger 1991). In other words, caring nurses will not use their own sets of values to judge patients but will be open to valuing whatever it is the patient values. It requires nurses to respect the traditional care practices of whatever cultural group the patient identifies with and it seems to associate care and caring practices with activities that are consistent with the beliefs and values of the patient. This approach has had an influence beyond acceptance of the theory itself – and may be an important antecedent of other generally accepted ideas within nursing regarding the idea of nursing as caring such as, for example, holistic care – although the underlying relativism would seem to present some difficulties regarding how distinctions are to be made between acceptable and unacceptable cultural practices. The theory also assumes that within cultures, humans care for, as well as about, one another, an assertion that can be difficult to sustain in the light of some specific cultural practices that may pose significant moral challenges to nurses including but not limited to the practice of female genital mutilation.

Perhaps the nursing theorist most often associated with nursing as caring is Jean Watson. Her theory of transpersonal caring (Watson 1979, 1988) places care firmly as the central feature of nursing practice. According to Watson, nursing has lost its way as part of the general modernist/technical-rational dominance in health care and needs to reestablish its role regarding human healing aspects of nursing. In emphasizing a strong relationship between nursing and caring in the sense of healing, Watson goes further than most other nurse theorists in distinguishing nursing from medicine. To the extent that these ideas resonate with the views of many nurses, Watson's writings to this point are well known and popular. However, the extent to which those same nurses accept the full requirements of her theory remains unclear as Watson makes strong claims regarding the requirement for nurses to enter into intense relationships with patients for the purpose, among other things, of healing the soul of the patient. For Edwards (2001) this requirement is neither reasonable nor desirable as it places excessive demands on both nurses and patients and seems to require the ontological impossibility of experiencing things as another person experiences them. Watson's emphasis on healing leads her toward a spiritualism and mysticism that remains unattractive to nurses unconvinced by the vision of nursing as an all-absorbing form of work. For such nurses, caring is a professional obligation to be distinguished from their personal lives. Thus, Watson's strong version of nursing as caring cannot be said to be universally accepted or acceptable.

No discussion of the idea of nursing as caring would be complete without reference to the work of Patricia Benner. Perhaps most famous for her book *From Novice to Expert* (Benner 1984), it is the title of a later, jointly authored, book *The Primacy of Caring* (Benner and Wrubel 1989) that illustrates the central position she holds for caring in the practice of nursing. Following Heidegger, Benner and Wrubel note that care is part of the very nature of what it is to be human extending this to nursing by claiming nursing to be a "caring practice" (Benner and Wrubel 1989, p. xi) and by, for example, identifying the need for nurses to take into account what it

is that any given patient cares about before giving care: although it is not clear that recourse to Heidegger is necessary to support this requirement. In other words, for Benner and Wrubel, expert nursing requires an understanding of the meaning to a patient of whatever ailment it is they are experiencing. This requires not merely the application of routine nursing care for any particular condition because what that condition may mean in the life of one individual may have a very different meaning for another person, thus requiring the nurse to respond to the individual and not the disease. This approach, long understood as a move toward individualized nursing care, presupposes a previous regimen in which the responses of nurses to the individual needs of a patient with a given illness were considered less important than responding to the condition itself. In nursing this has often been characterized as being of, for example, "the appendix in bed 6" approach to nursing care purported to be symptomatic of the biomedical model of health and illness so reviled by many in nursing. In this way, Benner can be read as attempting a technical definition of caring for the purpose of distinguishing nursing from medicine while retaining everyday understandings of the nature of care. The success or otherwise of this maneuver remains to be seen for it is not clear why caring in Benner's technical sense cannot be invoked for other health-care, or even non-health-care-related, occupations. In other words Benner's account does not explain nursing as caring so much as reiterate the importance of caring in nursing.

Perhaps a reflection of the increasing rejection by women of the subservient role allocated to them by repressive societal expectations and presumably as a part of a growing feminist movement in the Western world, Carol Gilligan (1993[1982]) and Nel Noddings (2013[1984]) each in their own way set out alternatives to male-dominated ways of thinking about care and caring, and both have been influential in the way that nurses perceive the value of what they do. Gilligan, often cast as a reaction to the male-dominated ideas on moral development offered by Kohlberg, set out to study moral decision-making in action. In attempting to understand how women made decisions regarding whether or not to have an abortion, she found that the women in her study approached moral decision-making in ways quite different to those described in Kohlberg's theory. Kohlberg's theoretical schema of stages of moral development with abstract notions of justice at the pinnacle was based on the responses of young males to a set of hypothetical situation questions (Kohlberg 1984). In the real-life situations of women trying to decide whether or not to have an abortion, Gilligan found little resemblance among her respondents to the claims of Kohlberg's theory. Rather she found that women were more concerned with the effects of their decisions on the myriad and often complicated relationships around them. For these women abstract principles were less important than caring relationships in making moral decisions, and rather than reflecting an inferior level of moral development as Kohlberg's theory would have us believe, Gilligan suggested that women approached moral decision-making in different, but no less important or refined, ways to men. Previously a research assistant to Kohlberg, Gilligan outlined her rejection of Kohlberg's theory in her influential book *In a Different Voice*, first published in 1982 but reflecting work started at the beginning of the 1970s. The ideas in her book resonated with many nurses whose own experience

of being rarely heard was a source of increasing frustration as the abstract principle-based approach to health-care ethics in general and to medical ethics in particular was fast becoming entrenched as dominant. Credited as one of the influential texts in the rise of care-based ethics as a viable alternative to principlism and taken up with some enthusiasm in nursing, care-based ethics has, in some places, had some influence on mainstream medical ethics. However, because principlism remains such a strong feature of medical ethics and because of the popularity of care-based ethics among nurses, there is a danger that these preferences are seen as merely reflecting and reinforcing the gender and role stereotypes that associate doctors with masculine traits of abstract rationality in the pursuit of cure and nurses with feminine emotional traits of caring.

Conclusion

Nursing's attempt to appropriate caring as a term that extends beyond everyday notions of what it means to care is fraught with difficulties precisely because caring is a term that has a variety of technical and everyday meanings largely determined by context. Within the health-care context, caring continues to be understood in different ways even as some nurses attempt to claim caring has a special place in, and a particular relevance to, nursing. It is difficult to imagine any definition of caring that marks it as unique to nursing without doing violence to general nontechnical and everyday understandings of the term. Even if there were to be a technical sense in which caring might be claimed as nursing's own, it is difficult to see how such a definition could withstand attempts by other professional groups to claim similar ground as part of their work. Social workers, doctors, psychotherapists, occupational therapists, teachers, and many others may also advance strong claims in regard to the caring nature of their work. Nor is it easy to imagine why a technical definition of caring for nursing would want to exclude other caring relationships such as those between a parent and child. For nursing to disregard the claims regarding caring in the context of other professional groups or the everyday understandings of caring within family relationships indicates a preference for placing professional interests ahead of the interests of those whom professional nurses profess to serve.

As Edwards (2001) notes, adopting the idea of nursing as caring requires accounts of caring that "avoid being so weak as to allow the kinds of caring acts appropriate for inanimate objects. But they must also avoid being excessively demanding, or they become either professionally inappropriate, ontologically impossible, or practically implausible" (Edwards 2001, p. 114).

Edwards goes on to suggest a modest proposal of what he describes as "intentional caring" as requiring three parts:

(a) Sympathetic awareness of the plight of the patient
(b) An emotional component – given acceptance of the existence of an emotional dimension to human being
(c) A response to what are perceived to be the needs of the patient

(Edwards 2001, p. 132)

This modest proposal for the nature of professional caring as it applies to nursing seems to be uncontroversial although, by design, it neither makes caring a unique feature of nursing nor does it exclude other professional groups from claiming similar ground. For this reason, it is unlikely to satisfy those with a vested interest in promoting nursing as caring in the sense that it distinguishes nursing from other health-care practices.

Definitions of Key Terms

Care/caring: Care is variously understood. For example, care is employed in a technical sense that equates with, among other things, the provision of services such as health care or nursing care; it is employed in the everyday vernacular in ways that require context to provide meaning: 'caring for' is often differentiated from 'caring about'; and some nursing theorists have attempted to adopt the idea of care or caring as phenomenon specific to nursing.

Summary Points

- Caring is used in both technical and nontechnical ways within health services language which makes it difficult to be clear exactly what is meant by the term.
- Care as women's work versus cure as men's work is a misleading way to try to differentiate the role of nurses compared to the role of doctors.
- Gender assumptions regarding nurses and doctors reinforce stereotypical understandings of nursing and medicine to the detriment of both.
- Nursing theorists emphasize the central importance of caring to nursing.
- Nursing as caring requires a definition of caring that does justice to everyday understandings of what it means to care.

References

Baly ME (1995) Nursing and social change, 3rd edn. Routledge, London

Benner P (1984) From novice to expert: excellence and power in clinical nursing practice. Addison-Wesley, Menlo Park

Benner P, Wrubel J (1989) The primacy of caring: stress and coping in health and illness. Addison-Wesley, Menlo Park

Coldwell Foster P (2010) Self-care deficit nursing theory: Dorothea E Orem. In: George JB (ed) Nursing theories: the base for professional nursing practice, 6th edn. Pearson, Boston, pp 113–145

Day RA, Paul P, Williams B, Smeltzer SC, Bare B (2010) Brunner & Suddath's textbook of Canadian medical-surgical nursing, 2nd edn. Wolters Kluwer Lippincott Williams & Wilkins, Philadelphia

Duxbury L, Higgins C, Schroeder B (2009) Balancing paid work and family caregiving responsibilities: a closer look at family caregivers in Canada. Human Resources and Skills Development Canada, Ottawa

Edwards SD (2001) Philosophy of nursing: an introduction. Palgrave Macmillan, Basingstoke

George JB (2010) Theory of culture care diversity and universality: Madeleine M Leininger. In: George JB (ed) Nursing theories: the base for professional nursing practice, 6th edn. Pearson, Boston, pp 404–434

Gilligan C (1993[1982]) In a different voice: psychological theory and women's development. Harvard University Press, Cambridge, MA

GMC (General Medical Council) (2014) List of registered medical practitioners – statistics. http://www.gmc-uk.org/doctors/register/search_stats.asp. Accessed 25 July 2014

Kohlberg L (1984) The psychology of moral development: the nature and validity of moral stages. Harper & Row, San Francisco

Leininger MM (1991) Culture care diversity and universality: a theory of nursing. National League for Nursing, New York

London ML, Ladewig PW, Ball JW, Binder RC, Cowen KJ (2011) Maternal and child nursing care. Pearson, New York

MacPhail J (1996) Men in nursing. In: Ross Kerr J, MacPhail J (eds) Canadian nursing: issues and perspectives, 3rd edn. Mosby, St Louis, pp 74–81

Nightingale F (1860) Letter to John Stuart Mill (12 September 1860), published in McDonald L (ed) (2003) Florence Nightingale on society and politics, philosophy, science, education and literature. Wilfred Laurier University Press, Waterloo

Nightingale F (1980[1859]) Notes on nursing: what it is and what it is not. Churchill Livingstone, Edinburgh. First published in 1859 by Harrison & Sons

NMC (Nursing and Midwifery Council) (nd) The nursing and midwifery council statistical analysis of the register 1 April 2006 to 31 March 2007. http://www.nmc-uk.org/Documents/Statistical%20analysis%20of%20the%20register/NMC-Statistical-analysis-of-the-register-202006-202007.pdf. Accessed 25 July 2014

Noddings N (2013[1984]) Caring: a relational approach to ethics and moral education, 2nd edn, updated. University of California Press, Berkley

Potter PA, Perry AG, Stockert PA, Hall AM, Ross-Kerr JC, Wood MJ, Astle BJ, Duggleby W (2014) Canadian fundamentals of nursing, 5th edn. Elsevier Canada, Toronto

RCP (Royal College of Physicians) (nd) Palliative medicine. https://www.rcplondon.ac.uk/specialty/palliative-medicine. Accessed 28 July 2014

Watson J (1979) Nursing: the philosophy and science of caring. Little Brown, Boston

Watson J (1988) Nursing: human science and human care: a theory of nursing. National League for Nursing, New York

Goals of Medicine

9

Thomas Schramme

Contents

Abstract

This chapter discusses different philosophical theories regarding the goals of medicine and places this debate within the context of the moral limits of the proper use of medical means. Two approaches are distinguished: first, a teleological approach, which sees medicine as a practice with an inherent telos and second, a consensual approach, which aims at assembling a list of goals of medicine that are identified in a deliberative process. This chapter also discusses the concept of medicine and scrutinizes whether it has any bearing on the debate regarding the goals of medicine. It is argued that the goals of medicine are still contested and will probably remain so. They cannot be used in a direct way to solve normative questions regarding the proper use of medicine.

T. Schramme (✉)
Department of Philosophy, University of Liverpool, Liverpool, UK
e-mail: t.schramme@liverpool.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_5

Introduction

Medicine is both a theoretical and a practical endeavor. Occasionally, medicine is referred to as a science and an art. Considered as a practical endeavor, or art, medicine apparently aims at particular goals, briefly the treatment of disease and restoration of health. In addition, medicine has become a service institution in many countries. Medical means are used for ends other than just treatment of disease or restoration of health. Enhancements of specific desired features, improvements of fitness, and so on, as well as treatment of non-pathological conditions, such as short stature or normal deterioration of performance due to aging, have become targets of medical intervention. For some philosophers of medicine, but also for many practitioners and citizens, this expansion of the remit of medicine is a worrying development, sometimes called medicalization. The term is usually used with a negative connotation and refers to an undesirable use of medical means to tackle social and individual problems or desires. Medicalization may even involve an usurpation of traditional ways of solving problems in living, for instance, when people experiencing unbearable working conditions take antidepressants instead of challenging their environment. On the other hand, medicine does not seem to have an imbedded scope of proper use, which would speak against, or even disallow, its employment for other than individual diseases.

Some philosophers of medicine have criticized medicalization and the use of medical means for aiming at desired conditions by arguing that medicine has particular intrinsic goals, which restrict the proper use of medical means to the pursuit of these goals (Pellegrino 2001; cf. Arras 2001; Veatch 2001). According to such a view, the goals of medicine, such as treatment of disease and relief of suffering, are intrinsic goals insofar as they are implied by the practice itself. This is partly a traditional argumentation, going back as far as ancient philosophical ideas about actions and practices. Every action seems to aim at a goal; otherwise, we would perhaps not even call it an action. Practices are iterated actions and also seem goal oriented. They have, in philosophical terms, a telos, or a telic structure. For the argument about the goals of medicine, this general idea is important, as it paves the ground for assuming particular goals of the practice of medicine. It is more difficult, though, to establish these goals as intrinsic to the practice of medicine itself. Such reasoning seems to rely on an assumption of a particular nature or essence of medicine.

A prominent philosopher of medicine, who has argued the case for intrinsic goals of medicine, is Edmund Pellegrino. Alternatively, a task force at the Hastings Center, which is a leading bioethics academic institution, published a report in 1996 and determined such goals of medicine in a process of identifying an international consensus. Although there might not be huge differences in terms of the mentioned goals of medicine between a teleological and a consensual approach – prevention of disease and avoidance of premature death were indeed items on the Hastings Center list – there are notable methodological differences. A consensual approach allows for the goals of medicine to change historically and socially, whereas a teleological approach aims at a universal and nonrelative determination of the proper goals of

medicine. In the following, a closer analysis of these two approaches will be pursued.

The Concept of Medicine

Discussion of the goals of medicine relies on a particular conception of medicine. But it is not quite clear what medicine encompasses or how to define the nature of medicine (Pellegrino and Thomasma 1981; Nordenfelt 1998). At the beginning of this chapter, it was stated that medicine is considered to be a science and an art. In other words, medicine has a theoretical part, which has mostly to do with gaining knowledge about the functions and dysfunctions of the organism. Medicine, understood as an art, is the application of such knowledge in specific contexts, such as diagnosis, treatment, or prognosis. So the extension of the concept of medicine seems to be fairly broad, as many scientific endeavors and also quite a few practices seem to be aspects of medicine. In addition, there are other terms, which are occasionally used synonymously with the concept of medicine, such as "health care." Health care includes practices such as nursing or rehabilitation. Finally, there is the discipline of public health which includes practices that aim at the health of the population. It uses the science of epidemiology, and different practical means, such as information or health education, policies, or the intentional shaping of the circumstances of people's choices. One might wonder whether public health is part of medicine or whether it has a broader remit. In any event, it seems that there are indeed many practices that aim at health and the prevention of disease, including medicine, health care, and public health. Even if some of these practices are not to be counted as medicine proper, it seems obvious that the goals of aiming at health and preventing disease are not restricted to medicine. So there is a problem for determining the concept of medicine by reference to its alleged goals (Nordin 1999). "Medicine" does not have clear-cut boundaries, and it is impossible to conceptually separate medicine from other practices by referring to its alleged goals, because these goals are shared with several other disciplines.

Another way to discuss the goals of medicine in relation to the concept of medicine might be to focus on the means of the practice. It might be said, for instance, that public health, in contrast to medicine properly conceived, uses political and pedagogical means, whereas medicine uses certain skills of doctors, communication, and diagnostic tools. In general, one might want to restrict medicine to the clinical encounter between a patient and a doctor (Cassell 1991). The goals of medicine, according to this point of view, are identical to the goals of treatment or care (Marcum 2008; Kaldjian 2014). As will be seen, especially the philosophical approach that aims at extracting goals of medicine from its practice, or rather application, is prone to such a view. Yet, again, it is not clear why the concept of medicine should be restricted to the clinical encounter. At least historically, there have been examples of other conceptualizations. For instance, Galen defined medicine by its aspect of gaining knowledge about the organism. There were also attempts to explicitly exclude any therapeutic intervention from medicine, as it

was deemed to do more harm than nonintervention (Temkin 1966). Accordingly, the group of people called "medical nihilists" by the historian Owsei Temkin, saw medicine as essentially including science. Medicine therefore was not the exclusive domain of physicians.

Altogether, the discussion of this section undermines a straightforward definition of the concept of medicine. Since the nature of medicine is hence underdetermined, it cannot simply be assumed that medicine is a specific discipline or practice with clear-cut goals.

A Teleological Approach

One way to determine goals of medicine is by interpreting it as a practice, which is structured by aiming in a certain direction. This is a traditional idea that goes back at least to ancient philosophy. It has followers especially in modern virtue theory. Here, the aims of practices also specify certain excellences or virtues. It should be noted that the concept of practice here refers quite generally to types of actions, not necessarily to the use of tools or something similar. Gaining knowledge in a scientific endeavor can be a practice, according to this understanding. In a teleological approach, the telos, or end, of a practice determines the good it aims at. Virtues are accordingly the excellent ways to perform such practices. When discussing the goals of medicine, such an account requires some idea of the specific goods which medicine aims at. An obvious goal of medicine is health.

A well-known defender of such a teleological account of the goals of medicine is Edmund Pellegrino. He claims: "[W]e must assert the obvious: medicine exists because humans become sick. It is an activity conceived to attain the overall end of coping with the individual and social experience of disordered health. Its end is to heal, help, care and cure, to prevent illness, and cultivate health" (Pellegrino 1999, p. 62).

It should be noted that Pellegrino allows for some level of change in using medicine for specific social purposes. Yet these purposes always need to be linked to the inherent ends of medicine (Pellegrino 1999, p. 65 f.). Hence there is no scope for taking medicine outside its proper remit, which is intrinsically set. The ends of medicine are determined by the practice of medicine, and these ends are essentially focused on sick patients.

Similarly, Leon Kass also maintains that there are proper goals of medicine. These set the norm as to how medical means are properly used. "I am rather inclined to the old-fashioned view that health – or if you prefer, the healthy human being – is the end of the physician's art. That health is *a* goal of medicine few would deny. The trouble is, so I am told, that health is not the only possible and reasonable goal of medicine, since there are other prizes for which medical technique can be put in harness. Yet I regard these other goals – even where I accept their goodness as goals – as false goals for medicine, and their pursuit as perversions of the art" (Kass 1985, p. 159).

As has been discussed earlier, it is not quite obvious that Pellegrino and Kass can make good their claim regarding proper goals of medicine. It is not even clear how exactly to draw the boundaries of the practice called medicine. In addition one might wonder in what way the specific goals of medicine and with it the assumption of a teleological structure of the practice can be philosophically justified, especially given historical variations.

As has been explained in the introduction of this chapter, reference to the alleged goals of medicine can often be found in contexts where certain contested ways of using medical means are being discussed. However, it does not seem easy, and perhaps impossible, to circumvent the normative debate about the proper use of medical means by a philosophical account of the proper goals of medicine. Indeed, it might not even be altogether obvious that within Pellegrino's approach all real developments in modern societies that can be summarized under the label of medicalization would be identified as improper uses of medical means. After all, the cultivation of health, for instance, might be understood as to imply an increasing societal demand for fitness and capacity to perform, which, again, could well be fostered by medical means. Yet it is clear that Pellegrino and Kass see their approach as a bulwark against modern developments of using medical means for purposes, which are alien to medicine proper according to their point of view. Still, despite the debatable real-life repercussions of such a teleological approach, there is a need for discussing the philosophical virtues and vices of their methodology. The general philosophical issue is whether practices really have intrinsic goals. Although some critical considerations have been raised in this chapter, this methodological discussion has not reached a final decision. Hence a teleological approach regarding the goals of medicine can still be defended.

A Consensual Approach

A less metaphysically charged approach was put forward by a group of scholars that discussed the goals of medicine at the Hastings Center. Here the idea was to use philosophical argument and empirical evidence to assemble a list of plausible goals of medicine, without assuming that it collects the only proper items of such a list. The methodology of such an approach can be called consensual, as it aims at an international consensus regarding the goals of medicine. Such a consensus requires deliberation and exchange of philosophical argument.

The group drafted the following list of four core goals of medicine (Hastings Center 1996, Executive Summary):

- The prevention of disease and injury and promotion and maintenance of health
- The relief of pain and suffering caused by maladies
- The care and cure of those with a malady and the care of those who cannot be cured
- The avoidance of premature death and the pursuit of a peaceful death

These goals are obviously not too different from the ones put forward by the defenders of the teleological approach. One pertinent difference, though, might be implied by the final goal of avoidance of premature death. Depending on what exactly is meant by "premature" death, there might be medical interventions, which Pellegrino and Kass would probably not see within the remit of medicine proper, for instance, the treatment of biologically normal deteriorations of fitness. In other words, the goal of avoiding premature death might justify enhancements – as opposed to treatment of disease.

Methodologically, the Hastings Center group seems to allow for revisions of their list, should there be considerable changes in social value judgments. After all, they assert "crucial points of contact between medical goals and social goals" (ibid., p. S6). Hence it is not quite clear as to how the setting of goals can establish an independent norm of the proper use of medicine when using the methodology of consensus. Even what many of us today regard as an instance of medicalization might change its status if the viewpoints within society change accordingly. To be sure, reasonable exchange would still be needed within such methodology, not just a majority vote or the like. But be that as it may, it seems that the key issues would still be found in the normative debate. There would be no external standard of the proper use of medical means, set by particular goals. This is to be expected within a consensual approach. It necessarily involves an element of conventionalism.

Alternative lists of goals of medicine have been proposed in the relevant literature (Miller and Brody 1998; Brülde 2001; Boorse 2016). These are fairly similar to the mentioned list, though they include additional aspects such as the improvement of healthy conditions in the environment, i.e., tackling the social determinants of health and reassuring the "worried well." Still, such similarity suggests a widespread convergence in normative assessments of the point of medicine and its remit – at least within a certain shared cultural background and at a particular point in time. It seems adequate to expect an ongoing debate about the goals of medicine in philosophy of medicine. This is at least partly due to the continuing dispute regarding the moral limits of the use of medical means to treat undesired ailments and to enhance desired conditions.

Conclusion

The main target of the debate on the goals of medicine has been to establish normative conclusions about the moral limits of the use of medicine for individual or social purposes. It has been claimed that the proper goals of medicine exclude certain medical practices, for instance, enhancements of fitness, the pursuit of aesthetic goals, the use of medicine in hastening death, or other contested aims. It has been shown in this chapter, however, that it is neither methodologically nor substantively straightforward that such a conclusion can be reached via an exploration of the goals of medicine. The debate on such goals is philosophically significant in its own right, but it is doubtful that the ethical issues can be solved on its basis.

Definition of Key Terms

Medicine	For the purposes of this chapter, medicine is understood both as a science and an art. It is a practice that contains numerous means in relation to the advancement of health. An exact definition of the concept of medicine is not forthcoming.
Medicalization	The use of medical means for improper purposes.
Teleology	An attempt to explain features or things by reference to purposes or goals.
Consensus	An attempt to find a coherent solution by means of deliberation in a group.

Summary Points

- The debate on the goals of medicine is usually concerned with the proper scope of medicine.
- A debate on the concept of medicine, and hence on its nature, might provide for a list of the goals of medicine.
- However, the concept of medicine has contested boundaries.
- Some scholars assume medicine to have a teleological structure and hence to aim at specific goals.
- Others have attempted to draft a list of the goals of medicine in a consensual approach.
- The philosophical debate on the goals of medicine is unlikely to disappear.
- An account of the goals of medicine will probably not solve the normative debate on the proper moral limits of the use of medicine.

References

Arras JD (2001) A method in search of a purpose: the internal morality of medicine. J Med Philos 26 (6):643–662

Boorse C (2016) Goals of medicine. In: Giroux E (ed) Naturalism in the philosophy of health: issues and implications. Springer, Cham, pp 145–177

Brülde B (2001) The goals of medicine. Towards a unified theory. Health Care Anal 9:1–13

Cassell EJ (1991) The nature of suffering and the goals of medicine. Oxford University Press, Oxford

Hastings Center (1996) Special supplement: the goals of medicine: setting new priorities. Hastings Cent Rep 26(6):S1–S27

Kaldjian LC (2014) Practicing medicine and ethics: integrating wisdom, conscience, and goals of care. Cambridge University Press, Cambridge

Kass LR (1985) Toward a more natural science: biology and human affairs. The Free Press, New York

Marcum JA (2008) Humanizing modern medicine. An introductory philosophy of medicine. Springer, Heidelberg
Miller FG, Brody H (1998) The internal morality of medicine: explication and application to managed care. J Med Philos 23:384–410
Nordenfelt L (1998) On medicine and health enhancement – towards a conceptual framework. Med Health Care Philos 1:5–12
Nordin I (1999) The limits of medical practice. Theor Med Bioeth 20:105–123
Pellegrino E (1999) The goals and ends of medicine: how are they to be defined? In: Hanson MJ, Callahan D (eds) The goals of medicine: the forgotten issue in health care reform. Georgetown University Press, Washington, DC, pp 55–68
Pellegrino E (2001) The internal morality of clinical medicine: a paradigm for the ethics of the helping and healing professions. J Med Philos 26(6):559–579
Pellegrino ED, Thomasma DC (1981) A philosophical basis of medical practice: toward a philosophy and ethic of the healing professions. Oxford University Press, New York/Oxford
Temkin O (1966) The meaning of medicine in historical perspective. Reprinted in Temkin O. The double face of Janus and other essay in the history of medicine. Johns Hopkins University Press 1977, Baltimore, pp 41–49
Veatch RM (2001) The impossibility of a morality internal to medicine. J Med Philos 26 (6):621–642

Suffering: Harm to Bodies, Minds, and Persons

10

Bjørn Hofmann

Contents

Abstract

Suffering is a basic human experience, and the concept of suffering can be defined in several ways. Typically, it is defined in terms of threats to human agency, loss or threat to an individual's value system, or as an experienced negative feeling. There are several types of (human) suffering. Suffering has been studied as a bodily, a mental, a social, and as an existential or spiritual experience. Suffering is frequently considered to be personal and subjective,

B. Hofmann (✉)
The Norwegian University of Science and Technology, Gjøvik, Norway

Centre for Medical Ethics, University of Oslo, Oslo, Norway
e-mail: bjorn.hofmann@medisin.uio.no

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_63

though not necessarily incomprehensible to others. Philosophy of language and phenomenology provide frameworks to understand other persons' suffering. While there is some agreement that suffering is something bad (an evil) toward which we have moral obligations to alleviate, there is less agreement on how this should be done and how far we should go. This is because there are several conceptual and ethical challenges with suffering, such as how to define it and where to set the limits to our duties and aspirations to alleviate suffering. What kind of suffering should be alleviated and by whom. For example, should social or existential suffering be alleviated by health care? What measures are acceptable in alleviating suffering, e.g., are modifications of personality, reduction of autonomy, or killing acceptable measures to alleviate suffering? Another basic question is who can suffer, e.g., whether animals, embryos, fetuses, or severely demented humans can suffer.

Introduction

Suffering is a basic human experience. Most people suffer during their lives, and suffering is seen as a basic part of the human condition (Spelman 1997; Morris 2002; Amato 1990; Schopenhauer 2008). As most people try to avoid suffering, it is considered to be one of the basic (negative) values of life (Spelman 1997; Cassell 1992; Hudson 2012). Suffering is also related to moral obligations: normally, we are obliged to help and care for those who are suffering and alleviate suffering where it is possible (Cassell 1982; Mayerfeld 2005; Wilson and Brown 2009; Andorno and Baffone 2014; Taylor and Watson 1989).

The term "suffering" is used in a wide range of contexts with many different and partly overlapping meanings (Cassell 2004; Edwards 2003; Green 2014). The term may have different meanings in a personal encounter, in a hospital setting, in a social group, in a piercing context, and in an existential and spiritual setting. Hence, there can be many concepts that cover the area of negative (human) experience. The term "suffering" can be used in a specific sense, e.g., in terms of physical suffering, and in a broad or vague sense. Common to many uses of the term "human suffering" is that it involves unpleasant sensory or emotional experience (DeGrazia 2014) and that it refers to states of severe human distress that threaten human agency (Cassell 1982, 2004). Suffering is also frequently (but not necessarily) related to events in time, i.e., suffering as a temporal event (Chapman and Gavrin 1992; Toombs 1999; Edwards 2003).

This chapter will start by reviewing the difference between suffering and pain as well as several common definitions of suffering in order to highlight some key features. Some traditional perspectives on (human) suffering will be presented. One section deals with the relationship between illness and ethics, and one addresses the question of whether suffering is comprehensible to others. The chapter ends with reviewing the ethics of suffering and some challenges with suffering as a concept.

Suffering and Pain

Suffering is related to pain, and the terms are often used synonymously (Scarry 1987; Wall 2000; Green and Palpant 2014). Persons in pain are said to suffer. However, suffering is considered to be distinct from pain (Cassell 2004; Clarke 2011; DeGrazia 2014). Although pain can result in or be part of suffering, pain is not a necessary condition for suffering (Fordyce 1988; Clarke 2011). A person can suffer without pain or when pain is relieved. Some people also seem to inflict pain on themselves in order to try to control suffering (Holm and Severinsson 2008). Pain is not a sufficient condition of suffering either, as people can have great pain without being distressed or feel their agency threatened. Fire walking and various types of passage rituals may cause pain, but not always suffering. Pain is considered to be a sensory experience, while suffering is not limited to sensory experience.

Definitions of Suffering

A wide range of definitions of suffering can be found in the literature (Malpas and Lickiss 2012; Green 2012). For the purpose of this overview, the definitions can be divided into three groups. One group of definitions of suffering emphasizes its *threats to human agency*: Suffering is a disruption, destruction, or loss of a person's dignity, integrity, intactness, or autonomy (Charmaz 1983; Shweder et al. 1997; Cassell 2004). This group of definitions may vary significantly with how agency is threatened (by disruption or loss) and what is threatened (autonomy, authenticity, or dignity). Despite these differences, they appear to cover some of the common ground: threats to human agency (Amato 1990; Green 2012).

Another group of definitions emphasizes that suffering is a *profound loss which impairs life* (Hoffmaster 2014). In particular, suffering may threat a person's existing values, transform his or her values, or enforce the elaboration of new values (Amato 2014). Accordingly, suffering is conceived of as an altered life, defined in terms of a constrained, compromised, or ephemeral life. A lot of opportunities are lost, and values are threatened or lost (Amato 1990). This group of definitions highlights a person's threatened value system. A third group of definitions of suffering highlights the *experienced negative sensation*, emotion, or feeling: there is an unpleasantness, which may threaten or shatter a personal felt reality (DeGrazia 2014; Davies 2012). For example, suffering can be experienced as torture without a torturer (Green and Palpant 2014).

One important question is whether suffering is subjective or objective. Some conceptions of suffering emphasize the *subjective experience*, such as pain, and see suffering as a personal matter. Suffering is experienced by the individual person (Cassell 2004). On the other hand, suffering is considered to be something that can be assessed by others, something *objective*. Accordingly, suffering is seen as something that frustrates the fulfillment of the being of an individual, e.g., biological needs, the need for food, shelter, sex, relationships, pursuing projects, and meaning in life (van Hooft 1998b).

Hence, there are many definitions of suffering within which three broad groups can be identified. For specific purposes, e.g., for certain types of research or care, it may be convenient to apply specific definitions, developed in certain contexts for particular purposes. In other settings, the three broad types of definitions above may be helpful. To the question of whether suffering is subjective or objective, some potential solutions exist (see below).

Perspectives on (Human) Suffering

In his *Epistulae morales ad Lucilium*, the Stoic philosopher Seneca (4 BC – 65 AD) describes three elements of disease: bodily pain (*dolor corporis*), suspension of joy (*intermission voluptatum*), and fear of death (*metus mortis*) (Seneca 2004). This illustrates that there exist long traditions of distinguishing between different types of human suffering. As shown, suffering extends beyond bodily pain and involves several other aspects. In describing how human suffering is experienced, it has been common to distinguish between four types of suffering. Most of these are covered by the first group of definitions mentioned above.

The first type of suffering is *bodily suffering*, which includes various types of pain, as well as nausea, dizziness, and feeling of cold or heat (Bakan 1968; Chapman and Gavrin 1992; Cassell 1999). The second type of suffering, *mental* (psychological, emotional) *suffering*, includes feelings of hopelessness, grief, sadness, anger, as well as guilt, regret, and embarrassment (Gilbert 1989; Byock 1996; Ratcliffe 2008; Davies 2012). *Social suffering*, the third type of suffering, involves (unwanted) dependency, shyness, withdrawal, and isolation (Kleinman et al. 1997; Carel 2014). Additionally, suffering is considered to be an *existential* (or spiritual) experience (Strang et al. 2004), e.g., in terms of fear of death, uncanniness, as well as loss of joy and meaning (van Hooft 1998b). Table 1 gives an overview over four traditional types of suffering and some examples of each type.

These traditional perspectives on suffering can be seen as reinforcing a traditional mind-body dualism in that suffering is experienced differently by the body and the mind. It may also reinforce a reductionistic conception of the world in general and of the human being in particular. That is, suffering can be studied in bits and pieces, in the same manner as we study the human being in terms of body and mind, as well as various social and existential faculties. Other conceptions may provide more holistic perspectives on human suffering, e.g., systems biology may offer fruitful ways to integrate physical, mental, and social aspects of human suffering (Federoff and Gostin 2009; Kirkengen et al. 2015). However, whether psychoneuroimmunological perspectives on suffering (Ulvestad 2012) offer something more fundamental than the integration of various perspectives is still open for debate (Gatherer 2010). Moreover, the different perspectives on suffering may also indicate that suffering is a complex phenomenon that cannot easily be grasped within one perspective or with one concept.

Table 1 Various traditional types of human suffering. The categories are overlapping, and it may not always be easy to distinguish between them

Bodily experience of suffering	Mental experience of suffering	Social experience of suffering	Existential or spiritual experience of suffering
Various types of pain	Various types of pain	Feeling vindicated	Fear of death
Weakness, bodily betrayal	Reduced ability	Learn to let others go or do	Deterioration, decline
Reduced ability	Anger	Learn to receive help	Shrinking of the world
Reduced function	Betrayal (of function)	Give up (friendship)	Modified meaning
Self-attention: awareness of body objectification	Self-attention: awareness of (function of) mind, objectification	Reinvent your life	Uncanniness, homelessness (Unheimlichkeit)
Sensation of excessive heat, excessive cold	Altered expectations	Learning to cope	Helplessness
Itching	Self-blaming, guilt	Surrender vanities	Grief (existential)
Hunger	Self-pity	Dependency	Uncertainty, unpredictability
Thirst	Self-censoring	Altered encounters	Fear for future
Nausea	Transformation	Isolation	Loss of joy
Suffocation, air hunger	Reduction of freedom	Shyness, awkwardness	Loss of meaning
Sleep deprivation	Changed self-conception	Lack of easiness	Loss of hope
Nausea	Changed self-perception	Shrinking of social space	Loss of flourishing
Disorientation	Curtailed possibilities (spatial, emotional)	Curtailed possibilities (relational, social)	Curtailed possibilities (existential)
Bodily disability	Mental disability	Social disability	Existential disability
Loss of balance	Modified embodiment, being in the world	Withdrawal	Loss of possibility, potential, capability
Fatigue	Grief	Isolation	Biographical break
Restlessness (legs)	Despair	Disgust	Crisis of identity
Feeling of “pins and needles”	Feeling of unfairness	Humiliation, embarrassment	Feeling of chaos
Twitching	Enervation	Loneliness	Awareness of finitude

(*continued*)

Table 1 (continued)

Bodily experience of suffering	Mental experience of suffering	Social experience of suffering	Existential or spiritual experience of suffering
Hiccups	Incapacitation	Loss of relation	Perception of fallibility
Sneezing	Anxiety, fear, panic,	Rejection	
Dizziness	Worry(ing)	Being expelled	
Tics	Confusion		
"Dysappearance" of the body (coming in the foreground)	Anguish		
	Sadness, depression		
	Restlessness		
	Regret		
	Envy, jealousy		
	Trying (beyond reach)		

Illness and Suffering

Suffering may result from injury, defects, infirmity, weakness, wounds, disability, disorder, disease, illness, or sickness. Malady is used as a common word for these types of human ailments (Clouser et al. 1997) and is defined as "a condition that involves the suffering or the increased risk of suffering an evil" (Clouser et al. 1981). Accordingly, suffering defines and results from malady. However, suffering is not the same as malady. In the same way that Huntington's disease is more than the 39 repetitions on the HD gene, suffering is more than the symptoms describing the disease. As indicated above, suffering may have a series of bodily, mental, social, and existential characteristics, and it may result from a wide range of other causes than malady.

Nevertheless, suffering is more closely related to illness than to any of the abovementioned types of malady (Kleinman 1988; Morse and Johnson 1991; Ware 1992). There is a tradition to distinguish between disease, illness, and sickness in the philosophy of medicine (Marinker 1973). *Disease* is the professional perspective on human malady, *sickness* is the societal perspective, and *illness* is the personal perspective on human malady. When talking about *disease*, we often try to give an objective account of human malady, while the languages of *sickness* and *illness* give voice to intersubjective and subjective accounts of human ailment, respectively. The basic entities defining and describing *disease* are biological (including biochemical and biomolecular phenomena) and mental events. The attributes of *sickness* are social role and status, while the basic feature of illness is suffering (Hofmann 2002). Figure 1 illustrates the relationship between disease, sickness, and illness. The figure demonstrates that the concepts do not always overlap and illustrates that there can be conflicts between them (Hofmann 2001, 2014).

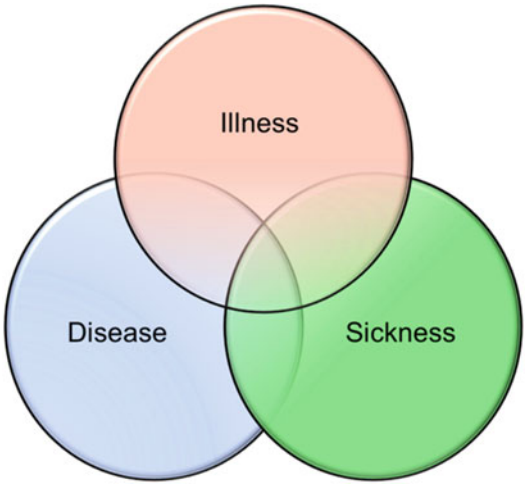


Fig. 1 The relationship between three core concepts of human ailment: illness, sickness, and disease (Adapted from Hofmann (2002))

Hence, illness is the personal aspect of human malady, e.g., a person might feel dizzy, weak, or exhausted, e.g., due to an infection. A person can be ill without being diseased or sick, e.g., when feeling lonely. While *disease* and *sickness* are considered to be external factors that can result in or from suffering, *illness* is more intimately related to suffering. Illness can be defined in terms of a negative experience as conceived of by the individual, such as in the case of severe nausea (Marinker 1975). Illness may well be conceived of as severe distress that threatens agency.

Despite the intimacy between illness and suffering, illness is not what obtains most attention in health care. Health professionals appear to be most preoccupied with disease (Usherwood 1990; Marinker 1975; Cassell 2004). For example, in the so-called medically unexplained physical symptoms (MUPS) (Kornelsen et al. 2015), persons have felt neglected by health professionals (Nunes et al. 2013). Health policy makers, employers, insurers, and legislators are primarily focusing on sickness. For example, they focus on whether the person can work or not and what measures are best suited to help (within or without the health-care system). More follows on the ethics of suffering below.

Personal but Not Inaccessible

As suffering is a personal experience, a key question has been whether it is accessible or comprehensible by others than the person himself. As pointed out above, some believe suffering to be a genuinely subjective concept (Marinker 1975; Cassell 1999; Evans 2003). This poses the question whether health-care professionals can know and address patients' suffering if suffering is something personal

and subjective (Frank 2001). Traditionally, there are at least two affirmative answers to these questions (Hofmann 2014). The first answer is based in philosophy of language and the second on phenomenology.

According to the first argument, all members of a linguistic community share a common language, by which they share the experiences that the language expresses (Wittgenstein 2001, §293). Ludwig Wittgenstein's beetle in the box thought experiment illustrates this. Wittgenstein invites people to imagine a community where all individuals each have a box containing a "beetle." The members all talk to each other about their "beetle," but no one can look into anyone else's box. The point of the experiment is to show that the talk of private experiences ("beetles" as an analogy of pain, in Wittgenstein's work) is learned through public experiences (when learning language and communicating). Accordingly, people may have access to the suffering of other persons through their language. By using a common language, apparently based on common experience, a person can express his or her experience of suffering. Hence, people in general and health professionals in particular may be able to understand other persons' suffering without "being under their skin" or "in their head." However, strictly speaking, it may well be the case that persons speaking different languages may experience suffering differently, unless suffering is something basically human which is experienced and expressed equally in all languages. We may have universal concepts in the same manner as we may have a universal grammar (Chomsky 1965).

A second argument emphasizing the possibility of having access to other persons' suffering comes from phenomenology. By systematic reflection, it is thought possible to determine the essential properties and structures of humans' experience of suffering (Carel 2014; Toombs 1999; Svenaeus 2014; Morrissey 2011; Leder 1985; Ratcliffe 2008; Merleau-Ponty 1962). Because all persons are living experiencing bodies in a common world, where they actively seek meaning, they can grasp basic aspects of suffering by introspection (looking into themselves) and by empathy (experiencing of another person's body as one's own) (Husserl 1963). In every person's suffering, there is something that is common to all human beings as all are beings in a common (life) world.

What then is common to people's suffering? Phenomenological studies have revealed several characteristics of suffering, several already mentioned in Table 1. Many people report that they experience uncanniness and that they lose joy, meaning, control, and balance when they are asked about their suffering. They experience feelings of rejection, dependence, and vulnerability. They feel that life is making a significant turn (a biographical break), and many report of an identity crisis. They cannot do as before. They cannot wish and aspire, or even hope as before. And they cannot be as before, both morally and in terms of character traits. Suffering makes people different. Correspondingly, specific characteristics of suffering can be found for particular diseases. For example, women diagnosed with and treated for breast cancer experienced a "field of force" that affected everything in their lives, including their views of themselves and their relationships (Arman et al. 2002).

Hence, even though we accept that suffering is a personal experience, it can be argued that we can still recognize, acknowledge, and understand other people's suffering. As shown this can be done by referring to the philosophy of language or by phenomenology, e.g., by empathy. According to optimistic neuroscientists and philosophers we will soon be able to understand people's suffering in terms of their brain states. Let us now return to a basic observation at the outset of this section "Ethics of Suffering."

Ethics of Suffering

Suffering is frequently considered to be something basically bad, i.e., an intrinsically negative value (Spelman 1997; Cassell 1992; Hudson 2012; van Hooft 2012). Correspondingly, all positions in bioethics consider suffering to be a (prima facie) moral obligation to be alleviated or ameliorated. Suffering evokes emotions, e.g., sympathy and compassion.

According to utilitarianism, suffering calls for action to reduce pain and increase pleasure (Beauchamp 1991). With deontology, suffering calls for duties to help and to care for the vulnerable (Kant 1780; Broad 1930), and virtue ethics refers to empathy and care for the sufferer (MacIntyre 1981). The suffering visible in the other's face has also been viewed as an ethical demand toward help and care, making the suffering of another person the basis for relationship and ethics as such (Levinas 1988).

In all these ethical approaches, suffering calls for attention, especially from health professionals. It directs the attention toward the person and not only toward the disease. Accordingly, the goal for health professionals is to treat a person, not only a disease.

From an evolutionary perspective, pain warns of threats, motivates coping, and avoids certain behavior (through punishment). Thereby suffering also is (loosely) related to learning (Decety 2014). Moreover, parental care can be seen as a moral response to suffering, which further has evolved into empathic arousal and concern for others. Hence, alleviating suffering is considered to be a basic moral obligation from a wide range of positions in ethics, including evolutionary biology (Krashin et al. 2014).

Accordingly, relieving suffering is considered to be a key goal for health professionals since the emergence of the professions, e.g., as expressed in the Hippocratic Oath: "I do solemnly swear by all I hold most sacred: . . . that I will exercise my art solely for the benefit of my patients, the relief of suffering, the prevention of disease and promotion of health, and I will give no drug and perform no act for an immoral purpose" (Adams 1891). Since then the obligation to relieve suffering is a stated goal in a wide range of professional codes of ethics (Davis 2003). In these codes, however, it is not always clear what is meant by suffering, and suffering is seldom explicitly on the curriculum in medical schools. Hence, although suffering is a core concept in medicine, it is rarely explicated beyond treating disease and palliating pain (Marinker 1973; Cassell 2004). Suffering has been explicitly on the agenda (and on the curricula) in nursing (Ferrell and Coyle 2008).

Although alleviating suffering is considered to be a core goal for all health professionals, it has been important to explore whether there are limits to this

obligation. Coherent with the unanimous agreement about a moral obligation to relieve suffering, there is agreement that there are limitations to this obligation (Rawlinson 1986; Hanson and Callahan 2000). Hence, the obligation to relieve suffering is prima facie. Consequentialists will argue that relieving suffering is acceptable only as long as it relieves the overall suffering. Hence, when relieving suffering in one person requires actions that impose suffering in others increasing the total suffering, it is not acceptable. Deontologists will normally argue that the duty to relieve suffering is valid only as it does not infringe or conflict with other basic duties, such as the duty not to kill, to cheat, to lie, etc. Correspondingly, it can be argued that where it leads to modification of personality (lobotomy, deep brain stimulation), or reduces autonomy (extreme sedation), alleviating suffering is controversial (Rousseau 2001).

Likewise, it is argued that the moral obligations to relieve others' suffering do not require sacrifice that violates personal concerns or other moral obligations. For example, a person is not obliged to care for a suffering stranger if that will result in neglect in the basic care for his or her own child, siblings, or parents (Nortvedt and Nordhaug 2008). Accordingly, the obligation to relieve suffering can be considered to be supererogatory. More on the limits to the types of suffering relevant for health professionals follows in the paragraph on *challenges* below.

Another ethical issue related to (the relief of) suffering is the problem of double effect (Cavanaugh 2006). The key issue is whether it is acceptable to relive suffering if this inevitably results in a higher risk of also ending the person's life. The "principle of double effect" has been applied as a set of criteria for when this can be acceptable.

Hence, the obligation to relive suffering is considered to be a moral obligation from a broad range of positions in moral philosophy, and there appears to be agreement that there are limitations to this obligation. However, exactly where the limits go, and how to justify them, is not always clear.

The Mystery of Suffering

Suffering is a basic human experience considered to be an intrinsic evil. It has aroused curiosity and speculation throughout centuries. As such, suffering is brimmed with mystery and has been subject to elaborate religious reflection and speculation (Green and Palpant 2014) in addition to existential challenges (Sartre 2003). In a wide range of religious traditions, suffering is attributed to a meaning or function, such as purifying the soul, ennobling the person, or a path to spiritual growth. Moreover, it is viewed as something human beings have to endure.

Buddhist traditions consider relief of suffering (dukkha) to be a key part of obtaining the supreme bliss (nirvana) (Schlieter 2014). In Hinduism suffering follows naturally from negative behaviors in a person's current or past life. Suffering is also at the core of Christian teachings, where the suffering of one person is considered to have freeing implications for many (Green and Palpant 2014). Table 2

Table 2 The meaning of suffering in some religious and secular perspectives (see van Hooft 1998a)

Perspective	Conceptions of suffering	Is suffering good or bad?
From antiquity	Suffering is a result of offending the divine order	Suffering was inevitable, negative, and necessary. However, as it was ordained by the supernatural order, it ultimately is positive
Stoic	Indifference: suffering is something morally irrelevant	Neither good nor bad
Christian	Suffering is part of God's salvific plan for humanity. That is, suffering is a kind of reparation	In the overall story, suffering is good
Humanist	Suffering grounds the possibility of ethics through compassion	Suffering creates bonds between human beings
Nietzschean	Suffering ennobles the human spirit and makes possible human advancement and personal self-validation	Suffering is a creative source

gives a short outline of some religious and secular meanings of suffering. Hence, suffering has strong philosophical (Nussbaum 1994) and religious connotations (Green and Palpant 2014), and the boundary between religious and medical handling of suffering has been unclear, especially as human ailment has been addressed and cared for by religious organizations and as the hospital movement stems from such organizations.

The boundaries between the medical and the religious aspects of suffering are not always clear and are subject to constant reflection (Best et al. 2014; Boston et al. 2011; Kissane 2012). Health professions vary in the attention to spiritual aspects of suffering. Health professionals have been considered to "have a calling," a professional altruistic obligation to help suffering persons based on a religious analogy. In secular societies, the bodily, mental, and social aspects of suffering appear to obtain more attention than the spiritual or existential aspects (Cherry 2014). How the various aspects of suffering are assessed and addressed is an issue for continuous debate.

Challenges

The basic questions following from the obligation to alleviate suffering are whether there are limits and what kind of suffering health professionals are to alleviate, and, if there are limits, where do they go (Rousseau 2001). The answer to the first question is normally yes, as there are many types of suffering that are not subject to health care. For example, if people suffer due to incompetence in a hi-tech environment, health professions may have little to offer, while other professionals may be able to help. Correspondingly, there are a wide range of social conditions that may result in suffering, which are not the subject matter for health professionals.

The second question, where the limits to health-care intervention are, is more challenging. There are a wide range of ordinary human experiences that can become pathologic, e.g., grief is an ordinary experience, which may warrant attention from health professionals (Wakefield 2013). Social suffering due to protruding ears may be alleviated by surgery, so may suffering for single persons for not having a partner or a child, e.g., with various types of assistive reproduction (Lauritzen 2014). Hence, the core question is where to set limits to avoid medicalization of human suffering. The traditional answer to this question has been that health care should limit its activity to what is related to malady. However, to delimit disease is in itself a normative question. On the other hand, the question is how far we should go to relieve suffering. In particular, is it right to kill (a person) in order to alleviate suffering? The question is pertinent in euthanasia, physician-assisted suicide, and selective abortion (Burgess 1993). These are difficult questions related to the concept of suffering, which merit separate analyses.

Another question that has troubled philosophers dealing with suffering is whether an unequivocal moral judgment in response to (human) suffering requires a common moral position (Hoffmaster 2014; Green and Palpant 2014). Can various stances in moral philosophy result in the same response to suffering? One answer is that human suffering is such a basic bad (evil) in most ethics, resulting in aversion and moral reactions in all. Another answer is that moral responses to suffering vary greatly, depending on peoples' moral stance. It can be useful to differentiate between suffering moving moral concern and the particular moral responses to suffering. While the first will be common and unequivocal to most positions in ethics, the latter may not.

Yet another important question is "who can suffer?"(DeGrazia 2014). Several of the definitions presented above require that a sufferer has to be a person. So, who counts as a person? Do embryos, fetuses, children, severely demented, or animals count as persons? These questions hinge on issues of moral status and consciousness. Depending on how one defines personhood, embryos, fetuses, and animals may or may not suffer. According to the third group of definitions presented earlier, several types of animals may very well suffer, as they have negative feelings (beyond instinctive sensations) (Dawkins 1980; Singer 1990). How the second type of definitions, highlighting the altered value system in suffering, considers animals, embryos, and fetuses to suffer depends on the specific definition. Several positions attribute more obligations toward nonhuman persons (animals) than human nonpersons (e.g., humans with severe dementia) (Singer 2011). In addition to arguments from various definitions of suffering, it is also argued from analogy that animals suffer: they behave as if they suffer, and therefore we have obligations toward animals. Neurological studies of animals have also increasingly revealed animals' experience of pain. We may of course very well have obligations toward other entities in nature, such as plants, fetuses, and animals without emotions, even when they do not suffer. However, such obligations need other moral foundations than suffering.

According to the various perspectives, types, and definitions described above, there are many languages of suffering. For example, there are biomedical, religious, existential, moral, and political languages of suffering (Brinkmann 2014), all framing social relationships and power. Moreover, suffering calls for moral sentiments such as sympathy, compassion and pity, and result in social bonding (Halpern 2002), and suffering resembles power in many ways (Eriksson 2006). Hence, suffering poses social and cultural challenges (Kleinman and Kleinman 1996) well beyond the realm of health care.

Summary

Accordingly, suffering is a basic human experience. It can be defined in many ways, but suffering is typically defined in terms of i) threats to human agency, ii) loss or threat of value system, and/or iii) experienced negative feelings. Hence, suffering is a negative experience of some kind. However, several types of (human) suffering are studied and discussed in the literature on suffering. Suffering can be studied as a bodily experience, mental experience, social experience, an existential or a spiritual experience.

Suffering may be personal, but not necessarily incomprehensible by others. The philosophy of language and phenomenology provides ways to understand persons' suffering. There is unanimous agreement that suffering is a moral bad (evil) that we have moral obligations to alleviate. However, there is less agreement on how this should be done and how far we should go.

There are several challenges with suffering, e.g., where to set the limits to alleviate suffering, both for social (nonmedical) issues and with respect to ending life (killing, euthanasia). Another basic question is who can suffer (animals, embryos, fetuses, severely demented). There are no general agreements on these questions.

Definition of Key Terms

Disease	There is no agreement on how to define disease, but many definitions refer to physiological, biochemical, genetic, and mental entities, and events are conceived of as being of negative value by the medical profession who therefore strives to classify, detect, control, and treat such entities and events in order to explain, palliate, and ultimately to cure.
Illness	Most definitions of illness refer to bodily, behavioral, or mental entities or events that the person experiences and mainly consider to be of negative value.
Sickness	Many definitions of sickness refer to bodily, behavioral, or mental entities or events that are conceived of by the society and/or its institutions to be of negative value.

Summary Points

- Suffering is a basic human experience.
- Suffering can be defined in many ways.
- Suffering is a negative experience of some kind.
- Suffering is most frequently defined in terms of threats to human agency, loss or threat of value system, and experienced negative feelings.
- Suffering can be studied as a bodily experience, a mental experience, a social experience, and an existential or a spiritual experience.
- Suffering is most frequently conceived of as a moral bad that should be alleviated.
- There are several challenges related to the concept of suffering.

References

Adams F (1891) The genuine works of hippocrates. William Wood and Company, New York. Here cited from http://www.thefullwiki.org/Hippocratic_Oath#cite_note-9. Accessed Oct 2014
Amato JA (1990) Victims and values: a history and a theory of suffering. Praeger, New York
Amato JA (2014) Suffering, and the promise of a world without pain. In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, pp 61–87
Andorno R, Baffone A (2014) Human rights and the moral obligation to alleviate suffering. In: Green R, Palpant N (eds) Suffering and bioethics. Oxford University Press, New York, pp 182–200
Arman M, Rehnsfeldt A, Lindholm L, Hamrin E (2002) The face of suffering among women with breast cancer-being in a field of forces. Cancer Nurs 25(2):96–103
Bakan D (1968) Disease, pain, and sacrifice: toward a psychology of suffering. University of Chicago Press, Chicago
Beauchamp TL (1991) Philosophical ethics: an introduction to moral philosophy, 2nd edn. McGraw Hill, New York
Best M, Aldridge L, Butow P, Olver I, Price M, Webster F (2014) Assessment of spiritual suffering in the cancer context: a systematic literature review. Palliat Support Care 11:1–27
Boston P, Bruce A, Schreiber R (2011) Existential suffering in the palliative care setting: an integrated literature review. J Pain Symptom Manage 41(3):604–618
Brinkmann S (2014) Languages of suffering. Theor Psychol 24:630–648
Broad CD (1930) Five types of ethical theory. Harcourt, Brace and Company, New York
Burgess MM (1993) The medicalization of dying. J Med Philos 18(3):269–279
Byock IR (1996) The nature of suffering and the nature of opportunity at the end of life. Clin Geriatr Med 12(2):237–238
Carel HH (2014) Illness and its experience: the perspective of the patient. In: Handbook of concepts in the philosophy of medicine. Springer, Dordecht
Cassell EJ (1982) The nature of suffering and the goals of medicine. N Engl J Med 306:639–645
Cassell EJ (1992) The nature of suffering: physical, psychological, social and spiritual aspects. In: Starck P, McGovern J (eds) The hidden dimension of illness: human suffering. National League for Nursing Press, New York, pp 1–10
Cassell EJ (1999) Diagnosing suffering: a perspective. Ann Intern Med 131(7):531–534
Cassell EJ (2004) The nature of suffering and the goals of medicine, 2nd edn. Oxford, New York
Cavanaugh TA (2006) Double-effect reasoning: doing good and avoiding evil. Clarendon, Oxford
Chapman CR, Gavrin J (1992) Suffering and its relationship to pain. J Palliat Care 9(2):5–13
Charmaz K (1983) Loss of self: a fundamental form of suffering in the chronically ill. Sociol Health Illn 5(2):168–195

Cherry MJ (2014) Human suffering and the limits of secular bioethics. In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, pp 337–355
Chomsky N (1965) Aspects of the theory of syntax. MIT Press, Cambridge, MA
Clarke BM (2011) On suffering: pathways to healing & health. Dartmouth College Press, Hanover
Clouser KD, Culver CM, Gert B (1981) Malady: a new treatment of disease. Hastings Cent Rep 11(3):29–37
Clouser K, Culver C, Gert B (1997) Malady. In: Humber J, Almeder R (eds) What is disease? Humana Press, Totowa, pp 173–218
Davies J (2012) The importance of suffering: the value and meaning of emotional discontent. Routledge, London
Davis M (2003) What can we learn by looking for the first code of professional ethics? Theor Med Bioeth 24(5):433–454
Dawkins MS (1980) Animal suffering: the science of animal welfare. Chapman and Hall, London
Decety J (2014) Social neuroscience meets philosophy: suffering, empathy, and moral cognition. In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, pp 89–104
DeGrazia D (2014) What is suffering, and what kind of beings can suffer? In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, p 138
Edwards SD (2003) Three concepts of suffering. Med Health Care Philos 6(1):59–66
Eriksson K (2006) The suffering human being. Nordic Studies Press, Chicago
Federoff HJ, Gostin LO (2009) Evolving from reductionism to holism: is there a future for systems medicine? JAMA 302(9):994–996
Ferrell BR, Coyle N (2008) The nature of suffering and the goals of nursing. Oncol Nurs Forum 35 (2):241–247
Fordyce WE (1988) Pain and suffering: a reappraisal. Am Psychol 43(4):276
Frank AW (2001) Can we research suffering? Qual Health Res 11(3):353–362
Gatherer D (2010) So what do we really mean when we say that systems biology is holistic? BMC Syst Biol 4:22
Gilbert P (1989) Human nature and suffering. Psychology Press, Hove
Green R, Palpant N (eds) (2014) Bioethics and suffering. Oxford University Press, New York
Halpern C (2002) Suffering, politics, power: a genealogy in modern political theory. State University of New York Press, Albany
Hanson MJ, Callahan D (2000) The goals of medicine: the forgotten issues in health care reform. Georgetown University Press, Washington
Hoffmaster B (2014) Understanding suffering. In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, pp 31–53
Hofmann B (2001) Complexity of the concept of disease as shown through rival theoretical frameworks. Theor Med Bioeth 22(3):211–236
Hofmann B (2002) On the triad disease, illness and sickness. J Med Philos 27(6):651–674
Hofmann B (2014) [What is disease? In Norwegian] Hva er sykdom? Gyldendal Akademisk, Oslo
Holm AL, Severinsson E (2008) The emotional pain and distress of borderline personality disorder: a review of the literature. Int J Ment Health Nurs 17(1):27–35
Hudson W (2012) Historicizing Suffering. In J. Malpas & N. Lickiss (Eds.), Perspectives on Human Suffering. Dortrecht: Springer, pp. 171–180.
Husserl E (1963) Ideas: a general introduction to pure phenomenology (trans: Boyce Gibson WR). Collier Books, New York
Kant I (1780/1998) The Metaphysical Elements of Ethics (T. K. Abbott, Trans.). Auckland: Floating Press
Kirkengen AL, Ekeland TJ, Getz L, Hetlevik I, Schei E, Ulvestad E, Vetlesen AJ (2015) Medicine's perception of reality – a split picture: critical reflections on apparent anomalies within the biomedical theory of science. J Eval Clin Pract. [Epub ahead of print]
Kissane DW (2012) The relief of existential suffering. Arch Intern Med 172(19):1501–1505
Kleinman A (1988) The illness narratives: suffering, healing, and the human condition. Basic Books, New York

Kleinman A, Kleinman J (1996) The appeal of experience; the dismay of images: cultural appropriations of suffering in our times. Daedalus 125:1–23
Kleinman A, Das V, Lock MM (eds) (1997) Social suffering. University of California Press, Berkeley
Kornelsen J, Atkins C, Brownell K, Woollard R (2015) The meaning of patient experiences of medically unexplained physical symptoms. Qual Health Res. [Epub ahead of print]
Krashin D, Murinova N, Howe C, Ballantyne J (2014) Biology and suffering. In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, pp 106–131
Lauritzen P (2014) Reproductive technology in suffering's shadow. In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, pp 357–373
Leder D (1985) Toward a phenomenology of pain. Rev Existent Psychol Psychiatry 19:255–266
Levinas E (1988) Useless suffering. In: Bernasconi R, Wood D (eds) The provocation of Levinas: rethinking the other. Routledge, London
MacIntyre A (1981) After virtue: a study in the moral theory. Duckworth, London
Malpas J, Lickiss N (eds) (2012) Perspectives on human suffering. Springer, Dordrecht
Marinker M (1975) Why make people patients? J Med Ethics 1(2):81–84
Mayerfeld J (2005) Suffering and moral responsibility. Oxford University Press, New York
Merleau-Ponty M (1962 [1945]) Phenomenology of perception (trans: Smith C). Routledge, New York
Morris DB (2002) The culture of pain. University of California, Berkley
Morrissey MB (2011) Phenomenology of pain and suffering at the end of life: a humanistic perspective in gerontological health and social work. J Soc Work End Life Palliat Care 7 (1):14–38
Morse JM, Johnson JL (1991) The illness experience: dimensions of suffering. Sage, Newbury Park
Nortvedt P, Nordhaug M (2008) The principle and problem of proximity in ethics. J Med Ethics 34 (3):156–161
Nunes J, Ventura T, Encarnacao R, Pinto PR, Santos I (2013) What do patients with medically unexplained physical symptoms (MUPS) think? A qualitative study. Ment Health Fam Med 10 (2):67–79
Nussbaum M (1994) The therapy of desire: theory and practice in Hellenistic ethics. Princeton University Press, Princeton
Ratcliffe M (2008) Feelings of being: phenomenology, psychiatry and the sense of reality. Oxford University Press, Oxford
Rawlinson MC (1986) The sense of suffering. J Med Philos 11(1):39–62
Rousseau P (2001) Ethical and legal issues in palliative care. Prim Care 28(2):391–400
Sartre JP (2003 [1943]) Being and nothingness (trans: Eyre R). Routledge, London
Scarry E (1987) The body in pain: the making and unmaking of the world. Oxford University Press, New York
Schlieter J (2014) Adapt, endure, or overcome? The concept of "suffering" in Buddhist bioethics. In: Green R, Palpant N (eds) Bioethics and suffering. Oxford University Press, New York, pp 309–336
Schopenhauer A (2008). The World as Will and resentation, trans. Richard E. Aquila in collaboration with David Carus. Longman, New York
Seneca (2004) On the shortness of life (trans: Costa CDN). Penguin, London
Shweder RA, Much N, Mahapatra M, Park L (1997) The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In: Brandt AM, Rozin P (eds) Morality and health. Routledge, London, pp 119–169
Singer P (1990 [1975]) Animal liberation. New York Review Books, New York
Singer P (2011 [1979]) Practical ethics. Cambridge University Press, New York
Spelman EV (1997) Fruits of sorrow: framing our attention to suffering. Beacon, Boston
Strang P, Strang S, Hultborn R, Arnér S (2004) Existential pain – an entity, a provocation, or a challenge? J Pain Symptom Manage 27(3):241–250

Svenaeus F (2014) The phenomenology of suffering in medicine and bioethics. Theor Med Bioeth 35(6):407–420
Taylor RL, Watson J (eds) (1989) They shall not hurt: human suffering and human caring. Colorado Associated University Press, Boulder
Toombs KS (1999) The meaning of illness: a phenomenological account of the different perspectives of physician and patient. Kluwer, Amsterdam
Ulvestad E (2012) Psychoneuroimmunology: the experiential dimension. Methods Mol Biol 934:21–37
Usherwood TP (1990) Responses to illness – implications for the clinician. J R Soc Med 83 (4):205–207
van Hooft S (1998a) The meanings of suffering. Hastings Cent Rep 28(5):13–19
Van Hooft S (1998b) Suffering and the goals of medicine. Med Health Care Philos 1:125–131
van Hooft S (2012) Ethics of suffering: modern law, philosophy and medicine/Marinos Diamantides. Monash Bioeth Rev 20(4):50–52
Wakefield JC (2013) DSM-5 grief scorecard: assessment and outcomes of proposals to pathologize grief. World Psychiatry 12(2):171–173
Wall PD (2000) Pain: the science of suffering. Columbia University Press, New York
Ware NC (1992) Suffering and the social construction of illness: the delegitimation of illness experience in chronic fatigue syndrome. Med Anthropol Q 6(4):347–361
Wilson RA, Brown RD (2009) Humanitarianism and suffering: the mobilization of empathy. Cambridge University Press, Cambridge
Wittgenstein L (2001 [1953]) Philosophical investigations. Blackwell, Oxford

Disability as Medical and as Social Category 11

Steven Edwards

Contents

Abstract

This chapter draws attention to several philosophical issues raised by the phenomenon of disablement and then focuses on two main ones pertinent to philosophy of medicine: the definition of disability and the relationship between disability and identity. Two kinds of approach are identified in relation to the question of the definition of disability, one of which focuses on the individual concerned and is sometimes described as a "medical model" of disability and another approach which places more emphasis on the environment beyond the individual. The World Health Organization's (WHO) taxonomy is presented below to represent the first kind of approach, and the theory devised by Professor Lennart Nordenfelt is presented as representing the second. The chapter then turns to discuss disability and identity. It is shown that on standard ways of conceiving of the identity relation, disability seems closer to a contingent (and so non-identity-constituting) characteristic of persons as opposed to an essential,

S. Edwards (✉)
Philosophy, History and Law, Swansea University, Swansea, Wales, UK
e-mail: s.d.edwards@swansea.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_3

identity-constituting one. However, another strategy is also described in which certain kinds of contingent properties can be identity-constituting. So if the latter strategy proves successful, then it may be true that disability can indeed be identity-constituting.

Introduction

Disability raises many philosophical questions regarding, for example, its definition, its normative character, its causes, its relationship with a person's identity, the question of a person's moral status and entitlement to life, as well as their entitlement to social justice. Since this handbook's primary focus is on philosophy of medicine, the two main issues addressed are two which seem most pertinent from a medical or health-care perspective. The first of these is the very definition of disability. This discussion highlights two distinct approaches, one which focuses on the individual concerned and is sometimes described as a "medical model" of disability and another approach which places more emphasis on the environment beyond the individual. The World Health Organization's (WHO) taxonomy is presented below to represent the first kind of approach, and the theory devised by Professor Lennart Nordenfelt is presented as representing the second. The question of how disability should be defined has obvious relevance to philosophy of medicine in particular since if it is true that the causes of disability lie in the individual concerned, then attempts to diagnose and respond therapeutically to disability will be directed primarily at the individual. But if the real causes lie beyond the individual, the role of medical practitioners in responding to disability is much less clear. It may even turn out that disability could be completely severed from the domain of health care since it would, at most, be a social problem as opposed to a medical one – something more in common with poverty or sexual discrimination as opposed to appendicitis or schizophrenia. In the latter part of the chapter, the question of the relationship between disability and identity is discussed. Somewhat paradoxically, theorists who claim that disability is largely caused by factors beyond the person maintain that disability is an essential part of who they are. This sounds prima facie paradoxical since it seems more likely that a person's identity would be determined by factors intrinsic to them, such as their genetic constitution, as opposed to extrinsic factors, such as social conditions.

Defining Disability: A Medical Model

No human being is wholly able, so to speak, able to do everything. There are some kinds of activities which a person may not be able to accomplish, yet this does not entail they have a disability or are disabled (Nordenfelt terms these "nonabilities"). Yet, historically, human beings have been identified as lacking certain abilities

considered typical of humans and identified as such. Thus, Braddock and Parish (2001) note that a distinction was frequently drawn between those people with what would now be thought of as physical impairments (e.g., lacking limbs), those with sensory impairments, those with what would now be thought of as intellectual disabilities, and those with severe mental health problems. Braddock and Parish emphasize the significance of a distinction between impairment and disability since they endorse the view that while impairments are grounded in actual physical difference and thus to a significant extent independent of social context, this is not true of the category of "disability" which they say is much more closely bound up with local values and norms.

The view that there is indeed at least this level of complexity in the classification of disability became widely recognized in the twentieth century and led to increasingly sophisticated attempts to taxonomize the phenomenon of "disablement" (e.g., Nagi 1965; WHO 1980; Oliver 1990; Nordenfelt 1983/1997).

Of the various approaches attempted, two general emphases can be discerned – reflected in the division referred to by Braddock and Parish – between on the one hand a primary focus on the physical constitution of the person concerned and approaches which have contextual factors as their primary focus on the other. Between these two extremes, as might be anticipated, are approaches which seek to acknowledge both internal and external factors as opposed to focusing narrowly on either one or the other.

The World Health Organization's (WHO) taxonomy from 1980 can be taken to represent approaches which at least appear to have the individual person as their primary focus. The conditions with which it is concerned are those which it describes as the "consequences of disease" (1980, p. 1). In their *International Classification of Impairments, Disabilities and Handicaps* (ICIDH), it is observed that disease can lead to impairment, which can lead to disability, which can lead to handicap. This is represented in a schema presented thus which illustrates the sense in which impairment, disability, and handicap are conceived of as *consequences* of disease:

> "disease >> impairment >> disability >> handicap" (WHO 1980, 11; see also Altman's depiction of Nagi's schema (2001, 113))

Disease is defined in terms of "aetiology >> pathology >> manifestation" (WHO 1980, p. 10). Thus, consider the former Olympic and Paralympic athlete Oscar Pistorius. Due to a genetic anomaly, he was born without fibulae; the aetiology would be the anomaly itself (a genetic difference or more strongly disease), the pathology it leads to is the lack of fibulae, and its manifestation is the missing fibulae.

Impairments are defined as follows: "Impairment: In the context of health experience, an impairment is any loss or abnormality of psychological, physiological or anatomical structure or function" (ibid. p. 27) and are said to arise at the level of "parts of the body" (ibid. p. 28). So the missing fibulae would qualify as impairments because their absence constitutes an abnormality – statistically speaking – in anatomical structure since it is statistically typical for humans to be born with fibulae.

According to the ICIDH, impairments are consequences of disease and disabilities consequences of impairments. So consider the definition of disability offered. "Disability: In the context of health experience, a disability is any restriction or lack (resulting from impairment) of ability to perform an activity in the manner or within the range considered normal for a human being" (ibid. p. 28). Whereas impairments arise at the level of "body parts" such as organs, disabilities are said to arise at the level of the individual person (ibid.). So to focus on Oscar Pistorius (OP) again, since walking is an activity within the range considered normal (statistically) for humans, the impairment leads to a lack of the ability to walk (unaided). Also, of course the relevant comparator group here is humans at a particular chronological stage of development: it is not statistically abnormal for babies to be unable to walk.

Turn now to the third consequence of disease as presented in the ICIDH, that of handicap. "Handicap: In the context of health experience, a handicap is a disadvantage for a given individual, resulting from an impairment or a disability, that limits or prevents the fulfilment of a role that is normal (depending on age, sex and cultural factors) for that individual" (ibid. p. 29). As mentioned, impairments are properly attributed only to body parts and disabilities only to persons, so it is said that handicap is a "social phenomenon" (ibid.). For in contrast to the other two consequences of disease, this category makes explicit reference to social and cultural factors. So it might be said that OP's missing fibulae constitute a "disadvantage" to him since they limit the range of activities and social roles open to him just as being unable to see, for example, may be said to be a handicap since it limits the social roles open to the person.

It is evident, then, that on the ICIDH schema, impairment may be a necessary condition of disability but is not a sufficient condition. Fused toes may count as impairments as they comprise a structural abnormality, but providing they don't limit the person's ability to walk, they fall short of disabilities. Here there are structural abnormalities with no functional consequences. (So impairments differ from diseases.) Also, as the case of OP illustrates, there may be provision to mitigate adverse functional consequences of impairments such as walking blades, spectacles for the shortsighted, insulin for diabetes, and so on. So again here we have impairments but no disability, provided relevant "external" compensating conditions are present, such as the availability of walking blades, spectacles, or insulin. Other kinds of contingent cultural factors can also be shown to be relevant to the question of whether a disability leads to a handicap. In the island known as "Martha's Vineyard," situated off the coast of North America (Sacks 1989), a high proportion of the residents were deaf, and most residents of the island were competent in using sign language. In this context deafness may be a disability but does not lead to a handicap (Oliver 1990, p. 16; Sacks 1989).

As may be anticipated perhaps, the ICIDH taxonomy generated substantial amount of criticism (see Oliver 1990); the most influential criticism focused on an interpretation of the kind of causal claim made in the document (see esp. Wasserman (2001) for philosophical discussion of this). Critics claim that according to the ICIDH, because it is a "consequence of disease," it follows that the causes of disability lie in the individual. But, it is argued that the causes of disability lie wholly

in the social environment not in the individual with the impairment (UPIAS 1975). The force of this claim is captured well in the front cover picture of Oliver's *The Politics of Disablement* which depicts a person in a wheelchair at the foot of a flight of steps of a building which is a polling station. The clear message is that the wheelchair user is perfectly able to vote but is "disabled" from doing so by the wheelchair-unfriendly social environment.

Moreover, it is argued that if it is held that the cause of disability is to be found within the "diseased" individual, they will be conceived of as suffering from a medical problem which requires a medical response (indeed Harris has claimed that a disabled person "will inevitably suffer" (2000)). This so-called "medicalization" of disability (Oliver 1990) fosters the impression that disability and handicap are inevitably accompanied by illness and this is a mistake it is said. A person might have an impairment which leads to a disability (say lack of ability to walk) but consider themselves perfectly healthy. One can think of many Paralympic athletes in this context to reinforce the conceptual separation between disease, impairment, and health. Thus, it may be held that even impairments which lead to deviations from species-typical functions as fundamental as seeing, hearing, and walking need not impede health (see esp. Nordenfelt's work, of which more below). So, there is no necessary relationship between disability and illness, contrary to the impression fostered by a medicalized view of disability. Moreover, the view that the cause of disability lies in the individual has clear implications regarding the focus of attempts to remedy it, namely, the "diseased" person themselves. Hence, from the perspective of the contrasting "social model" (Oliver 1990) according to which the cause of disability lies in the social environment, the medicalized ICIDH model is seriously flawed. Lastly, the claimed medicalization of disability generated by the ICIDH led to it being regarded as a medical problem requiring the input of medical and other health-care personnel, and this, it is argued, is wholly inappropriate given the contingent nature of the relationship between impairment, disability, and illness (Oliver, op. cit; Swain et al. 1993).

Many of these criticisms of the ICIDH and the claimed "medical model" of disability it presents were accepted, and a revised WHO schema appeared in 2001 (WHO 2001). The *International Classification of Functioning, Disability and Health* (usually abbreviated to ICF) stressed the multidimensional nature of disablement and emphasized that "a person's functioning and disability is conceived of as a dynamic interaction between health conditions (diseases, disorders, injuries, traumas, etc.) and contextual factors" (ibid. p. 10); these contextual factors include "features of the physical, social and attitudinal world" (ibid.). So even if the earlier schema did express a much simpler causal picture, the latter model is explicitly of the view that the causes of disability are manyfold and include factors beyond the body of the individual concerned. Despite this major change of emphasis, the basic categories remain more or less unchanged; impairment is much as before what were disabilities are now labeled "activity limitations" (ibid. p. 191), and what were labeled handicaps are renamed "participation restrictions" (ibid.).

Also, as was the case in the ICIDH, each of the main categories is defined with reference to statistical norms: the definition of impairment in ICF is explicit that "Abnormality here is used strictly to refer to a significant variation from established

statistical norms" (ibid. p. 190); the relevant comparator group in relation to activity limitations is the "manner or to the extent to that is expected of people without the health condition" (ibid. p. 191); and in the case of participation restrictions, the comparison is with "that which is expected of an individual without disability in that culture or society" (ibid). The kinds of norms these three categories appeal to are much narrower than biostatistical norms and make reference to cultural or local norms, but, nonetheless, the possibility arises that a person may be categorized as having an activity limitation or participation restriction even if they themselves deny this.

Nordenfelt's Theory

In an interesting series of publications, the philosopher Lennart Nordenfelt argues that the question of whether or not a person has a disability is not, in most cases, separable from the values of the person themselves. So rather than tie classification to statistical norms, Nordenfelt argues for a theory in which the categorization of a person as having a disability depends to a large extent on what is of value to that person (see Nordenfelt 1983/1997, 1993, 1995, 2000, 2001). His approach is one which is part of his general theory of health in which the question of a person's health is similarly bound up with the person's own views regarding what is a good life (Nordenfelt 1995; 2001). The approach tries to tread a fine line between being a subjectivist one in which the question of whether or not a person is disabled is wholly determined by them and what might be described as an objectivist one – such as that which is found in the WHO schemas – and in accordance with which a person can be categorized as disabled irrespective of their own values or opinion on the matter.

A central component of Nordenfelt's theory is that of a vital goal. The definitions of disability and handicap given by Nordenfelt are these: "A disability, as well as a handicap, is a non-ability – given a specified set of circumstances – to realize one or more of one's vital goals (or any of its necessary conditions)" (1993, p. 22). By "vital goal" Nordenfelt means "a state of affairs that is a necessary condition for the realization of A's at least minimal happiness" (1993, p. 20). It is made plain that vital goals may include activities which are important to an individual such as cinema-going, bird-watching, and, more controversially perhaps, even sports-related goals involving high levels of achievement (Nordenfelt 2007; Schramme 2007). The idea here is familiar enough, namely, that most people perform everyday acts with some longer-term goal in mind and these acts and the goals they aim for manifest a particular view about what is a good life – or a happy life as Nordenfelt uses the term. He does not equate happiness with mere pleasure or even to preference satisfaction in a crude sense; rather, it is closer to an Aristotelian conception such as flourishing. Satisfaction of the physiological needs necessary for survival (such as needs for oxygen and water) is also necessary of course since satisfaction of these is a necessary condition for the realization of goals such as cinema-going, bird-watching, or even sporting success. So a person has a disability in a specific set of circumstances if, due to an impairment, they are unable to pursue their vital goals.

To give an example to try to illustrate an important implication of Nordenfelt's theory, recall again the example of Oscar Pistorius. Clearly, he appears to be physically fit and healthy and reportedly does not consider himself to be disabled. True he has an impairment of course; but then, so do those with mild hearing loss or shortsightedness. Just as the provision of spectacles addresses the vision of the shortsighted, Pistorius' prostheses remedy his inability to walk. It follows from Nordenfelt's theory that he is not disabled.

A further interesting implication of Nordenfelt's approach is one such as the following. Consider a person with a severe intellectual disability accompanied by severe physical disabilities which have a negative effect on the person's mobility. Suppose further the person lives in a supported environment with 24-h individual care. As it happens the person is a great fan of the cinema and would happily spend up to 4 h per day watching films in their local cinema. This brings so much pleasure to the person that it is reasonable to state that cinema-going is one of their vital goals. The person feels that life is much less rich if they are deprived of the opportunities to pursue their favorite hobby. Happily, due to the provision of 24-h support, they are able to pursue their vital goal of cinema-going. Were the level of support for this person to be reduced to the extent that regular visits to the cinema are not possible, then the person becomes disabled – due to the nature of this change in the "social" environment – namely, the economic decision to withdraw such a high level of support to the person. So it can be seen that in contrast to the "medical model" allegedly found in the ICIDH, Nordenfelt's theory recognizes a clear role for the social environment, broadly construed, as a causative factor of disablement, among other factors, including properties of the individual concerned, such as impairment. In contrast to the ICF, it is sensitive to the personal values reflected in the vital goals of the individual; the emphasis in the ICF is to be found on statistical norms and not personal values.

With reference to the so-called social model of disability (Oliver 1990), if it is interpreted straightforwardly as the view that the cause of disability lies wholly in the social environment, with no causal role acknowledged for factors intrinsic to the person, then there is a clear difference between it and Nordenfelt's theory. However, as Wasserman observes, there is good reason to suppose that "the claim of exclusive social causation is a calculated overstatement, a corrective for the opposing and more damaging misrepresentation [that the cause of disability lies in the individual]" (2001, p. 228). In the early days of the social model, in order to emphasize the role played by the social environment in the causation of disability – especially its physical architecture – it is suggested that causal role was perhaps overemphasized for purely tactical purposes. Also, the social model has been subjected to criticism for its apparent neglect of the significance of impairments, especially the debilitating effects of chronic symptoms of phenomena such as pain and fatigue (Shakespeare 2013; Wendell 2013). The reality of life with some impairments, it is pointed out, is that there are accompanying health problems. So altering the physical environment will leave the significance of those aspects of disablement at risk of neglect. Moreover, as described earlier, the cover of Oliver's influential and important book in 1990 strongly suggests that by modifying the social architecture, apparent

disability evaporates. But critics point out this way of addressing difficulties resulting from disabilities and impairments applies less straightforwardly in conditions such as sensory disabilities and intellectual disabilities. So the social model has itself been subjected to some challenging criticisms.

Criticisms of Nordenfelt have tended to focus on his appeal to vital goals. Critics find a tension between respecting the values of the individual in the formation of vital goals while resisting the collapse into a subjectivist approach which he would obviously reject (1987, pp. 90–91). Nordenfelt appeals to a notion of welfare, which he uses as a synonym for "happiness." But Nordenfelt uses this term in a way that has more in common with flourishing in a purportedly objective sense, and this is perhaps the best way to think of it. Further specification of that brings further problems of course, but it at least indicates that, as far as Nordenfelt is concerned, his theory does not rest upon subjectivism with respect to vital goals.

A further kind of criticism leveled at Nordenfelt concerns the more general theoretical problem of whether "disease" is a normative or nonnormative, purely descriptive concept. This is relevant since as seen above the WHO classificatory schemes regard impairments as consequences of disease and as purely descriptive. The most well-known non-normativist account of health has been developed by Boorse (esp. his 1975). According to him disease classifications are purely scientific descriptions of abnormal functions. So if one defines health as a state involving the absence of disease, wherever there is abnormal function, there will be ill-health – irrespective of the view of the person whose health status is in question. By contrast, in Nordenfelt's theory, the question of whether or not a person is healthy is not independent of the values of that person – as manifested in their vital goals – and the same applies in relation to the question of whether or not they have a disability (Nordenfelt 2001). Of course, while maintaining this, Nordenfelt accepts that some statements about, for example, the incidence of disease are descriptive, e.g., one might observe that there is a high prevalence of a particular disease in a particular region, and that would be a descriptive claim. However, the very identifying of certain physical states as diseases – according to Nordenfelt – is due to their propensity to impact negatively on the capacity of people to be "happy," i.e., achieve their vital goals (2001, p. 78). The same is true with regard to physical states referred to by the term "impairments." So there are these two main positions with regard to the kinds of physical states which, for example, taxonomies such as the ICIDH and ICF claim to be necessary for disability: on a nonnormative approach, they are purely descriptive; on a normative approach, they involve tacit reference to values and so are inherently normative. Nordenfelt's theory is of the latter kind due to the connection – albeit a contingent one – between impairments and ill-health.

Disability and Identity

Several commentators make the point that disability is an intrinsic part of who they are – part of their very identity. For example, Oliver argues against the expression "people with disabilities" because it implies the separability of the disability from the

person; this is not the case he claims because "far from being an appendage, disability is an essential part of the self" (1990, p. xiii; also Toombs 1995). Similar views can be found in the ICIDH (28), though that discussion moves on to distinguish "being [disabled] rather than having [a disability]" (ibid.); the former expression aligns with Oliver's position, whereas the latter clearly implies that disability is a contingent rather than identity-constituting aspect of the person. Also, as Davis explains (2013), a significant strand of disability scholarship has sought to emphasize the view that there is a distinct "disability identity," such that disability really is an identity-constituting feature of the person. This strategy has been exploited in part to fuel what has become known as "identity politics" (see, e.g., Davis, ibid; Siebers 2009, 2013) in which establishing a distinct identity for a group serves as a political expedient to securing increased recognition and respect. This can also help to attack discriminatory practices relating to employment opportunities and health and education. So if it could be shown that disability is on a par with race and gender – accepting for the sake of argument for the moment that these are identity-constituting – then the political strength of those who campaign for the rights of disabled people would be considerably strengthened.

In spite of these proposals, from a philosophical perspective, the claim that disability is identity-constituting does not seem a promising one. When one thinks of people who become disabled due to bodily trauma, it looks more plausible to claim that numerically the same person has undergone a qualitative change albeit sometimes a radical change. The contrary claim – that a numerically different person now emerges – does not sound very plausible. But the "identity claim" as it can be called may be easier to sustain in relation to genetically caused disabilities. If personal identity is considered to involve bodily continuity, then in so far as the disability-causing genetic characteristics figure in that then they might be considered identity-constituting.

However, two kinds of objections have been raised against this view. It has been pointed out that in attaching so much significance to physical properties, the importance of psychological ones is implausibly neglected. What lies behind such objections is the kind of considerations which might lead one to adopt the main alternative approach, relying on psychological continuity. What seems essential to our identity is our values, together with the sense of continuity of existence which we characteristically possess through our memories. Also, it is possible to induce radical changes in a person's physical nature; for example, by using gene therapy genetic makeup can be altered. More radically, one's body parts could very gradually over time be replaced by synthetic substitutes (see Haraway 1991) to the point when one loses all genetic human identity, only psychological continuity. By contrast some commentators have focused on one's origin, but it is fair to say that similar difficulties have been leveled at that: it neglects what is really important in identity and faces an "infinite regress" charge in the search for origins (Edwards 2007).

In light of these kinds of problems faced by attempts to develop theories of personal identity which are constrained by the search for a particular set of essential properties, alternative approaches have emerged which attach less significance to the search for a set of essential properties which are identity-constituting. These

approaches have tended to rely on concepts such as that of narrative identity (Ricoeur 1991; also DeGrazia 2005). They stem in part from work undertaken by MacIntyre in his classic *After Virtue* (1981; see also Schechtman 1996). MacIntyre reminds us that, normally, identity judgments are relatively straightforward and that they rest upon a less strict understanding of the identity relation than that which constrains the philosophical literature on the problem. The suggestion is that the philosophical concept of "strict" identity is a distillation of a more familiar concept which serves its function perfectly adequately, except when required to determine the kinds of problem cases thrown up by the philosophical discussions – e.g., regarding brain swapping, time travel, etc. (see, e.g., Parfit 1985; Williams 1973). If this is accepted, then the kind of work undertaken by, for example, Ricoeur (1991) and Taylor (1989) can be exploited to show how disability might indeed be understood as something more than a mere contingent aspect of the person. This "something more" can then be exploited to support the kinds of identity claims quoted above and thus to sustain the distinction between being a "person with a disability" and a "disabled person."

According to Ricoeur's approach to the problem of self-identity, any plausible theory must include the resources capable of answering the question "Who?" when asked of the person whose identity is in question. It can be seen that answering that question with a proper name is somewhat inadequate unless one already knows something about that person. An adequate answer, by contrast, would be parasitic upon a narrative conception of the self: a conception of the self which possesses the kind of unity characteristic of persons. The unity of a human life can be understood in terms of a narrative – with a beginning, middle, and end. Certain central "structural concepts" provide a framework within which the narrative is filled out. These constitute the form of the narrative which is given content from the life of the particular person. Candidates for such structural concepts would be those such as space, time, embodiment, a "self-conception" (understood as a sense of the kind of person one wants to be (cf. Taylor 1989)), and a "self-project" (van Hooft 1995) where this latter is one's actual attempt to enact or operationalize one's self-conception. Of course, all lives of persons can be understood in relation to these five structuring concepts. But when disability colors the mode in which each of them is experienced, then it may be said plausibly that disability is identity-constituting: that one is disabled as opposed to being a person with a disability.

An example of this kind of approach can be found in the philosophical autobiographical writing of S. K. Toombs (1993, 1995; also Carel 2008, 2013; Merleau-Ponty 1945). She is a philosopher who developed multiple sclerosis and writes philosophically about it. She describes how, due to tiredness and muscle fatigue, the way she experienced space was gradually transformed. Distances hitherto considered short, now were perceived as much greater, thus her conception of space "as lived" became fundamental to her way of being in the world. A similar account is provided in relation to time: activities which previously took little time, such as dressing, became major tasks taking up much more time. With reference to embodiment, as many have observed, when one is healthy, one's body is almost unnoticed, but illness changes that. The body can seem as if it is an obstacle to pursuance of

one's self-projects. In a chronic condition such as multiple sclerosis, Toombs documents how this is experienced by her in terms of the body as lived. With respect to her self-conception and self-project, radical changes were needed to this due to the levels of concentration needed to pursue analytical philosophy at the highest level which were jeopardized by the effects of MS.

So this is one illustration of a theoretical framework in which apparently contingent aspects of one's life can be characterized as identity-constituting. This approach thus makes sense of claims to be a "disabled person" as opposed to a "person with a disability" which are made by others who become disabled in later life.

Definitions of Key Terms

Disability (medical model)	A physical or psychological abnormality, resulting from impairment and associated with disadvantage.
Disability (social model)	A disadvantaged state caused by social factors, possibly in conjunction with the presence of impairments.
Impairment	Physical or psychological state which is associated with reduced function.
Identity-constituting	A characteristic which is fundamental to a person's identity.

Summary Points

- Two main approaches to defining disability can be distinguished, one which focuses primarily on the individual concerned, another which focuses on factors beyond the person such as the social environment.
- In contrast to the theories of disability developed by the WHO, Nordenfelt has developed a theory which defines disability by reference to the values of the person concerned. On such a theory, a person defined as disabled by either WHO taxonomy may not be disabled.
- Some commentators hold that disability is part of their identity, is identity-constituting.

References

Altman BM (2001) Disability, definitions, models, classification schemes, and applications. In: Albrecht G, Seelman KD, Bury M (eds) The handbook of disability studies. Sage, Thousand Oaks, pp 97–122

Boorse C (1975) On the distinction between disease and illness. Philos Public Affairs 5:49–68

Braddock DL, Parish SL (2001) An institutional history of disability. In: Albrecht G, Seelman KD, Bury M (eds) The handbook of disability studies. Sage, Thousand oaks, pp 11–68

Carel HH (2008) Illness. Acumen, Stocksfield
Carel HH (2013) Illness, phenomenology, and philosophical method. Theor Med Bioeth 34:345–357
Davis LJ (ed) (2013) The disability studies reader. Routledge, London
DeGrazia D (2005) Human identity and bioethics. Cambridge University Press, Cambridge
Edwards SD (2007) Disablement and personal identity. Med Health Care Philos 10:209–215
Haraway S (1991) Simians, cyborgs and women: the reinvention of nature. Routledge, New York
Harris J (2000) Is there a coherent social conception of disability? J Med Ethics 26(2):95–100
MacIntyre A (1981) After virtue, a study in moral theory. Duckworth, London
Merleau-Ponty M (1945/1962) The phenomenology of perception (Smith C., trans.), Routledge, London
Nagi SZ (1965) Some conceptual issues in disability and rehabilitation. In: Sussman M (ed) Sociology and rehabilitation. American Sociological Association, Washington, DC, pp 100–113
Nordenfelt L (1983/1997) On disabilities and their classification. University of Linkopping, Linkopping
Nordenfelt L (1993) On the notions of disability and handicap. Soc Welfare 2:17–24, p. 17
Nordenfelt L (1987) On the nature of health: an action-theoretic approach. Reidel, Dordrecht
Nordenfelt L (1995) On the nature of health. Kluwer, Dordrecht
Nordenfelt L (2000) Action, ability and health. Kluwer, Dordrecht
Nordenfelt L (2001) Health, science and ordinary language. Rodopi, Amsterdam
Nordenfelt L (2007) The concepts of health and illness revisited. Med Health Care Philos 10:5–10
Oliver M (1990) The politics of disablement. Macmillan, London
Parfit R (1985) Reasons and persons. OUP, Oxford
Ricoeur P (1991) Narrative identity. Philos Today Spring 1991:73–81
Sacks O (1989) Seeing voices. Picador, London
Schechtman M (1996) The constitution of selves. Cornell Univ Press, Ithaca
Schramme T (2007) A qualified defence of a naturalist theory of health. Med Health Care Philos 10:11–17
Shakespeare T (2013) The social model of disability. In: Davis L (ed) The disability studies reader, 4th edn. Routledge, London, pp 214–221
Siebers T (2009) Disability theory. University of Michigan Press, Ann Arbor
Siebers T (2013) Disability and the theory of complex embodiment-for identity politics in a new register. In: Davis L (ed) The disability studies reader, 4th edn. Routledge, London, pp 278–297
Swain J, Finkelstein V, French S, Oliver M (eds) (1993) Disabling barriers, enabling environments. Sage, London
Taylor C (1989) Sources of the self. Cambridge University Press, Cambridge
Toombs SK (1993) The meaning of illness, a phenomenological account of the different perspectives of the physician and patient. Kluwer, Dordrecht
Toombs SK (1995) Sufficient unto the day: a life with multiple sclerosis. In: Toombs SK, Barnard D, Carson RA (eds) Chronic illness, from experience to policy. Indiana University Press, Indianapolis, pp 3–22
UPIAS (Union of the physically impaired against segregation) (1975) Fundamental principles of disability. UPIAS, London
van Hooft S (1995) Caring, an essay in the philosophy of ethics. Colorado University Press, Boulder
Wasserman D (2001) Philosophical issues in the definition and social response to disability. In: Albrecht G, Seelman KD, Bury M (eds) The handbook of disability studies. Sage, Thousand Oaks, pp 219–251
Wendell S (2013) Unhealthy disabled: treating chronic illnesses as disabilities. In: Davis L (ed) The disability studies reader, 4th edn. Routledge, London, pp 161–176
WHO (2001) The International classification of functioning, disability & health. WHO, Geneva
Williams B (1973) Problems of the self. Cambridge University Press, Cambridge
World Health Organization (WHO) (1980/1993) International classification of impairments, disabilities and handicaps. WHO, Geneva

Subjective and Objective Accounts of Well-Being and Quality of Life

12

Thomas Schramme

Contents

Abstract

The chapter aims to provide a classification of different philosophical theories of well-being. A very common issue of contestation is whether well-being is subjective or objective. However, ontological and evaluative perspectives in this regard need to be disentangled. The ontological perspective is concerned with the problem whether well-being is a mode of consciousness or of existence. The evaluative perspective focuses on the criteria of well-being. There are then altogether four different accounts: (i) experience theories (ontological subjectivism), (ii) state-of-being theories (ontological objectivism), (iii) desire-fulfillment theories (evaluative subjectivism), and (iv) essence theories (evaluative objectivism). This classification is applied to a particular philosophical and social dispute, namely, whether and, if so, in what way disability undermines the quality of life of persons with disability.

T. Schramme (✉)
Department of Philosophy, University of Liverpool, Liverpool, UK
e-mail: t.schramme@liverpool.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_7

Introduction

The concept of well-being refers to conditions of persons and their circumstances. It is a gradual notion that allows for different levels. An individual level of well-being is determined in relation to the extent to which all elements that are good for a person are accessible and enjoyable to a person. The elements of well-being are also called prudential goods or values, as they refer to what should be pursued for prudential reasons, as opposed to moral reasons, in order to live a good life (Griffin 1986).

Since well-being is one important aim of medical practice, it is of significance how the concept is understood. One of the most important contested aspects about the notion of well-being is whether its elements are subjective or objective. If they were objective, then it seems that what is good for patients can be determined without taking their individual perspective into account. Yet, as will become clear in this chapter, it is not clear from the outset what exactly subjectivity and objectivity in relation to well-being mean. Here, different ways how well-being may be subjective or objective need to be distinguished.

In recent times, "quality of life" has become a more common term in medicine, but is also used in philosophy (Nussbaum and Sen 1993). On first sight, it has a more objective ring because it refers to elements of life, especially circumstances that are factual, not to the experienced quality of such elements, which seems to be the main reference point for the concept of well-being. Indeed, in medicine and related areas, metrics to measure quality of life have been introduced and widely used. Still, these metrics often contain subjective evaluations of the conditions of persons' lives and of their health status.

In philosophy, there are still other terms that have some relation to the concepts of well-being and quality of life. In ancient philosophy, the idea of a good human life was referred to by the Greek term *eudaimonia*, which has often been translated with "happiness," though this notion again carries heavy subjective connotations – it seems to refer to a mental state – whereas *eudaimonia* was, at least in several ancient accounts, an objective notion. It has therefore become more common to use the term "flourishing" as translation.

Still another term, which is more common in economics, is "welfare" (Rescher 1972). It is arguably more narrowly concerned with the quality of the circumstances of life, not with how far these external conditions are enjoyable to a person. But, be that as it may, the different mentioned concepts – well-being, quality of life, welfare, the good human life, *eudaimonia*, and flourishing – are somewhat related and occasionally used interchangeably, despite the mentioned differences.

The Philosophical Debate on Theories of Well-Being

It has become common for philosophers to start debates about human well-being with a tripartite distinction of theories that Derek Parfit provided in an appendix of his book *Reasons and Persons* (Parfit 1984, p. 493ff.). According to this description, there are three kinds of theories: hedonism, desire-fulfillment theories,

and objective list theories. Thomas Scanlon (1993) replaced the third category by substantive good theories. As will become clear later in some respects, this is an inadequate distinction. Still, it is a valid starting point for discussions of well-being.

Hedonism is an account that has a long pedigree and was for a long time the dominant point of view. It was explicitly endorsed by Epicurus (1940), who held that something can only be good for us if it is experienced, and later on by Bentham (1789) and Mill (1861), the founding fathers of utilitarianism. Pleasure and the absence of pain are, according to hedonism, the only constitutive elements of human well-being or happiness. What causes pleasure can of course vary, though whether something indeed causes pleasure in us is not up for us to decide. Desire-fulfillment theories were developed in order to deal with theoretical problems of hedonism that have to do with its lack of "worldliness," as it were. Hedonism allows for something to be of prudential value simply because it pleases, for instance, when someone enjoys a hot bath; however, there can be pleasures that are not at all based on reality, and they are merely imagined or artificial, for instance, due to drug abuse. It would not normally be said that something can be good for someone if it is not real.

Desire-fulfillment theories deal with this proviso by basing prudential value on facts about the world, namely, whether a desire is fulfilled or not. Again, this allows for individual varieties of desires, but the criterion of something being good for us simply is that it fulfills a desire. Many modern utilitarian theorists as well as many economists believe in this theory. As it was just described here, it has been deemed in need of qualifications though. It seems that people often desire strange and irrational things, such as remembering the first 500 digits of the number pi by heart, which then calls into doubt whether it would actually be good for them if such a desire were fulfilled. The desire-fulfillment theory has accordingly been developed to include certain normative criteria that every desire has to meet, such as being voluntary, informed, and rational.

Objective list theories take a different stance; they deem things good for us because they are objectively good, meaning that their value is not due either to our subjective experiences or desires. Objective list theories also go back to ancient philosophy; Aristotle is regarded as its founding father.

Parfit's distinction has become influential, and it is therefore not surprising that the most general distinction philosophers usually draw between different theories of well-being is to classify them into subjective as opposed to objective accounts. Hedonism and desire fulfillment are regarded as subjective in contrast to the objective list theory. It will be seen shortly, though, that there is an important ambiguity in the subjective-objective distinction. But it is nevertheless helpful to use this general distinction. An element of our well-being can accordingly be either due to some subjective stance or pro-attitude of a person, or it can be good for her without her appreciating it as such. In the former case, it is due to the person herself what constitutes her well-being, so is in a way her invention; in the latter case, it is something that can be discovered. A fitting way to describe the distinction is therefore to call the subjective account "taste model" and the objective account "perception model" (Griffin 1996, p. 20 ff.). Either something is valuable because it is desired (taste model) or something is desired because it is valuable (perception

model). It is obvious that within the taste model, the elements of human well-being will differ widely, and it is this very fact that makes it plausible, as people indeed differ in their tastes and what they regard as good for them. Still, there might be elements of human well-being that are universally valued and which could for this reason be called objective values – though in this case it would be more adequate to call them intersubjective values. Intersubjectively valued elements of human well-being are not the same as objective ones, because the latter are not up to our choices – or tastes for that matter. It is the source of value that is important for this distinction, and for an objectivist about well-being the source is not to be found in individual persons but in objective facts about humans more generally, or indeed external sources, such as a deity or a natural order. A good example is health. Many theorists maintain that health is an element of human well-being and that this fact is not due to our evaluation. Thus, health is regarded as valuable in itself; it is good for us in virtue of facts about ourselves. Other examples that philosophers deem objective elements of well-being are, for example, friendship and knowledge.

Subjectivity and Objectivity of Well-Being

The most foundational contentious issue between different theories of well-being is whether individual well-being can be determined objectively or only subjectively. Subjective theories seem to have won the day right from the start, though, because it seems obvious that something can only be prudentially valuable if it is valuable *for* someone. This implies a subject-relativity of the prudentially good. Yet, it does not imply that something can contribute to individual well-being only because persons themselves judge it to be good (Darwall 2002), and this aspect – concerning the source of prudential value – is the contentious issue. Still, the mentioned subject-relativity of the notion of well-being seems to be speaking in favor of subjective accounts (Sumner 1996).

To avoid cross-purposes, it is important to sort out different ways of speaking about the subjectivity or objectivity of well-being. First, there can be theories that explain individual well-being by subjectively experienced mental states. These theories are in contrast to approaches that see well-being determined by elements that are independent of individual feeling, such as material resources. This kind of quarrel is concerned with the question whether human well-being is wholly or foremost explained by subject-relative mental states or not. Here, the issue concerns what kind of condition or state well-being is: a mode of consciousness or of existence. Accordingly, the debate is concerned with ontological subjectivity or objectivity.

Second, theories of well-being might see the relevant elements as either depending on subjective evaluations or assessments. In contrast, they can put forward objective or intersubjective determinants. In this case, the quarrel is concerned with the criteria of well-being and whether it is evaluatively subjective or objective.

In both cases, relating to ontological or evaluative subjectivity or objectivity, there can also be a discussion as to whether persons themselves can be regarded as the best or even unfailing judges of their own well-being or whether one can

actually be wrong about one's own well-being. This concerns epistemic aspects. Yet this way of understanding objectivity and subjectivity does not require a classification in its own right, as it is not concerned with determining the elements of well-being but with our knowledge about these elements.

In addition to the already confusing terminology, there is a claim found in empirically minded social psychology whereby scientists measure objective happiness or well-being (Kahneman 1999). It is true that given a certain metric of well-being – that is, by assuming specific components of the human good – the level of individual well-being can be assessed objectively. But here the account of well-being might be, and often is, determined by subjective factors, as the measured elements of well-being are either of a mental kind or particular, generally assumed preferences.

By using the twofold distinction between ontological and evaluative forms of theories of well-being and adding the two variants of subjectivism and objectivism, four types of theories are gained. There are no established names for all of these, so the following labels are partially new inventions: (i) experience theories (ontological subjectivism), (ii) state-of-being theories (ontological objectivism), (iii) desire-fulfillment theories (evaluative subjectivism), and (iv) essence theories (evaluative objectivism).

According to the introduced classification, the first version of accounts of well-being is experience theories. These are ontologically subjective, and obviously hedonism is the best-known variant of this version. The second group of theories, state-of-being theories, are ontologically objective as they focus on objective facts of the life of human beings or their life circumstances, including such things as resources, opportunities, or income, generally speaking items that John Rawls called primary social goods (Rawls 1971, p. 78 ff.). The third variant is the well-known and widespread desire-fulfillment theory. Well-being, for its proponents, is a life according to one's own desires and ideals. It is therefore an evaluatively subjective approach. Finally, there are essence or genus theories that base well-being on an objective standard of basic or necessary elements of the good human life. This is an evaluatively objective approach. Aristotle, for instance, assumed a teleological structure of human striving. *Eudaimonia*, for him, is determined by an activity in accordance with the specific ergon (function) of human beings. Modern Aristotelian essence theories usually reject teleological considerations, but agree that what is good for human beings is determined by matters regarding the human form of life (Nussbaum 1995).

When considering the mentioned distinction between different ways of accounting for the objectivity or subjectivity of well-being, Parfit's mentioned classification does not seem the most fitting anymore. He restricts objective theories to evaluatively objective ones, which are then additionally required to list items – supposedly constitutive elements of well-being – though this seems hardly true even for a theory that most clearly falls into this category, namely, Aristotle's. He rather gives a general description of *eudaimonia*, which is roughly to live in accordance with the virtues. So a list alone does not add up to a theory of human well-being, unless it contains a criterion, which explains why the listed elements are on the list. In addition, Parfit does not allow for ontologically objective theories, and

he restricts ontologically subjective theories to hedonism, though there can certainly be alternatives, which might not see pleasure as the decisive conscious experience (cf. Kagan 1992), but, for instance, religious experiences.

If taking the fourfold classification into account, hedonism is actually difficult to categorize. It is ontologically subjective because pleasure is a subjective experience. Prudential value here has to be instantiated in the conscious experience of a person. Pleasure, hence hedonistic value, only exists subjectively. However, what is pleasurable to us apparently is not up to us to decide, but due to our nature. Again, this can vary; there can be individual "natures" of individual persons, but there might also be universal pleasures, for instance, the pleasures of sex and eating. It cannot always be simply decided what to find pleasurable. It seems therefore to be, at least partly, a matter of discovery what is good for us according to hedonism, and consequently in this respect it would be wrong to call it a subjective account. In fact, it might be more adequate to call it an objective list theory with one element on the list, namely, pleasure (and possibly absence of pain as a second element).

The main points of introducing this classification are first to stress that there is no single or unified approach to human well-being, and hence no clear single basis for assessing elements of human life in respect to well-being, and second to provide a rational basis for evaluation that does not beg the question regarding a standard of assessment. In the next section, it will be illustrated, by virtue of using the pertinent example of disability, how different philosophical theories of well-being lead to different assessments of the impact of disability on the quality of life of people with disabilities.

Different Theories of Human Well-Being in Practice

Philosophical theories of well-being determine assessments of certain conditions of human life as good or bad for human beings. In order to give more substance to these theories, in this section one particular example of a medical condition will be used to see how differences in theory might pan out in practice.

According to experience theories, disability is bad for the person if it is experienced as something bad, for instance, because it is painful. It is obvious that on such a basis, there can be no generalization regarding the badness of disability; it cannot be said that disability is always bad for the disabled person. If a person does not experience it thus, there is no basis for saying that her well-being is impaired by a disability.

A possible rejoinder is the "disability paradox," which is due to the fact that people normally disvalue disability but, if they become disabled, report a relative high level of subjective well-being after a while. A variant of this rejoinder can be found in John Stuart Mill's discussion of the "satisfied fool" (Mill 1861, p. 212), which states that people might be content with being cognitively and intellectually "foolish," but that this does not mean that they are deemed to be happy. The rejoinder accordingly insists that people sometimes experience something as pleasurable, or fail to experience it as harmful, due to the circumstances, such as being mentally adapted to a harmed state of being. This thesis indeed challenges the

adequacy of experience theories as such, not merely the assessment of disability on its basis. The challenge implies that human well-being cannot only be based on ontologically subjective elements, but requires a standard of how people should normally experience certain things or aspects of their life. On grounds of experience theories alone, it seems nevertheless unlikely that it can be claimed that disability is always bad for the disabled person.

On the basis of state-of-being theories, it can be argued that disability is bad, because it prevents opportunities, leads to lower income, or generally hinders one in the pursuit of resources. It is especially the value of opportunities or options and the claim that disabled people lack a normal range of opportunities (Daniels 1985, p. 27 f.) that heavily influences the discussion about the evaluative assessment of disability. It seems obvious that indeed disabled people have fewer options within their lives than nondisabled people. The very term "disability" already contains the notion that there are less abilities, hence less options, for people with disabilities. Therefore, disability seems clearly and in every instance a harmed condition. But there are at least two objections to this straightforward view. Firstly, the theory itself is less convincing than it might seem. State-of-being theories focus on resources and generally the external life circumstances of a person, but not on the significance resources have in the life of persons. It can lead to a kind of "resource fetishism" (Sen 1980, p. 216) that tends to ignore the individuality of human beings. Disabled people might have exactly those opportunities and resources they deem of significance. Whether they could have more options if they were nondisabled might not be important to them. To insist that they are nevertheless harmed seems to ignore their individual point of view. This of course should not be read to imply that disabled people are always happy with the options they have. Still, the opposite generalization can be challenged. Secondly and relatedly, many restrictions in the opportunity range of disabled people are due to environmental conditions – hence societal decisions – not to the physical or mental condition of a person herself. So even if there is harm involved in the state of being of a person, it is not always due to the disability, and the harm can actually be overcome. This is of course a point that has been put forward on several occasions by proponents of the social model of disability.

Desire-fulfillment theories result in a view that disability is bad, if it is undesired. Again, there can be no generalization regarding the assessment of disability, as it is based on individual desires. There is also a rejoinder similar to the one mentioned in relation to experience theories, which can be applied to the assessment of disability. Empirical research shows that people often adapt their preferences according to their real situations. If they cannot reach a certain good they normally desire, they might solve the problem by adapting their preferences accordingly, like the fox in the well-known fable, which dismisses the grapes he cannot reach as inedible. The rejoinder regarding adaptive preferences can therefore be called the "sour grapes" objection. One could then claim that people with disabilities who assert not be harmed fall foul of a similar sour grape preference, or rationalization, as it is often considered. However, just as in the case before, this objection points at a general problem of the theory under scrutiny. It seems that it is based on an implicit standard of what should be desired and strived at by human beings.

The repeatedly mentioned reference to standards of normal happiness or desire fulfillment speaks in favor of a nonsubjective theory of well-being. Essence theories offer such an evaluatively objective account. According to this perspective, disability is bad, because it prevents or restricts basic elements of the good human life, for instance, self-determination, liberty, and independence. It is indeed plausible to say that many people regard restrictions of disabilities as bad because they see them as undermining basic human capacities, such as seeing, hearing, speaking, walking, abstract thinking, remembering, and so on. So they seem to endorse an essence theory. This ties in also with the fact that many people do not see minor impairments, such as a missing finger or lack of imagination, as disabilities, although they can have a considerable impact on some people's lives. But even if common sense seems to be in congruence with essence theories, this does not mean that they share the same philosophical basis. Common sense seems to be just that – a sense that is common – whereas essence theories would need an argument as to why certain characteristics of human beings are necessary requirements of the good human life. This is much harder than identifying a widespread agreement and what people find important. Indeed, many philosophers would see it as a hopeless task, as it seems to commit an is-ought fallacy by determining what is required for a good human life through an account of human nature. Yet only normative argument would allow to reject as wrong the judgments of people with disabilities who challenge the received view and do not see their disability as harmful. Otherwise these people would simply have an unusual point of view that clashes with common sense. But even if such a philosophical theory were to succeed, it will only bear on severe impairments, i.e., conditions that restrict essential elements of the good human life, and will not allow for a sweeping claim according to which disability is always incompatible with the good life of human beings.

In this section, disability has been an example of a putative impairment of well-being to give further substance to the four different types of theories introduced above. All accounts provide reasons to call disability an instance of harm, but none seems to allow for a general claim about the impact of medical impairment on the quality of life of individual human beings. There cannot be a straightforward identification of disability and harm on the basis of these theories. This shows that such theories of well-being need further discussion and also that there is usually no straightforward evaluation of individual conditions or circumstances of human lives in terms of these theories. Still, they form the backdrop, often implicit, of many real debates about quality of life and well-being.

Conclusion

The philosophical debate on well-being has important practical repercussions. How people conceive of the prudentially good for human beings determines their assessments of the quality of life of people and eventually their ethical judgments. In this chapter, a fourfold classification of theories of well-being has been introduced: (i) experience theories (ontological subjectivism), (ii) state-of-being theories

(ontological objectivism), (iii) desire-fulfillment theories (evaluative subjectivism), and (iv) essence theories (evaluative objectivism). When applying this classification to concrete cases in order to determine assessments of the quality of life of specific groups of patients, it can be seen how these depend on underlying accounts of well-being. This has been illustrated by using the apparently straightforward case of disability.

Definition of Key Terms

Desire-fulfillment theories	A group of philosophical theories of human well-being that see the fulfillment of differently qualified desires as constitutive of well-being. It is an evaluatively subjective approach.
Essence theories	An evaluatively objective group of theories that put forward objective criteria of human well-being.
Experience theories	An ontologically subjective group of theories that claim well-being to be a mode of conscious experience.
Hedonism	A group of philosophical theories of human well-being or happiness that sees pleasure and the absence of pain as the only constitutive elements.
Objective list theories	A group of philosophical theories of human well-being that claims particular elements of a life to be objectively good, for instance, because they belong to a genuinely human life form.
Perception model	Sees elements of well-being as valuable in their own terms.
Prudential goods/values	Things that are good for us in terms of our interests, not morally.
State-of-being theories	An ontologically objective group of theories that claim well-being to be a mode of a human life, including its circumstances.
Taste model	Sees elements of well-being as valuable because they are desired.
Well-being	The extent to which all elements that are good for a person are accessible and enjoyable to a person.

Summary Points

- Well-being is a term that refers to conditions of persons and their circumstances. The elements of well-being are also called prudential goods.
- The notion of quality of life is mostly used in medical contexts.

- The philosophical debate on well-being is mainly concerned with the problem whether it is subjective or objective.
- An ontological and an evaluative perspective need to be distinguished. The ontological perspective is concerned with the problem whether well-being is a mode of consciousness or of existence. The evaluative perspective focuses on the criteria of elements of well-being.
- There are different theories of well-being that are pursuing these different perspectives and develop either subjective or objective accounts.
- These theories form the backdrop of variations in the assessment of the quality of life of patients.

Acknowledgments This chapter draws on material that has previously been published in the article "Comparative and Non-Comparative Perspectives on Disability" in Jerome Bickenbach, Franziska Felder, Barbara Schmitz (eds.), Disability and the Good Human Life, Cambridge University Press 2014, 72–92.

References

Bentham J (1789) An introduction to the principles of morals and legislation. Oxford University Press, Oxford, 1996

Daniels N (1985) Just health care. Cambridge University Press, Cambridge

Darwall S (2002) Welfare and rational care. Princeton University Press, Princeton

Epicurus (1940) Letter to Menoeceus. In: Whitney JO (ed) The stoic and epicurean philosophers. The Modern Library, New York, pp 30–34

Griffin J (1986) Well-being. Its meaning, measurement, and moral importance. Clarendon, Oxford

Griffin J (1996) Value judgement: improving our ethical beliefs. Clarendon, Oxford

Kagan S (1992) The limits of well-being. Soc Philos Policy 9(2):169–189

Kahneman D (1999) Objective happiness. In: Kahneman D, Diener E, Schwarz N (eds) Well-being: the foundations of hedonic psychology. Russell Sage, New York, pp 3–25

Mill JS (1861) Utilitarianism. In: Robson J (ed) Collected works. Toronto University Press 1969, Toronto, pp 203–255

Nussbaum M (1995) Human capabilities, female human beings. In: Nussbaum M, Glover J (eds) Women, culture, and development. Clarendon, Oxford, pp 61–104

Nussbaum M, Sen A (eds) (1993) The quality of life. Clarendon, Oxford

Parfit D (1984) Reasons and persons. Clarendon, Oxford

Rawls J (1971) A theory of justice, rev edn 1999. Harvard University Press, Cambridge, MA

Rescher N (1972) Welfare: the social issues in philosophical perspective. Pittsburgh University Press, Pittsburgh

Scanlon T (1993) Value, desire, and quality of life. In: Nussbaum M, Sen A (eds) The quality of life. Clarendon, Oxford, pp 185–200

Sen A (1980) Equality of what? In: McMurrin S (ed) Tanner lectures on human values 1. Cambridge University Press, Cambridge, pp 197–220

Sumner LW (1996) Welfare, happiness & ethics. Clarendon, Oxford

Pain as a Subjective and Objective Phenomenon

13

Wim Dekkers

"Pain is as elemental as fire or ice. Like love, it belongs to the most basic human experiences that make us who we are."
(Morris 1993, 1)

Contents

W. Dekkers (✉)
IQ Healthcare, Radboud University Medical Center, Nijmegen, The Netherlands
e-mail: wim.dekkers@radboudumc.nl

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_8

Abstract

Pain belongs to human life. Although pain is not a medical phenomenon per se, reflections on pain touch upon the philosophical foundations of medicine. Pain confronts us with basic questions such as the tension between an objective and a subjective approach, the concept of brain disease, human consciousness, and the relationship between body and mind. In this contribution, pain is placed in the context of the philosophy of medicine. Attention will be paid to some basic medical-biological and behavioral theories about pain and their underlying presuppositions. For about four decades, several holistic approaches of pain have existed. It appears that the meaning of pain is hard to understand from a scientific perspective. Pain is a sensory and emotional experience, the quality of which is difficult to express in words. Pain is a mystery; it cannot be explained as having just a signaling function. It has also an ontological and an existential dimension.

Introduction

Pain belongs to human life. Since the beginning of its existence, mankind in all its historical and cultural diversities has been suffering pain, and there are no signs that this will ever change. Every human being knows what pain is based on his or her own experience. Pain is not a medical phenomenon per se, though it is still a key element in the practice and theory of medicine. The relief of suffering is considered one of the primary ends of medicine (Cassell 1991). Reflections on pain touch upon the philosophical foundations of medicine. Pain confronts us with basic questions such as the goal of medicine, the tension between an objective and a subjective approach, the concept of brain disease, human consciousness, and the interaction between physical and mental determinants, that is, between the body and mind.

It is difficult to provide an adequate definition of pain, a fact that is reflected in the clinical wisdom that "pain is what the patient states it is." Nevertheless, the International Association for the Study of Pain (IASP) describes pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (IASP 2014). This definition is widely accepted and often quoted. It emphasizes that pain is not only a sensory but also an emotional experience and that it can occur without tissue damage.

The word "pain" and related terms are derived from the ancient Greek *poine* and the Latin *poena*, which mean "penalty" or "punishment." From this etymological perspective it can be seen that the original meaning of the word pain must be traced outside the medical context. Since time immemorial, the language of pain has been embedded in a much broader context in which cultural, philosophical, and religious factors play an important role. Pain has attracted the attention of a wide diversity of disciplines inside and outside medicine, ranging from basic biomedical sciences to behavioral and cultural sciences and to philosophy and theology.

Pain is a complex phenomenon. The French surgeon René Leriche (1937) insisted upon the difference between pain as being studied in physiological and psychological laboratories and pain as being experienced in daily life. He claimed that pain studied in a laboratory bears little resemblance to "living pain," that is, pain as it is experienced by individual patients encountered by practicing doctors. A distinction is often made between acute pain and chronic pain. Acute pain is of recent onset and probably limited duration. It usually has an identifiable temporal and causal relationship to injury, disease, or medical intervention. Chronic pain commonly persists beyond the time of healing of an injury, while there may frequently not be any clearly identifiable cause (Loeser 1991). Moreover, chronic pain is often difficult to treat.

For about six decades there has been an increasing interest in pain and its treatment. Due to the human life span becoming increasingly longer, people suffer from chronic diseases such as cancer, arthrosis, and rheumatoid arthritis, often painful diseases. After centuries in which medicine was quite powerless to treat pain, a new era with potent pain killers has started. The first multidisciplinary pain team was established in 1948 by the American anesthesiologist John Bonica. Pain medicine has emerged as a clinical speciality since the 1950s. Clinical studies took into account the need for attention to chronic pain as well as acute pain and argued for a multidisciplinary approach to research and treatment. This transition involved a shift from the laboratory to the clinic and included the creation of a "world of pain," which would be addressed by teams of experts in specialist pain centers, whose interests were quickly to fragment into highly specialized subfields concerned with acute pain, terminal pain, cancer pain, chronic neuropathic pain, chronic postoperative pain, etc. (Baszanger 1998).

Another important development is the emergence of hospice medicine and palliative care in the 1970s. The World Health Organization (WHO) describes palliative care as "an approach that improves the quality of life of patients and their families facing the problem associated with life-threatening illness, through the prevention and relief of suffering by means of early identification and impeccable assessment and treatment of pain and other problems, physical, psychosocial and spiritual" (WHO 2014). The total pain concept as developed by Cicely Saunders, one of the founders of the modern hospice movement, has been particularly influential.

Pain and suffering are closely related, since pain is the most commonly considered cause of suffering. People in pain frequently report suffering from pain when they feel out of control, when the pain is overwhelming, when the source of the pain is unknown, when the meaning of the pain is dire, or when the pain is chronic (Cassell 1991). However, although the terms "suffering" and "pain" are generally coupled in the medical literature, they are phenomenologically distinct forms of distress (Pullman 2002). Pain does not necessarily involve suffering (in a pregnant sense), for example, when someone cuts himself with a knife. A woman may experience the pain of childbirth as severe yet "rewarding." Hence, she would not describe the pain experience as an experience of suffering. Conversely, it is possible to suffer without experiencing physical pain, for example, when someone suffers from homesickness, a depressive mood, or anxiety. A patient with an injured spinal

cord might suffer because of the loss of bodily functions, even though physical pain is absent or well managed. Pain can be considered a perceived threat or damage to one's biological integrity. Suffering is the perception of a serious threat or damage to one's personal integrity. It is a specific state of severe distress induced by the loss of integrity, intactness, cohesiveness, or wholeness of the person or by a threat that the person believes will result in the dissolution of his or her integrity (Cassell 1995; Chapman and Gavrin 1999).

The aim of this contribution is to place pain in the context of the philosophy of medicine and to present the most important philosophical aspects of pain. Attention will also be paid to some basic scientific theories about pain. Because suffering is also the topic of another article in this *Handbook of the Philosophy of Medicine*, the focus here is rather on physical pain than on the broader phenomenon of suffering.

Two Opposing Theories

Pain has been the object of fundamental laboratory research as well as practical clinical studies. From the beginning of scientific inquiry, there have been many diverse ideas about the neural mechanisms underlying pain. Scientific concepts about pain have emerged, evolved, and changed over time, dependent on cultural influences and philosophical presuppositions. The conceptual changes show that many current controversies have old roots (Perl 2011). In particular, two opposing views on the medical-biological explanation of pain exist: the *specificity theory* and the *pattern theory* (Cervero 2009).

The *specificity theory* proposes that a specific pain system carries messages from pain receptors in the skin to a pain center in the brain. Pain is considered to be a sense similar to sight, hearing, smell, taste, and touch, that is, a component of the sensory system that provides the brain with accurate information about injuries, warns us of impending damage, and helps us to heal. The specificity theory maintains that there are elements of the peripheral and central nervous system specifically and exclusively dedicated to the processing of pain-related information. The French philosopher René Descartes is often mentioned as the one who gave the first classical description of this theory. His observations on pain, illustrated by the famous drawing of the kneeling boy by the fire in his *Traité de l'homme* (Descartes 1963), have been an important point of reference in medical writings. His reputation for inaugurating a reductionist and dualist conception of pain seems to rest especially on this drawing. It is often said that Descartes conceived of the pain system as a simple straight-through channel from the skin to the brain.

It is, however, inaccurate to see the drawing of the kneeling boy by the fire as representative of Descartes' understanding of pain. His more detailed remarks about pain in his *Méditations touchant la première philosophie* suggest a more complex view. In the sixth meditation he argues that pain confronts us with the fact that we do not just *have* bodies, but that we *are* our bodies (Descartes 1967; Van Dijkhuizen and Enenkel 2009). On a general level, Descartes is rightly seen as a dualist philosopher, arguing that the body and mind are essentially two entirely different substances.

In this sixth meditation, however, he seems to refer to a non-dualist relationship between the body and mind, something that was later on explicitly brought forward by twentieth-century phenomenologists such as Gabriel Marcel and Maurice Merleau-Ponty. Descartes' writings also demonstrate that the phenomenon of pain plays a central role in his philosophical thoughts, particularly in one of the key modern reflections on the mind-body problem.

The *pattern theory* of pain denies that pain is just a sense, such as sight and hearing. It attaches to both pain and its opposite, pleasure, fundamental roles in shaping emotions and behavior. From this point of view, pain is considered a trigger of emotional states, a behavioral drive, as well as a highly effective learning tool. It is often mentioned that the pattern theory of pain dates back to Aristotle. According to Aristotle, pain and its opposite pleasure are not sensations but emotions, that is, "passions of the soul." In his theory, to have an emotion is to experience pain, pleasure, or both, this pain or pleasure being intentional and representational. Pain and pleasure also play a crucial role in his virtue ethics, especially in his doctrine of the mean, that is, the idea that virtue is the mean between two extremes (Aristotle 1976).

The choice between the specificity theory and the pattern theory is not just an academic question: it has implications for the experimental paradigms used to study the nervous system. If pain is regarded as a sense, one will look for sensors that are selectively activated by painful stimuli and for sensory pathways in the brain and spinal cord that carry pain information in the same way as photoreceptors in the retina and visual pathway to the cortex are identified. If pain is not considered a sense such as sight, hearing, etc., then there is no need to look for specific neural pain mechanisms. If pain is a "passion of the soul," one will need distributed networks and parallel interactive processing, rather than a specific pain pathway.

Scientific Approach

In his standard study *The Puzzle of Pain*, pain researcher Ronald Melzack (1973) focuses on pain as a neuroanatomical and neurophysiological phenomenon. His purpose was to explain the phenomenon of pain in a biological-scientific way. Melzack's gate control theory has influenced medical theories on pain for many years. Basically, the theory proposes that a neural mechanism in the dorsal horns of the spinal cord acts like a gate, which is able to increase or decrease the flow of nerve impulses from peripheral fibers to the central nervous system. Somatic input is therefore subjected to the modulating influence of the gate before it evokes pain perception and response. According to Melzack, the key to the puzzle of pain was thought to lay in medical-biological knowledge.

The gate control theory of pain is a good example of a pattern theory. Correlates of higher brain processes such as attention, anxiety, anticipation, and past experience exert a powerful influence on pain processes. However, this comprehensive pain theory still contains certain elements of the specificity theory, for example, the function and activity of various peripheral nerve fibers involved in pain processes.

Although the central nervous system plays a crucial role in the gate control theory, this theory is not able to explain some severe chronic pain problems that require a greater understanding of brain processes. Phantom limb pain, for example, occurs in the absence of a specific limb, that is, in the absence of peripheral nerves and other structures that underlie nociception. Therefore, a revised theory has been developed that conceives of a so-called neuromatrix that extends throughout selective areas of the whole brain (Melzack 1996). In this theory, the brain, in particular the neuromatrix, can generate painful sensations on its own in the absence of peripheral input. The origins of phantom limb pain are thought to lie in the brain. When someone loses a hand, he or she experiences a phantom hand because the central representation of the hand remains intact.

Pain is not only a medical-biological phenomenon. Many authors argue that we can explain the phenomenon of pain only by placing it in the context of the whole human existence. A well-known example of such a holistic approach is the pain model developed by the American pain specialist John Loeser in the 1970s. The key tenets of this model are identical to the so-called biopsychosocial model, which has also been influential for the theory and practice of medicine (Engel 1977). Loeser developed his pain model in order to describe "the universe of pain" via four nested circles that identify four components of pain. (1) The physiological basis of pain is *nociception*, that is, the process whereby noxious stimuli in case of tissue damage are transmitted and further modulated by specialized transducers to specific pain fibers. (2) The *perception of pain* is the awareness of a painful event, frequently triggered by a noxious stimulus such as an injury or a disease. Pain can also be generated without nociception, when the peripheral or central nervous system has been damaged. (3) *Suffering* is a negative personal response induced by pain, but also by fear, anxiety, stress, loss of loved objects, and other psychological states. (4) *Pain behavior* results from pain and suffering and consists of the things a person does or does not do as a response to experienced pain. This model has been heuristically useful and has been appropriated in diverse fashions (Loeser and Melzack 1999). It reminds us of the fact that "nerves exist in a patient, who is, first and foremost, a human being and not just a biological machine" (Loeser 2000, S2).

Melzack's gate control theory and Loeser's model are well-known examples of attempts to explain pain in a scientific way and to lay the ground for a scientifically based treatment of pain. Diagnostic and therapeutic interventions of physicians are based on a biomedical explanation of the pain symptom and/or a behavioral understanding of the pain complaint. This approach is based on the conviction that all illnesses and other pathological processes can be explained by the laws of the biomedical and behavioral sciences. However, this conviction can be regarded as one of the reasons why pain continues to be a problem for the practice of medicine. Pain does not entirely conform to the scientific, in particular, biomedical approach of health and disease. It is argued that pain is the principal reason for patients to go to their physicians, while pain is routinely undertreated in health care (Resnik et al. 2001).

Total Pain

Pain is one of the most common and distressing symptoms described by cancer patients. One of the reasons to initiate modern hospice care was a critique of the reductionistic way in which medicine in the mid-twentieth century dealt with cancer pain. In palliative care, the clinical management of patients suffering pain from advanced cancer is paramount. Innovations in this field can be summarized in three points: (1) the development of a patient-centered approach to analgesic evaluation, which resulted from the search for an alternative analgesic to morphine, (2) the reintroduction by John Bonica of the idea that pain is what the individual feels and thinks it is, and (3) the work of Cicely Saunders in establishing the foundations of the modern hospice and palliative care movement. The work of these clinicians must be considered in the context of their time, when new hopes emerged that cancer could be cured and, at the same time, the cancer patient began to be remolded from a passive participant in treatment and care to an active collaborator (Seymour et al. 2005).

Since the 1970s there has been a widening of interest from acute, mainly postoperative pain to the question of chronic pain. Some of this new interest focused on the apparent purpose or function of pain. In an acute context, pain is seen as functional in drawing attention to injury or as an indicator of the need for rest and recuperation. By contrast, chronic pain in advanced cancer appears dysfunctional and without adaptive purpose, while posing particular challenges at the level of meaning. The lack of clarity in how to understand pain for cancer patients contributed to the persistence of poor pain management. It was partly because of this that the study of chronic pain became open to influences from the human sciences (Clark 1999). It is in this context that Cicely Saunders in the 1960s developed the idea of a holistic approach of pain. In the course of her clinical work with dying people, she coined the term "total pain," one of the most powerful concepts of the modern hospice movement, to characterize the multidimensional nature of the patient's pain experience and to include the physical, psychological, social, and spiritual domains of pain (Saunders 1967; Clark 1999). The combination of these elements is believed to result in a "total pain" experience that is individualized and specific to each patient's particular situation.

The complexity of treating patients with total pain is often compounded by the patients' inability to distinguish exactly which component is causing pain, because all they can express is that "they just hurt." Patients may not be capable of expressing or even demonstrating an awareness of the fact that the pain they are experiencing is a result of a combination of factors. For example, pain manifested physically can be caused by a combination of a child not visiting, a despondent feeling that "God has left me," and a bedsore developed during hospitalization. These examples demonstrate the experience of pain as a total experience. Effective pain relief follows the acknowledgment and management of the physical, psychological, social, and spiritual dimensions (Mehta and Chan 2008).

Consciousness and the Brain

For a few decades pain researchers have been increasingly interested in the structure and function of the brain, especially in chronic pain. The cerebral cortex appears to play an active role in chronic pain. Working models have been developed outlining the mechanism by which acute pain transforms into a chronic state and by which distinct chronic pain conditions impact on the cortex in unique patterns. Such models incorporate knowledge of underlying brain structures and their reorganization, while also including specific variations as a function of pain persistence and injury type, thereby providing mechanistic descriptions of several unique chronic pain conditions (Apkarian et al. 2009). This type of brain research, especially brain imaging, has led to two problems which are relevant for the philosophy of medicine.

The first is the question whether the subjective pain experience can be objectified in a way by brain imaging. Today, some scientists believe that cortical imaging can provide an objective measure of the pain experience (Basbaum and Bushnell 2009, ix). Lee and Tracey (2010), for example, describe how neuroimaging techniques provide an account of neural activity in the human brain when pain is experienced. They admit that pain and suffering as subjective experiences are private and not directly quantifiable. Only behavioral responses and verbal communication can be observed and measured. However, according to them, the physiological recordings of brain activity during pain via neuroimaging are not merely surrogate measures of pain: they have informed us through which our experiences of pain, suffering, and relief emerge. They argue that, although nociception is most often the cause of pain and undoubtedly required for survival, it is neither necessary nor sufficient for the consciousness of pain. They conclude that pain and suffering are highly complex conscious experiences that are ultimately generated by the brain. In the brains of patients with chronic pain, neuroimaging has revealed subtle but significant structural, functional, and neurochemical abnormalities. Converging evidence suggests that the chronic pain state may arise from a dysfunction of the frontal-limbic system.

This conclusion gives rise to a second problem, namely, whether the cause of chronic pain can be discovered by neuroimaging and whether chronic pain can be considered a brain disease. Nowadays, it is quite a common understanding that chronic pain is not a symptom of disease, but rather a disease entity itself, namely, a disease of the nervous system function (Basbaum and Bushnell 2009, ix). This, however, is not an undisputed view. It is argued, for example, that disease is a clinical concept and that conceiving of chronic pain as a brain disease can have negative consequences for research and clinical care of patients with chronic pain (Sullivan et al. 2013). It cannot be simply assumed that the changes associated with chronic pain on neuroimaging are causal. Considered scientifically, one may be looking for the cause of chronic pain through neuroimaging, but considered clinically, one is in fact often looking to validate pain complaints. It is argued therefore that we should resist the temptation to validate pain with the magnetic resonance imaging scanner. Pain cannot be seen as caused by the brain alone. Pain is not felt by the brain, but by the person (Sullivan et al. 2013).

Pain Experience and Expression

The idea of pain as a subjective experience has given rise to further philosophical investigations which problematize standard models, such as Loeser's distinction between nociception, pain perception, suffering, and pain behavior. How do we know for sure, for example, that someone else is in pain? In tackling this question, Nelkin (1986) takes as his starting point the cases of lobotomized patients and patients who are given morphine after the onset of pain, as discussed earlier by Daniel Dennett. On the one hand, we believe that being in pain is being in what can be called a "transparent mental state," such that the person having the pain is in the best position to judge whether he or she really is in pain. On the other hand, we believe that one cannot be in pain without hurting and that hurting is tied up with certain kinds of affects, beliefs, and behavior, such as trying to do something to alleviate the pain. In fact, if grimacing, groaning, and similar behavior occur and if we have no reason for suspecting pretense, then we believe we are in a position to be certain that someone else is in pain.

It is these various sorts of intuition that come into conflict in the cases of the lobotomized and morphined patients. Both types of patients claim to feel pain but also state that it no longer hurts them. One way to solve this problem is to say that these patients, despite what they say, do not really have pain sensations. The changes in their brains brought about by lobotomy or morphine have caused them to be mistaken when they say they do have pain sensations. Another way is to say that such patients had been caused to forget how to use the concept of pain sensations or have even forgotten what the word "pain" means. Nelkin, however, argues that in these cases it would be better to sacrifice our intuition that one is in pain if one has a pain sensation: one can have pain sensations without being in pain and one can be in pain without having pain sensations. From a philosophical perspective, it is therefore crucial to be clear on what we mean when we say "being in pain" and what we mean by "pain sensation."

Pain is an unpleasant sensory and emotional experience. From a scientific point of view, one can analytically distinguish between nociception, perception, suffering, and behavior, but from a phenomenological perspective, for the person in pain, there is just the painful experience. The patient's experience of pain is lived as a whole, difficult to split up in several dimensions and difficult to express (Kleinman et al. 1992, pp. 7–8). Perhaps more than other somatic experiences, the experience of pain resists verbalization. Speaking about pain is one of the most difficult linguistic activities, as pointed out by Ludwig Wittgenstein. The language of pain, he argues, is something very different from the language of customary descriptions (Ehlich 1985).

Physical pain has no voice, but when it at last does find a voice, it begins to tell a story, by the person suffering, by his relatives and friends, and by his physician. In those pain stories, metaphors play a crucial role: the phrase "as if" is crucial. Patients say that it hurts "as if" needles have been stuck in the body, "as if" a hammer knocks on the head, etc. It is difficult or even impossible to describe the pain experience directly without the use of metaphorical or symbolic language. According to Scarry

(1985), pain defies verbal objectification: "Physical pain does not simply resist language, but actively destroys it, bringing about an immediate reversion to a state anterior to language, to the sounds and cries a human being makes before language is learned" (Scarry 1985, p. 4). It is not only the intensity but foremost the quality of pain which is difficult to express. Much cited in the literature on pain is the following quote from Virginia Woolf's essay *On Being Ill*: "English, which can express the thoughts of Hamlet and the tragedy of Lear, has no words for the shiver and the headache. [. . .] The merest schoolgirl, when she falls in love, has Shakespeare and Keats to speak for her; but let a sufferer try to describe a pain in his head to a doctor and language at once runs dry" (quotation by Melzack 1973, p. 45).

Pain and Narrative

Although it can be said that (severe) pain has no voice, almost all pain patients have personal stories to tell about their pain. While listening to patients' stories is as old as medicine, for a few decades much attention has been paid to so-called narrative medicine, in line with a much wider interest in narrativity in philosophy, theology, and the humanities. Basic to this new approach is the idea that human beings are essentially characterized by the fact that they tell and live through stories. A human being is a "storytelling animal," as Alasdair MacIntyre has put it, or a "self-interpreting animal," in the words of Charles Taylor. Narrative medicine is medicine practiced on the basis of "narrative knowledge" and with "narrative competence" (Charon 2006). Nowadays, a growing list of literature exists about the relevance of narrative competence for the treatment of pain patients (Morris 2002, 2012). Medical practitioners must learn to deal with the two most significant sources of predictable uncertainties basic to every narrative encounter and basic to every valid claim of narrative knowledge: interactivity and intersubjectivity. The recent medical literature on pain contains specific studies of patients' stories. Narrativity, therefore, makes a solid contribution to pain, both in research and in treatment. The value of narrativity to pain medicine can be traced in five specific areas: communication, diagnosis, treatment, ethics, and education (Morris 2012).

The Measurement of Pain

Pain as a subjective experience cannot be measured in a strict sense, that is, the way the heartbeat, blood pressure, or glucose levels can be measured. Noxious stimuli and nociceptive responses can be quantified, but not pain. Pain is usually accompanied by all kinds of quantifiable physiological parameters, such as an increase in stress hormones, blood pressure, and heartbeat, but the pain as such is not measurable. Even those closest to a patient cannot truly observe his pain or share in his suffering. Yet an assessment of the degree of pain of individual patients is a daily concern for the practicing physician. Pain, particularly chronic pain, thus challenges one of the central tenets of biomedical epistemology, namely, that there is objective

knowledge, knowable apart from subjective experience. Three strands of activity can be identified in the history of pain measurement (Noble et al. 2005).

The first, psychophysics, dates back to the nineteenth century and measures the effect of analgesia by quantifying the noxious stimulation required to elicit pain, as well as the maximum stimulation tolerated. Methods to measure the so-called single dimension of pain in the laboratory and in the clinic are used to assess the pain threshold as a response to a single painful stimulus. Gross changes in the pain threshold can be assessed in a quantitative way by administering a standard stimulus such as pricking with a safety pin and manual palpation and by asking whether the evoked sensation is painful. The pain threshold marks the transition from the absence of pain sensation to the presence of pain sensation and is quantified as the amount of stimulus needed to evoke a painful sensation (Gracely and Eliav 2009).

In daily life and medical practice, however, pain cannot be seen as a single dimension of sensory intensity. Pain refers to a category of complex experiences, not to a specific sensation that varies only along a single intensity dimension. The second strand of activity uses standardized questionnaires for patients, developed to categorize pain according to its emotional impact, distribution, character, and other dimensions. A number of approaches have extended the evaluation of pain to include all of the qualitative and affective dimensions such as emotions, cognitions, and autonomic and behavioral responses. The most popular and widely used multidimensional pain instrument is the McGill Pain Questionnaire (MPQ). The MPQ presents 20 categories of verbal descriptors with two to six descriptors per category for a total of 78 pain-related descriptors.

The third strand of activity asks patients to report on pain intensity using rating scales and is used in clinical trials where analgesics are evaluated and results can be combined to influence clinical guidelines and protocols. Although all three strands have found a place in modern clinical practice, it is the reporting of pain by patients undergoing treatment using simple scales of intensity which has emerged as the crucial method by which analgesic therapies can be evaluated and compared. Two of the most commonly used techniques to measure evoked and spontaneous pain are the Numeric Rating Scale (scale from 1 to 10) and the Visual Analog Scale (VAS pain). The VAS pain is a single-item scale for pain intensity. The scale is most commonly anchored by “no pain” (score of 0 in a scale of 100) and “pain as bad as it could be” or “worst imaginable pain” (score of 100 on a scale of 100).

The Meaning of Pain

A quite common presupposition in medical theory and practice is that pain has a biological cause. Pain is also explained with all kinds of psychological determinants, such as specific cognition patterns, emotions, stress, or anxiety (Price et al. 2009). However, there are many forms of pain that cannot be explained in biological or psychological terms. And even if one can trace a biological cause or a psychological determinant of pain, this does not mean that one understands pain. Scientific explanations fall short of understanding the phenomenon of pain. It is often argued

that pain is a phenomenon that transcends the borders of science. Pain is a mystery and cannot be explained as having just a signaling function. It also has an ontological and existential dimension (Bakan 1976).

Puzzle or Mystery?

Most of the scientific literature describes pain as a problem, a puzzle or riddle, something to be solved or unraveled. However, the conviction that pain can be explained entirely by the biomedical and social sciences is not self-evident. The increasing interest in pain and suffering, which started in the 1960s, can be attributed both to the increased attention being given to the experience of the sick person and to the fact that they defy explanation on a purely biomedical basis (Cassell 1995). Many authors have argued that the phenomenon of pain cannot be understood by means of the sciences and that we would do better to speak of the "mystery" of pain (Buytendijk 1962). Morris writes: "A true mystery, as opposed to a puzzle or riddle, cannot be known *apart from* the veil that separates us from a true understanding. [. . .] A mystery, then, is not something that exists principally to be solved" (Morris 1993, p. 24). Morris goes on by saying that while the doctor typically approaches pain as a puzzle or a challenge, the patient typically experiences it as a mystery. It seems likely that mystery can never be entirely eliminated from pain as long as pain remains a subjective experience. The term "mystery" here refers to the existential interpretation as provided, for example, by the French philosopher Gabriel Marcel. Pain is not considered to be a solvable scientific puzzle. It is rather a phenomenon that will never betray its secrets, but toward which human beings nevertheless must take a philosophical stance.

Beyond the Signaling Function

Pain is often considered to be a functional warning sign. By producing a retraction from the painful stimulus, the body tries to avoid further harmful situations and damages. There are rare cases of children who are born without the ability to feel pain. This pathological condition of a congenital insensitivity to pain clearly demonstrates the biological function of pain. Many of these children sustain extensive burns, bruises, and lacerations during childhood and frequently bite deeply on their tongue while chewing food, and only with difficulty are they able to learn to avoid inflicting severe wounds on themselves (Melzack 1973). We constantly employ the sensation of pain, even at very mild levels, to adjust our posture or shift our position. We learn how to feel pain and to learn what it means. Children unable to feel pain lack this adaptation mechanism and easily suffer from bodily damages and infections.

An explanation of pain in terms of a signaling function might be adequate in cases of trauma, injury, and infections such as an appendicitis or an inflammation of the ear. In general, however, this theory does not suffice. One might, for example, think of phantom limb pain, which is clearly at odds with this signaling theory.

Another example is the severe pain that accompanies some forms of cancer. In this case, the diagnosis of a malignant process which might be incurable has been well established. The patient is well aware of the diagnosis; he knows that he is incurably ill and is going to die soon. If one considers the function of pain as signaling (possible) threats only, then cases like this one cannot be understood. The experience of pain, especially chronic pain, includes much more than a physical sensation, a signaling function, or a psychological explanation: it creates problems of control and meaning.

Morris argues that we must proliferate the meanings of pain in order not to reduce human suffering to the dimensions of a mere physical problem, for which there is always a medical solution. Pain is not just a biological fact, but "an experience in search of an interpretation" (Morris 1993, p. 38). We experience pain only and entirely as we interpret it. Morris speaks about the "hermeneutics of pain" (Morris 1993, p. 33): "We experience our pain as it is interpreted, enfolded within formal or informal systems of thought that endow it with a time-bound meaning – whether theological, economic, scientific, or psychological. We make sense of pain in much the same way that we make sense of the world. Sometimes pain can even reveal to us beliefs and values we did not know we held" (Morris 1993, p. 45). In other words, pain can be considered to be a heuristic instrument.

Ontological Dimension

Buytendijk's classic study *Pain*. Its modes and function (1962) is a good example of giving pain a personal meaning based on scientific explanations, philosophical insights, and religious convictions. According to Buytendijk, pain can serve multiple purposes and hold multiple meanings beyond its basic function as a signal of tissue damage. He speaks of the ontological dimension of pain, referring to the way in which pain is one of the constituting factors of human existence. According to Buytendijk, the essence of pain is "a state where man is afflicted in his most intimate unity, his psychophysical nature: self is brought into conflict with the body while remaining bound to the body in its painfulness" (Buytendijk 1962, p. 148). Pain leads to a dissociation between the ego and the body. In daily life, when we are involved in all kinds of practical tasks and are close to the things in the world, we often tend to forget our body, as if we do not have one. In (severe) pain, for example, when my hand hurts, this body part takes over control and dominates the whole situation: "We are not tormented by some foreign agent, it is not an incident, a word, a thought, or even sickness or death, however we may acknowledge the power of these: it is our own body" (Buytendijk 1962, p. 26).

Pain does not have an intentional object. Although the capacity to experience pain is as primal a fact about the human being as is the capacity to hear, to touch, to desire, to fear, etc., it differs from these events by not having an object in the external world. While hearing, touch, desire, and fear refer to something in the outside world, pain "is itself alone" (Scarry 1985, p. 162). This objectlessness, the complete absence of referential content, almost prevents pain from being rendered in language. But it also impacts severely on our existence. According to Scarry (1985), suffering pain might be the most evident character of what we call certainty. Suffering pain leads to the

highest possible certainty that we are: *doleo, ergo sum*. The experience of pain is a breaking point in the obvious nature of our normal, healthy existence. Pain awakens us, not only literally but also metaphorically speaking, that is, out of our "metaphysical heedlessness" as Max Scheler has called it (quotation in Buytendijk 1962, p. 22).

Low Back Pain

The existential dimension of pain appears both at an individual and at a general level and can be illustrated by the phenomenon of chronic low back pain. Since time immemorial low back pain belongs to the "evergreen" of bodily complaints. It is already mentioned in the papyrus Edwin Smith (about 1500 before Chr.), in the Hippocratic writings, and in the works of Galen (Allan and Waddell 1989). Even today, low back pain is one of the most common reasons for consulting a primary care physician in industrialized countries. Although low back pain is one of the most common types of pain, it is often said that it is poorly understood (Melzack and Wall 1988). Low back pain is related to physical abnormalities such as arthrosis, spondylotic deformations, and intervertebral disk problems, but very often no biological cause can be found. There are people with anatomical degenerations of the spine who do not complain about their back. And, the other way around, people with low back pain often show no physical abnormality. At an individual level, patients with low back pain have to search for a meaning of their pain and to give it a place in the context of their life story (Dekkers 1998).

At a general level, low back pain is related to one of the characteristics of the human condition, that is, the "uprightness" of the human being. Since time immemorial an upright posture was considered to be one of the essential characteristics of man. Biologically oriented philosophers such as Helmuth Plessner, Adolf Portmann, and Buytendijk specifically described the human posture as upright. In their view, human posture cannot be considered solely from an anatomical and biological perspective. According to Buytendijk, man's upright posture is to be understood as a specifically human posture in a specifically human world. He writes: "It is nature and culture, expression of emancipation and independence. It is also a sign of being threatened: the righteous man is threatened by a collapse" (Buytendijk 1974, p. 230). When standing up, balancing on the small plane of his feet, the human being adopts a distinct positional relation *vis-à-vis* the world. This new relation is only possible through a loss of security, while at the same time offering a new freedom. From such a perspective, the existence of low back pain could be interpreted as a negative side effect of the human upright posture.

Historical and Cultural Aspects

Pain not only has a biological, psychological, and existential dimension but also a cultural dimension. The experience of pain is powerfully mediated by the cultural and historical context (Moscoso 2012). Studying pain is a way of studying the

intersections between the physical human body, the product of evolutionary processes, and the cultural body, that is, the human body as it is experienced and perceived by people in specific cultural and historical circumstances (Porter 1999; Van Dijkhuizen and Enenkel 2009). Cultural beliefs, social values, and religious traditions play an important role in the response to pain by affecting the way we interpret and attend to pain.

Algophobia and Medicalization

In some non-Western cultures, pain and suffering are regarded as facts of life that must be accepted. In Western cultures, many people tend to regard pain as a pathological condition that should be eliminated and prevented, if possible. There is some evidence that Western attitudes toward pain have changed over time. People have become less accepting of pain as medicine has provided them with more effective ways of controlling it (Resnik et al. 2001). Never before in human history has the explanation of pain fallen so completely to medicine. This has led to a seemingly distinctive paradox: although biomedical research has enormously expanded our knowledge of the anatomy, physiology, and pharmacology of pain, never before has pain, particularly chronic pain, reached its present proportions. We possess more knowledge and better remedies than ever before (Morris 1993). Pain has become a medical problem asking for a medical solution. The individual and social willingness to tolerate and accept pain has decreased.

According to Buytendijk, modern man takes offense at many things that used to be accepted by older generations with resignation. Modern man can be irritated by growing old, a long sickbed, and certainly by pain. Its occurrence is unacceptable. In modern society, the demand to do away with pain has become progressively stronger. This has led to the development of an "algophobia [. . .] which is itself an evil and sets a seal of timidity on the whole of human life" (Buytendijk 1962, p. 16). Morris (1993, p. 60) speaks about "medicalization of pain." By what Ivan Illich has called "cultural iatrogenesis," modern medicine has deprived us from our ability to suffer from pain. Pain has become a medical problem which doctors need to solve. Nowadays, adequate pain management is understood to be a fundamental human right and an integral part of the patient-centered part of modern medicine (Cousins et al. 2004).

Change in Pain Sensation or Attitude?

The question raised by the concepts of "algophobia" and "medicalization" of pain is whether modern algophobia can be attributed purely and simply to a decreased willingness to endure pain, that is, a changed mental attitude. Or are we conceivably observing a change in the tolerance of pain, an increase in the pain experience as such, and an increase in the painfulness of pain?

Leriche (1937) gives an affirmative answer to the question whether sensitivity to pain, conceived of as a physiological quality, may have increased in the course of time. His argument is based on his experience as surgeon serving at the front during the First World War and a historical analysis of well-known cases of people who have been suffering from very painful medical conditions. Leriche believed that the increase of sensitivity to pain must have come about in the nineteenth century, where it was strongly promoted by the introduction of surgical anesthesia and of aspirin. Also, the Dutch physician and philosopher Van den Berg (1963) argues on historical-phenomenological (metabletical) grounds that pain sensitivity has increased in the first half of the nineteenth century, more precisely between 1780 and 1845. Since then patients had much more need of analgesics than their fellow sufferers in previous centuries. A key point in Van den Berg's metabletical approach is that pain is linked to social isolation and loneliness: experiencing pain means a lack of human relationships. This view is open to criticism and can naturally not be taken literally. However, what Van den Berg wishes to point out is that the phenomenon of pain should always be studied in its historical and cultural context.

There is some evidence that the pain we feel today differs from the pain our ancestors felt. Most authors, however, come to the conclusion that it is a change of mentality rather than a physiological increase in sensitivity that has caused our changed attitude toward pain. In a historical-phenomenological study of bodily pain, the medical historian Daniel de Moulin argues that on historical grounds one cannot come to the conclusion that medieval man bore his pain in any way distinct from that of patients today. In view of the reaction of the patient, the attitude of the physician, the attention that medieval textbooks have given to the management of pain, as well as the interest at the time in the problem of surgical anesthesia, it is not plausible that pain perception was less acutely felt than today. Pain appears to have been experienced in the same way and with the same intensity as today. De Moulin comes to the conclusion that, finding no arguments that the sensitivity of the nervous system has increased in recent times, we have to attribute the rapidly dwindling readiness to accept pain to a change in mental attitude. The threshold of pain may vary from culture to culture, and even from time to time in the same person: "Pain is [. . .] a subjective way of being an experience as well as an evaluation of the actual situation, and as such the creation of a human being in his sense-giving existence" (De Moulin 1974, p. 570).

The question of how to understand our modern attitude toward pain and how to explain possible differences with the past sounds like an empirical question. In fact, however, this question is intermingled with philosophical problems to such an extent that it is hard to solve the question in an empirical way. From a philosophical perspective it is, for example, crucial to know what is meant in discussions like these by the terms which are central in Loeser's model: nociception, perception, suffering, and behavior. And what about the discussion previously mentioned about the relationship between being in pain and pain sensations? Nevertheless, putting these philosophical questions aside, the conclusion can be drawn that an explanation of our modern attitude toward pain in terms of a change in neurophysiological nociception is difficult to defend.

Definition of Key Terms

Pain: "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (IASP 2014).

Summary Points

- Two opposing views on the medical-biological explanation of pain exist: the *specificity theory* and the *pattern theory.*
- The emergence of modern palliative care, particularly the concept of "total pain," was an important impetus for a holistic approach to pain.
- It is debatable whether chronic pain is a brain disease and whether the subjective experience of pain can be objectified by neuroimaging.
- It is not only the intensity but foremost the quality of pain that is difficult to express in words. In pain stories, metaphors play a crucial role.
- Pain, particularly chronic pain, challenges one of the central tenets of biomedical epistemology, namely, that there is objective knowledge, knowable apart from subjective experience.
- Pain is a mystery; it cannot be explained as having just a signaling function. It also has an ontological and an existential dimension.
- An explanation of our modern attitude toward pain (algophobia, medicalization) in terms of a change in neurophysiological nociception is difficult to defend.

References

Allan DB, Waddell G (1989) Understanding and management of low back pain. The history of low back disability. Low back pain and disability since World War II. Acta Orthop Scand 60(Suppl 234):1–23

Apkarian AV et al (2009) Towards a theory of chronic pain. Prog Neurobiol 87:81–97

Aristotle (1976) The Nicomachean ethics (trans: Thompson JAK). Penguin Books, London

Bakan D (1976) Pain – the existential symptom. In: Spicker SF, Tristam Engelhardt H Jr (eds) Philosophical dimensions of the neuro-medical sciences. Reidel Publishing, Dordrecht/Boston, pp 197–207

Basbaum AI, Bushnell MC (eds) (2009) Science of pain, 1st edn. Elsevier, Amsterdam

Baszanger I (1998) Inventing pain medicine: from the laboratory to the clinic. Rutgers University, New Brunswick/London

Buytendijk FJJ (1962) Pain. Its modes and functions. The University of Chicago Press, Chicago

Buytendijk FJJ (1974) Prolegomena to an anthropological physiology. Duquesne University Press, Pittsburgh

Cassell E (1991) The nature of suffering and the goals of medicine. Oxford University Press, New York/Oxford

Cassell E (1995) Pain and suffering. In: Reich WT (ed) Encyclopedia of bioethics, vol 4. Simon & Schuster MacMillan, New York, pp 1897–1905

Cervero F (2009) Pain theories. In: Basbaum AI, Bushnell MC (eds) Science of pain, 1st edn. Elsevier, Amsterdam, pp 5–10

Chapman CR, Gavrin J (1999) Suffering: the contributions of persistent pain. Lancet 353:2233–2237
Charon R (2006) Narrative medicine: honoring the stories of illness. Oxford University Press, Oxford/New York
Clark D (1999) 'Total pain', disciplinary power and the body in the work of Cicely Saunders, 1958–1967. Soc Sci Med 49:727–736
Cousins MJ et al (2004) Pain relief: a universal human right. Pain 112:1–4
de Moulin D (1974) A historical-phenomenological study of bodily pain in western man. Bull Hist Med 48:540–570
Dekkers W (1998) Hermeneutics and experiences of the body. The case of low back pain. Theor Med Bioeth 19:277–293
Descartes R (1963) Le traité de l'homme. In: Descartes R (ed) Oeuvres philosophiques. Tome I. Garnier, Paris, pp 379–480
Descartes R (1967) Méditations touchant la première philosophie. In: Descartes R (ed) Oeuvres philosophiques. Tome II. Garnier, Paris, pp 404–505
Ehlich K (1985) The language of pain. Theor Med 6:177–187
Engel GL (1977) The need for a new medical model: a challenge for biomedicine. Science 196:129–136
Gracely RH, Eliav E (2009) Psychophysics of pain. In: Basbaum AI, Bushnell MC (eds) Science of pain, 1st edn. Elsevier, Amsterdam, pp 927–959
IASP (2014) http://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698. Accessed 4 Oct 2014
Kleinman A et al (1992) Pain as human experience: an introduction. In: DelVecchio Good M-J et al (eds) Pain as human experience: an anthropological perspective. University of California Press, Berkeley, pp 1–28
Lee MC, Tracey I (2010) Unravelling the mystery of pain, suffering, and relief with brain imaging. Curr Pain Headache Rep 14:124–131
Leriche R (1937) La chirurgie de la douleur. Masson, Paris
Loeser JD (1991) What is chronic pain? Theor Med 12:213–225
Loeser JD (2000) Pain and suffering. Clin J Pain 16:S2–S6
Loeser JD, Melzack R (1999) Pain: an overview. Lancet 353:1607–1609
Mehta A, Chan LS (2008) Understanding of the concept of "total pain". A prerequisite for pain control. J Hosp Palliat Nurs 10:26–32
Melzack R (1973) The puzzle of pain. Penguin, Harmondsworth
Melzack R (1996) Gate control theory. On the evolution of pain concepts. Pain Forum 5(1):128–138
Melzack R, Wall P (1988) The challenge of pain, 2nd edn. Penguin, Harmondsworth
Morris DB (1993) The culture of pain. University of California Press, Berkeley/Los Angeles
Morris DB (2002) Narrative, ethics, and pain: thinking with stories. In: Charon R, Montello M (eds) Stories matter: the role of narrative in medical ethics. Routledge, New York, pp 196–218
Morris DB (2012) Narrative and pain: towards an integrated model. In: Moore RJ (ed) Handbook of pain and palliative care: biobehavioral approaches for the life course. Springer, New York, pp 733–751
Moscoso J (2012) Pain. A cultural history. Palgrave Macmillan, Basingstoke
Nelkin N (1986) Pains and pain sensations. J Philos 83:129–148
Noble B et al (2005) The measurement of pain, 1945–2000. J Pain Symptom Manage 29:14–21
Perl ER (2011) Pain mechanisms: a commentary on concepts and issues. Prog Neurobiol 94:20–38
Porter R (1999) Western medicine and pain: historical perspectives. In: Hinnells JR, Porter R (eds) Religion, health and suffering. Kegan Paul International, London, pp 364–381
Price DD, Hirsh A, Robinson ME (2009) Psychological modulation of pain. In: Basbaum AI, Bushnell MC (eds) Science of pain, 1st edn. Elsevier, Amsterdam., pp 975–1002
Pullman D (2002) Human dignity and the ethics and aesthetics of pain and suffering. Theor Med 23:75–94

Resnik DB, Rehm M, Minard RB (2001) The undertreatment of pain: scientific, clinical, cultural, and philosophical factors. Med Health Care Philos 4:277–288
Saunders C (1967) The management of terminal illness. Hospital Medicine Publications, London
Scarry E (1985) The body in pain. The making and unmaking of the world. Oxford University Press, New York/Oxford
Seymour J et al (2005) Pain and palliative care: the emergence of new specialties. J Pain Symptom Manag 29:2–13
Sullivan MD et al (2013) What does it mean to call chronic pain a brain disease? J Pain 14:317–322
van den Berg JH (1963) Leven in meervoud. Een metabletisch onderzoek. Callenbach, Nijkerk
van Dijkhuizen JF, Enenkel K (2009) Introduction: constructions of physical pain in early modern culture. In: van Dijkhuizen JF, Enenkel KAE (eds) The sense of suffering. Construction of physical pain in early modern culture. Brill, Leiden, pp 1–17
WHO. http://www.who.int/cancer/palliative/definition/en. Accessed 4 Oct 2014

Death as Biological Category

14

Stephen Holland

Contents

Abstract

It is very important to know whether someone is alive or dead. Usually this is obvious, but there are difficult cases such as brain-dead patients on life support and brain-injured patients who are in a permanent vegetative state. The traditional way of determining death centered on the cardiorespiratory system: a patient was declared dead when breathing and heartbeat had stopped. Advances in medical technology, such as artificial ventilation and resuscitation techniques, brought this connection between death and the cardiorespiratory system into question. In response, "brain death" was proposed. Brain death is a criterion that is part of a biological paradigm for death. There are objections to both brain death and the biological paradigm. This has given rise to an alternative consciousness-based paradigm for death. There is objection to this too, which creates a quandary as to

S. Holland (✉)
Departments of Philosophy and Health Sciences, University of York, Heslington, York, UK
e-mail: stephen.holland@york.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_11

what death is and whether it has occurred in difficult cases. Arguably, this quandary was inevitable given the way the ordinary concept of death works. Some responses to this quandary are canvassed, but they are problematic, so the death debate is as yet unresolved.

Introduction

The subject area of this article is death, in particular how to go about deciding whether someone has died. This matters because how someone ought to be treated in large part depends on whether they are alive or dead: imagine, for example, cremating someone wrongly thought of as dead. In particular, it matters because, according to what is known as the "dead donor rule," vital organs should only be retrieved for transplant purposes from patients known to be dead.

The historical background to the subject is that advances in medical technology made it more difficult to decide whether a patient has died. Of course, it remains obvious in the majority of cases but two factors in particular made it more difficult. First, whether someone is alive or dead used to be determined by referring to their cardiorespiratory system, i.e., checking for a heartbeat or pulse, and whether they were still breathing. But lots of medical developments, such as pacemakers, resuscitation techniques, heart transplants, etc., severed the connection between the state of someone's cardiorespiratory system and whether they are alive or dead.

The second factor is that patients can now be maintained in various conditions that are ambiguous regarding whether they are alive or dead. Here arise some important exemplars of the problem discussed in this article. Consider, for example, a patient on a life support machine who has suffered "whole brain death" – i.e., their brain is irreversibly incapable of functioning – but their body is artificially maintained so they are capable of digesting and excreting food, responding to temperature changes, etc. Are they alive or dead? Another example is the brain-injured patient in a permanent vegetative state. Like the whole brain dead, the permanently vegetative retain lots of physiological functions (such as digesting food), but crucially, they are also capable of breathing unaided. However, the parts of their brain responsible for consciousness are destroyed, so they have lost any prospect of "waking up" and regaining conscious awareness. Again, are they dead or alive?

The structure of this article is determined by the shape of the debate. First, the development of the proposal known as "brain death" is traced. Brain death was a new criterion for death, so it belongs in what is known as a "definition-criteria-tests" model which comprises a biological paradigm for death. Problems with the biological paradigm are presented. The major alternative is the consciousness-based paradigm (or the "higher-brain" account). Problems with the consciousness-based paradigm are presented. Clearly, this creates a dilemma since neither of the main ways of deciding what death is and whether it has occurred is satisfactory. A reason for this is suggested, based on the ambiguous nature of the ordinary concept of death. Some responses are canvassed but they too are problematic so the controversy about death continues.

Brain Death

Treating a living patient on the basis that they are dead would be bad in various ways, from failing to provide potentially beneficial treatment to causing death by removing their vital organs. Conversely, it would be bad to treat a dead patient as if they are alive, not least because scarce medical resources that could have helped others would be wasted on them. So, it is important to know whether a patient has died or not.

It used to be very straightforward to determine whether death had occurred. In the vast majority of cases it was obvious: the patient in conversation with their doctor is alive; the corpse stored in the morgue is dead. And if any ambiguity arose, there were simple tests to resolve it. These focused on the patient's heart or, more generally, their cardiorespiratory system. For example, medical staff would feel for a pulse or listen for a heartbeat using a stethoscope; if heartbeat and breathing had stopped, the patient was declared dead.

In the postwar period this situation was complicated by developments in medical technology which challenged reliance on the cardiorespiratory system when deciding whether someone had died. For example, improved resuscitation techniques meant that a patient who would have been declared dead could be revived; given that death is irreversible, such patients were considered to have been alive throughout their ordeal as opposed to having died and been brought back to life. Innovations such as pacemakers and heart transplant surgery meant that a patient could survive irreparable damage to their heart. Most notably, developments in intensive care meant that, for patients on life-support machines, cardiorespiratory functions are performed – i.e., oxygenated blood is pumped around the body – albeit non-spontaneously and only with mechanical assistance.

The role of the cardiorespiratory system in deciding whether death has occurred was put under further pressure by concerns about transplant organs. In particular, patients on life support have healthy organs which could be donated to other patients. But the "dead donor rule" states that it is impermissible to kill anyone for transplant purposes. So, whether it is ethically permissible to retrieve organs from patients on life support turns on whether they are alive or dead. According to the traditional cardiorespiratory means of determining death, they are alive because, although incapable of pumping blood around the body and breathing spontaneously, their cardiorespiratory functions are performed with mechanical support. In turn, it would be impermissible to retrieve their organs for transplant. But transplant organs are very scarce medical resources so this situation seemed regrettable.

In the response to this pressure on the traditional focus on the cardiorespiratory system when determining whether a patient has died, two landmarks are notable. First, in 1968 an ad hoc committee of the Harvard Medical School undertook to examine the issue (Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death 1968). The committee proposed a new criterion of death, namely, irreversible coma: i.e., a patient with a permanently nonfunctioning brain is dead. But untenable confusion was created because only some jurisdictions adopted the committee's proposal so "someone could be dead in Kansas but alive,

for example, in the neighboring state of Missouri" (Lizza 2006, p. 9). In order to clear up the confusion, a President's Commission was set up to achieve uniformity. The Commission's report resulted in the Uniform Determination of Death Act (UDDA) which states: "an individual who has sustained either (1) irreversible cessation of circulatory or respiratory functions, or (2) irreversible cessation of all functions of the entire brain, including brain stem, is dead" (President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research 1981).

It is important to note that each of (1) and (2) in the UDDA is sufficient for death to be declared. In other words, if (1) occurs, then the patient is dead: i.e., if there is irreversible cessation of cardiorespiratory functions, then the patient is dead. And if (2) occurs, then the patient is dead: i.e., if the patient's entire brain is irreversibly defunct, then the patient is dead. But (1) is not necessary for death to have occurred, so a patient can be dead even if cardiorespiratory functions are performed. This is known as "brain death," and it is referred to as a neurological criterion for death. Versions of the brain death account have been very widely adopted but remain highly controversial.

The practical significance of these landmark judgments is that some patients who were considered to be alive turn out to be dead. Notably, in the terminology of the ad hoc committee, patients on life support are in an "irreversible coma"; since they meet the ad hoc committee's neurological criterion for death, they are dead. In the terminology of the UDDA, these patients meet condition (2) because their whole brain is dead, so they are dead even though they do not meet condition (1). Declaring patients on life support to be dead has two major practical ramifications. First, hospitals are no longer required to use resources to maintain brain-dead patients. Second, the organs of patients on life support can be retrieved without transgressing the dead donor rule (Truog and Robinson 2003).

That incorporating brain death has these resource implications motivates suspicion – never quite dispelled – that the policy was rather too convenient for those seeking to save money and increase the number of donor organs. Also, this development in diagnosing death occurred with very little public debate or protest (Singer 1995a, pp. 28–32). If such suspicions are well founded, the policy change got matters the wrong way round: ethically speaking, the way patients are treated is supposed to be based on their being alive or dead, as opposed to deciding whether they are alive or dead on the basis of how others want to treat them. But whether brain death was merely a convenient fiction to save resources is best left moot here, depending as it does on the murky business of quite what motivates a shift in medical policy and practice.

The Biological Paradigm for Death

The more telling way of evaluating brain death involves putting it into a wider context. In particular, there are three different but closely connected issues which must be clearly delineated:

First, the definition of death: what is death, what do or should people understand by death, and what does the concept of death mean?
Second, criteria to determine whether death has occurred: what has to happen to a creature for it to die and what change or changes take place in an individual when it goes from being alive to being dead?
Third, diagnostic tests: how can whether death has occurred be checked and what investigative methods are available to decide whether a creature has died?

This is known as the definition-criteria-tests model (Bernat et al. 1981). Here, a completed definition-criteria-tests model will be referred to as a paradigm for death. The relationship between the second and third components – criteria and tests – is simply that diagnostic tests determine whether criteria have been met in particular cases. For example, clinicians can run tests on a patient whose status is unclear in order to determine whether the criteria for death have been met. The relationship between the first two components – definition and criteria – is more involved: the definition of death justifies the criteria. In other words, what grounds criteria for death is their compatibility with the definition in question. So, of these three components of the paradigm, the first – the definition of death – is fundamental in that it provides the foundation to the others (Feldman 1992; Chiong 2005; Boniolo 2007).

In order to evaluate brain death, it has to be put in the context of a definition-criteria-tests model that will be referred to here as the biological paradigm for death. The crucial point is that the ad hoc committee and President's Commission proposed an additional *criterion* for death, i.e., the second of the three components of the definition-criteria-tests model. The literature on brain death is bedeviled by confusion because criteria and definitions get mixed up. For example, it is common to encounter phrases such as "brain death as death" in the literature. But brain death is not death – not a definition of death, an attempt to say what death is – rather it is a *criterion* for death: when the brain dies, the individual can be declared dead. The trouble started early on in the debate. For example, the President's Commission was asked to address "the matter of 'defining' death at the level of general physiological standards rather than at the level of more abstract concepts or the level of more precise criteria and tests." Although definition, criteria, and tests appear in this quote, terminological distinctions could have been clearer.

So, brain death was a change in the criteria for death. But, as already pointed out, the fundamental component of a paradigm is the definition of death because the definition grounds criteria, and tests simply confirm whether criteria have been met. What was the definition of death on which brain death was grounded? A standard version is: death is the irreversible breakdown in the functioning of the organism as a whole. This rather odd-sounding definition needs to be unpacked.

First, death is something that happens to organisms. Of course, it is often said of nonorganic things that they have died, as in "the batteries are dead." But according to the definition under discussion, this is a metaphorical, as opposed to a literal, use of the word "dead." Furthermore, death is something that happens to all organisms: the fate of all organisms is to undergo "irreversible breakdown in their functioning as a

whole." So this is a trans-species definition of death: i.e., death is the same thing for all organisms, from trees to fish, spiders, slugs, dogs, monkeys, and human beings.

Second, organisms function in a distinctive fashion. Of course, individual organs function, such as kidneys and eyes, and more or less discrete systems within an organism function (e.g., the cardiorespiratory system). But the functioning in the biological definition of death under consideration is of a different sort. A living organism functions in a holistic fashion in the sense that all components are integrated and interlocked to form a coherent organic whole. According to the definition, an organism's being alive is a matter of its functioning in this distinctive way; death amounts to its irreversibly ceasing to do so.

To see that this definition of death grounds brain death, recall the ad hoc committee and President's Commission's proposals. The reason they focused on the brain when devising criteria for death is that the brain is unique in governing and controlling the functioning of the organism as a whole. Compare, for example, the role of the brain and the kidney. Both are organs of the human body but the kidney plays a discrete role which can be performed by an organic (transplant organ) or inorganic (dialysis machine) alternative. By contrast, the brain's role is wide ranging in that it governs the operation of other parts of the body; and there is no organic or inorganic equivalent that can play this integrating role.

Analogies are often used to capture this relationship between the brain and the whole functioning organism; for example, the brain is said to relate to the body as the programming box relates to a central heating system. Note that, while on the definition under consideration, death is trans-species – i.e., the same for all organisms – the neurological criterion for death involving the brain is species specific in that it only applies to encephalic creatures (i.e., creatures with a brain); a tree's integrative and holistic functioning, for example, is obviously not governed by its brain.

There is a complication that recalls the President's Commission reference to ". . . functions of the entire brain, including brain stem . . .." The former, death of the entire brain, is referred to as whole brain death; the latter is known as brainstem death. The complication is that in some jurisdictions, notably the UK, brainstem death suffices for death to be declared, whereas elsewhere, notably in America, whole brain death is required. Whether anything hangs on this distinction is controversial. For example, that the brainstem is "the site where all integrative capacities for consciousness and involuntary integrative physiologic functioning reside" (Canadian Council for Donation and Transplantation 2003) suggests that whole brain death and brainstem death amount to the same thing. But the American insistence on death of the whole brain has been reiterated, for example, due to concerns about confirmatory tests and "super locked-in syndrome," i.e., the theoretical possibility that someone could retain conscious awareness despite irreversible loss of brainstem function (President's Council on Bioethics 2008).

This distinction between whole brain and brainstem death is elided here. This is because either of these can be inserted into the definition-criteria-tests model that comprises a biological paradigm for death, and it is more important to examine that paradigm than to wrangle over which of alternative neurological criteria to select.

So, in sum, the biological paradigm for death is (definition) death is an irreversible breakdown in the functioning of the organism as a whole; (criteria) this occurs when the whole brain/brainstem is irreversibly defunct; (tests) diagnostic investigation reveals whether this has occurred. Now that brain death has been positioned in the biological paradigm, the whole approach can be properly evaluated. (For useful summaries of developments, see Singer 1995a, ch. 2; Lamb 1996, ch. 2; Belshaw 2009, pp. 219–226.)

Evaluating the Biological Paradigm

Advocates support the brain death account in various ways (Bernat 1998). Two strategies are canvassed here. The first is to argue that brain death should not be considered controversial. For example, advocates point out that the same definition of death underpins both the traditional cardiorespiratory and the new neurological criteria for death. The definition is "irreversible breakdown in the functioning of the organism as a whole"; it used to be the case that the best criterion for whether this had occurred in an individual was irreversible cessation of cardiorespiratory functions, but this criterion needed augmenting because of the disconnect described above between the state of a patient's cardiorespiratory system and whether they are alive or dead; hence, a new neurological criterion, brain death (and attendant tests), was proposed. Since this merely adds a new criterion for the same phenomenon, brain death is not a major change so should not be thought of as controversial.

Another example of this strategy of defusing controversy is to claim that the two criteria for death – cardiorespiratory and neurological – are intimately connected. In particular, cessation of cardiorespiratory functions entails brain death because a brain starved of oxygenated blood will die. On the other hand, brain death entails the irreversible demise of the cardiorespiratory system because a dead brain cannot control and regulate cardiorespiration or any of the patient's other physiological functions. Again the thought is that, since these two criteria are so intimately connected, the proposal to allow a neurological criterion for death related to the brain is not controversial.

Another strategy for defending brain death is to emphasize putative advantages of the biological paradigm. For example, as already mentioned, the definition of death that underpins the paradigm – irreversible breakdown in the functioning of the organism as a whole – is trans-species because death turns out to be the same phenomenon for all organisms. And the biological account is said to achieve procedural clarity: as stated at the outset, it is ethically important to delineate clearly the living and the dead; the neurological criterion of brain death is said to help in this regard because there are clear tests for confirming the death of a brain and, in turn, the death of the patient.

None of these arguments for adopting the neurological criterion of brain death is fully convincing. To take each strategy in turn, the controversy surrounding brain death cannot be so easily defused. For one thing, as discussed more fully below, the relationship between brain death and organismic breakdown is much more vexed

than advocates suggest. For another, no doubt the cardiorespiratory system and the brain are intimately connected, but this does not establish that they are equally important in determining whether death has occurred.

Regarding the second strategy, putative advantages of the account are not as striking as advocates suggest. That death is trans-species is not self-evidently correct: one might think that the deaths of creatures as unalike as trees and persons are dissimilar (this motivates the consciousness-based paradigm for death discussed below). And whether brain death achieves procedural clarity is disputed. For example, sociological studies of the way brain death is established suggest that, in practice, investigative methods do not always concord (Kellehear 2008). More to the point, that a policy is easy to implement does not establish its correctness: a neurological criterion for death should be adopted because it correctly identifies the dead, not because it is easy to apply.

Not only are standard arguments for a neurological criterion for death unconvincing, there are strong arguments against brain death and the biological paradigm as a whole. To illustrate, consider a very distinctive patient: a pregnant woman who suffers an adverse event that results in whole brain death; she is maintained on life support for some time during which she gestates her fetus which is born by Caesarian section. Is this patient alive or dead, according to the biological paradigm? The patient is brain dead so she meets the neurological criterion in the biological paradigm and, in turn, she is dead. But clearly this patient has not yet suffered "irreversible breakdown in the functioning of the organism as a whole" because the whole organism is capable of integrative functioning sufficient to gestate a fetus. So, according to the definition of death in the biological paradigm, she is still alive. In this case, the criterion and definition of death in the biological paradigm have come apart: the patient is alive according to the definition of death but dead according to the criterion (cf. Halevy and Brody 1993).

The pregnant brain-dead patient is only a particularly striking case of the general point. All brain-dead patients on life support are capable of integrated physiological functioning – albeit only with mechanical support – that is at odds with their meeting the biological definition of death. As opponents of brain death often point out, such patients are capable of digesting food, excreting waste materials, regulating body temperature, and so on (for a list of such functions, see President's Council on Bioethics 2008, p. 56).

But it is worth emphasizing the precise point of the objection here. It is not that, intuitively speaking, artificially supported brain-dead patients are alive because they appear to be alive (e.g., healthcare staff do not think of them as dead). Such intuitions are compatible with the patient being, in fact, dead; after all, brain death was meant to revise the diagnosis of these patients. The precise point is that the biological paradigm is incoherent: artificially maintained brain-dead patients are alive according to the paradigm's definition but dead according to the paradigm's criterion.

Emphasizing this helps deal with the obvious rejoinder by advocates of brain death and the biological paradigm. The rejoinder is that the patient is not alive because machines, not the patients themselves, enable physiological functions to be

performed. The idea is that the patient contains an assortment of physiological "kit" that continues to operate by dint of mechanical assistance; take away the machines, and the kit stops working. So, for example, the pregnant patient is a sort of cyborg equivalent to an organic and inorganic hybrid gestation machine, not a living human being.

But while others might well question whether the brain-dead patient on life support is still alive, advocates of brain death cannot. This is clear when brain death is positioned within the biological paradigm. Brain death is a criterion grounded on a definition of death such patients simply do not meet. The fact that machinery is required to maintain the patient is no more relevant than the fact that soil is required to maintain a lettuce: take either away and the entity will die, but until then they are integrative functioning organisms that do not meet the definition of death fundamental to the biological paradigm.

Furthermore, the motivation for devising a neurological criterion for death disappears. The reason for focusing on the brain when devising criteria for death was that the brain was thought of as the organ that governs and controls the integrative functioning of the organism as a whole. To recall, the brain was supposed to relate to the body the way a programming box relates to the central heating system. But if brain-dead patients on life support are capable of integrative organismic functioning (as illustrated by the brain-dead pregnant woman), this account of the relationship between brain and organism must be wrong; in which case, why look to the brain for criteria for death (Shewmon 1998, 2001; Youngner et al. 1999; Capron 2001)?

The Consciousness-Based Paradigm

Evidently, brain death and the biological paradigm of which it is a part are vexed. One option is to retain the definition of death in the biological paradigm but reject the neurological criterion of brain death. In other words, death is irreversible breakdown in the functioning of the organism as a whole; but it is not the case that death has occurred just because someone's brain has irreversibly ceased to function (Belshaw 2009). But opponents argue that not even the definition of death is correct. Here arises the fundamental objection to the biological paradigm. Opponents argue that the fundamental problem is biological reductionism. In other words, the paradigm reduces life and death to the merely biological. And, intuitively, it seems wrong to think of phenomena of such significance as life and death as merely biological, on a par with other biological features such as digestion or photosynthesis.

Opponents go on to suggest an alternative paradigm, one that avoids biological reductionism. To introduce this, recall the point made in passing above that one might well wonder whether the death of creatures as unalike as trees and people must be the same phenomenon. What drives this intuition is that people are very different from trees because of their rich and complex psychology. Here, the concept of personhood is useful. Persons are individuals such as you and I are now, who have mental lives comprised of conscious awareness of one's environment, self-consciousness, memories and desires, and capacities for language use and

rationality. Note that the category of persons does not map onto similar sounding categories such as people and human beings because not all people or humans are persons. For example, babies have yet to acquire, and the severely demented have lost, the characteristics of personhood; the very severely learning disabled will never acquire, and severely brain-injured patients have tragically lost, personhood.

Personhood is what matters to us: I have a body, but I am a person. So it is natural to think that if I cease to be the person I am then I go out of existence. But what more could my death amount to than just this: my going out of existence? There is an immediate complication. Personhood is a contested notion. For one thing, the characteristics of personhood rather glibly listed above are much disputed (in fact, this was precisely the reason the President's Commission did not pursue a non-biological understanding of death when devising the UDDA). For another, there is much disagreement about the ontology of persons, i.e., the sense in which persons exist. For example, some people think personhood is simply a phase most human beings pass through, like adolescence; others think that persons are in some way ontologically distinct from the bodies to which they (in some way) relate.

So, it seems that the intuition that I die when the person I am goes out of existence, though initially plausible, cannot be developed because there is no consensus on personhood. But advocates of the consciousness-based paradigm have a strong response. Whatever else is true of persons – what characterizes them, their ontological status, etc. – consciousness is a necessary condition for personhood. In other words, there is personhood only if there is consciousness (or: no consciousness, no person). This retains the intuition about one's own death while avoiding all the controversies about personhood: if I cease to be conscious, then I cease to be a person; but a person is what I am; so if I cease to be conscious, then I go out of existence; and it is hard to see what more could be involved in my death than that (Rich 1997; Lizza 1993, 2006).

This is what motivates the consciousness-based paradigm for death. As clarified above, a paradigm for death has three components: definition, criteria, and tests. What are the three components of the consciousness-based paradigm for death? First, the definition of death that underpins the paradigm is: death is an irreversible loss of capacity for consciousness. Second, the criterion for death is that parts of the brain responsible for consciousness are no longer capable of sustaining consciousness. Third, tests establish that a patient's brain is damaged in such a way as meet the criterion.

The consciousness-based paradigm is obviously species specific because it does not apply to nonconscious creatures such as trees. In fact, it is even more specific than that; for example, it does not account for the death of a nonconscious human such as a young fetus. So the consciousness-based paradigm is not intended to replace the biological paradigm, which is required to account for the death of lots of creatures. Nonetheless, for specific individuals – those with a capacity for consciousness capable of sustaining personhood – death is other than that which is presented in the biological paradigm: not organismic breakdown but the irreversible loss of the capacity for consciousness that is the necessary condition for personhood.

The implication of the consciousness-based paradigm is that some people turn out to be dead that would be alive according to the biological paradigm. Two sorts of patients are particularly important. First, anencephalic infants are born with a brain that has not developed in such a way as to sustain consciousness; second, the brains of permanently vegetative patients have been damaged in such a way that they are no longer capable of sustaining consciousness. Both sorts of patients display autonomic physiological functions, such as breathing unaided. According to the biological paradigm, such patients are alive: they do not meet the neurological criterion for death because neither their whole brain nor their brainstem is dead; and they are not dead according to the biological definition of death because they display integrative organismic functioning. But according to the consciousness-based paradigm, they are dead: the person either did not come into or has gone out of existence, which is what death amounts to for such individuals (Veatch 1975; Green and Wikler 1980).

Evaluating the Consciousness-Based Paradigm

There are numerous worries – philosophical, pragmatic, and ethical – about the consciousness-based paradigm for death (DeGrazia 1999a; Fisher 1999). The central philosophical problem recalls the distinction between literal and metaphorical uses of the word "death." According to the consciousness-based account, persons are the kind of thing that can literally, not merely metaphorically, die. How plausible this is depends on the ontology of persons, i.e., the sense in which persons – as opposed to human beings or the organisms to which persons relate – exist. As mentioned above, there is considerable philosophical disagreement about this which cannot be sorted out here, but at least the outline of the dilemma for the consciousness-based account can be drawn.

There is a continuum from philosophically less contentious to philosophically more contentious ontologies of persons. For example, the phase view is that personhood is akin to adolescence in being a phase through which most human beings pass, one characterized by various psychological capacities such as self-consciousness and rationality (DeGrazia 1999b). The phase view is philosophically uncontentious because nothing dies when a human being stops being a person, just as nothing dies when a human being stops being an adolescent. But the consciousness-based view is that a death does occur when a human being ceases to be a person; even when, as in the case of permanent vegetative state, the human being continues to function as an organism (Lizza 2006). So persons must exist in such a way that they literally live and die independently of the organism (which can also die an independent death, which suggests that people such as you and I will die two deaths). This is much more philosophically contentious because it involves some form of dualism of persons and human organisms (Shrader 1986; McMahan 1995; Hershenov 2006).

On a pragmatic note, a major problem with the consciousness-based paradigm centers on the third component of the definition-criteria-tests model, namely, diagnostic tests. Can it be stated with confidence that a brain is irreversibly incapable of sustaining consciousness? The fundamental problem is that the understanding of

neurological bases of consciousness is far from complete. This is evident in current controversies over whether patients diagnosed as permanently vegetative are, in fact, conscious or have some prospect of recovering consciousness (Owen et al. 2006). And unless it can be accurately determined whether a brain is such that the individual's capacity for consciousness is irreversibly lost, there is a danger of misdiagnosing patients as dead.

The main ethical worry about consciousness-based accounts of death is a slippery slope argument. Some patients are conscious but, irreversibly, nonpersons. Examples include severely learning disabled patients who have never had, and patients suffering severe dementia who have lost, the characteristics of personhood. Advocates will point out that, according to the consciousness-based account, patients such as the learning disabled and Alzheimer's sufferer are conscious therefore alive. But the slippery slope worry is that loss of consciousness is said to be significant because it connects to loss of personhood, and the learning disabled and Alzheimer's patient, for example, also lack personhood; so there is a danger of treating them as if they too are dead, notwithstanding the fact that they are conscious.

The Ordinary Concept of Death

The upshot of the discussion so far is that there is a quandary about death because neither the biological nor consciousness-based paradigm is wholly satisfactory. One view is that this quandary is unavoidable. This starts by asking after the best approach to the question as to what death is. For example, one might take a scientific approach; but all that a scientific analysis will give us is more detail about the biology of death, not reveal what death itself is. An alternative is to ask experts other than scientists who work a lot around death, such as doctors and hospice staff; but why take their distinctive engagement with death as authoritative? Or perhaps it is a matter of doing more philosophy, pursuing the metaphysics of death in the way touched on above, for example, in the discussion of ontologies of persons; but metaphysics is notoriously contentious and removed from everyday life.

Such problems with the more obvious ways of addressing the question as to what death is suggest that the best way forward is to attend to the ordinary concept of death, i.e., to reflect on what people have in mind when they think and talk about death. But, arguably, in undertaking this, it turns out that the ordinary concept of death involves both of the ways of understanding death captured in the biological and consciousness-based paradigms, respectively. In other words, in thinking and talking about death, people have in mind the breakdown in functioning of organisms familiar from the biological paradigm. But if this were all that is meant by death then it is akin to other biological phenomena, such as digestion or photosynthesis, which is implausible. In fact, the ordinary concept of death involves other, non-biological thoughts such as: when I am dead it will never again be like anything to be me. This is a consciousness-based thought: death is a matter of never again being conscious, irrespective of what is happening, biologically speaking, to the organism to which I relate. Since, in the ordinary understanding, death is biological but also something

more than merely biological, the quandary reached at the end of the previous section was inevitable (Holland 2010).

This has implications for the ambiguous cases discussed in this article, such as brain-dead patients on life support, anencephalic infants, and the permanently vegetative. Dwelling on their biology invites the thought that they are alive because of their integrative organismic functioning. But in such cases, not only the person, but also the capacity for consciousness required for personhood, are missing; dwell on this and the thought that they are dead – or, at least, not straightforwardly alive – gets hold. So their condition is unavoidably unclear: they are neither straightforwardly alive nor yet simply dead (Wikler 1988). Empirical data from interviews with relatives and experts involved with permanently vegetative patients support this analysis: interviewees speak of such patients as being alive but also struggle to explain the ontological status of patients and even explicitly state that they are already dead (Holland et al. 2014).

Two Responses

One response to this quandary is to question how important it is that the patient is alive or dead in deciding how to treat them. A version of this response reinstates the traditional cardiorespiratory criterion for death. On this criterion, anencephalic babies and the permanently vegetative – and even brain-dead patients on life support, assuming the need for mechanical assistance is irrelevant (Kamm 2001) – are alive. But the life they have is that of a functioning organism and such a life is ethically insignificant. So, for example, it is permissible to retrieve their organs for transplant purposes and cease futile treatment. This does justice to the difficulties involved in establishing that such patients are dead while achieving the goals of saving medical resources (Singer 1995a).

But the obvious objection is that this suggestion flouts the dead donor rule by allowing people to be killed for transplant purposes. Perhaps it is no more than moral squeamishness to retain the rule (Arnold and Youngner 1993; Singer 1995b; Fost 2004; Veatch 2004; Potts and Evans 2005). But worries persist about stipulating that a patient does not have an "ethically significant life" (Rachels 1986; Lamb 1996, pp. 60–66). Who has the moral authority to decide whether a life is ethically significant? And is this the top of a slippery slope to more dubious killing for utilitarian gains?

Another response to the quandary is to say that, since defining death is unavoidably difficult, people should be allowed to decide for themselves whether death has occurred. This is known as pluralism about death because it allows a plurality of views about when death has occurred. People can decide either individually, regarding themselves or their loved ones, or in cultural (sub)groups as to whether to declare someone as dead (Lizza 2006, ch. 8; Bagheri 2007; Molina et al. 2008).

Again, this recognizes the quandary about death, and it avoids stipulation regarding whether death has occurred. Also, despite the attempt to achieve uniformity, diagnosing death has admitted of pluralism throughout the debate in that different

jurisdictions have adopted different criteria for death, and many have allowed cultural groups to retain nonstandard criteria. But there is something unsatisfactory about saying that when and whether death has occurred is a personal or cultural matter as opposed to a matter of discoverable fact.

Concluding Remarks

In conclusion, the death debate remains currently unresolved. There is something right about both of the main paradigms for death – the biological and the consciousness based – but neither is wholly satisfactory. Arguably, this quandary is inevitable because of an ineluctable complexity in the way death is ordinarily thought about. Attempts to come to terms with the quandary, either by saying that some sorts of lives are ethically insignificant or people should be allowed to decide for themselves whether death has occurred, are not wholly satisfying. So the debate goes on.

Definitions of Key Terms

Anencephalic infant	The brain has developed small or entirely missing hemispheres so the infant is capable of physiological functions, including spontaneous breathing, but not consciousness.
Biological paradigm	(i) A definition of death as irreversible breakdown in the functioning of the organism as whole; (ii) two criteria for death of human beings, namely, brain death (see below) and irreversible cessation of circulatory or respiratory functions; and (iii) appropriate tests to check whether criteria have been met in specific cases.
Brain death	Irreversible loss of brain function; according to the Uniform Determination of Death Act, brain death is sufficient to declare death.
Cardiorespiratory criteria for death	Irreversible cessation of circulatory or respiratory functions.
Consciousness-based paradigm	(i) A definition of death as irreversible loss of capacity for consciousness; (ii) criteria for death of human beings, namely, irreversible damage to parts of the brain responsible for sustaining consciousness; and (iii) appropriate tests to check whether criteria have been met in specific cases.
Criteria for death	Conditions met by a creature in virtue of having died.
Dead donor rule	It is ethically impermissible to kill someone for organ transplant purposes.
Definition of death	Tells us what death itself is, what “death” means.

Ethically insignificant life	The individual in question is alive but the sort of life they have is of no value to them.
Ontology of persons	The study of the sort of existence persons have, e.g., whether a person is simply a phase (most) humans pass through (like adolescence) or an entity, distinct from the organism to which they relate, which can literally live and die.
Paradigm for death	Comprised of a definition, criteria, and tests for death (the "definition-criteria-tests model").
Permanent vegetative state	Parts of the brain responsible for consciousness are irreversibly damaged, but the brainstem is sufficiently intact to retain physiological functions including spontaneous breathing.
Personhood	Characterized by (disputed) psychological characteristics such as consciousness, self-consciousness, rationality, and language use.
Pluralism about death	The policy of allowing people, individually or in cultural subgroups, to decide what death is and whether it has occurred in specific cases.

Summary Points

- It is ethically important to know whether patients are dead or alive (e.g., we are only allowed to take vital organs from dead patients).
- Traditionally, any difficulties in deciding this were resolved by testing the cardiorespiratory system (a patient without a pulse/heartbeat, or who had stopped breathing, was declared dead).
- Reliance on cardiorespiratory criteria for death was brought into question by medical advances such as resuscitation techniques and life support machines.
- So a new way of determining death was established, namely, brain death: a patient whose brain had irreversibly ceased to function is dead.
- But when this new criterion for death is positioned in the "biological paradigm for death," lots of problems emerge.
- This gave rise to an alternative, non-biological account, the consciousness-based paradigm: a patient whose brain is incapable of sustaining consciousness is dead even if they continue to function as an organism.
- But there are lots of problems with the consciousness-based paradigm.
- So we are in a quandary about death and about whether certain patients are dead or alive; arguably, this was inevitable given the way the ordinary concept of death works.
- We could say some patients are alive but their life is not ethically significant, but this seems like a stipulation and morally dangerous.
- We could let people decide for themselves whether death has occurred, but this will result in confusion and allow dubious decisions about how to treat patients.

References

Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death (1968) A definition of irreversible coma. JAMA 205:337–340
Arnold R, Youngner S (1993) The dead donor rule: should we stretch it, bend it, or abandon it? Kennedy Inst Ethics J 3:263–278
Bagheri A (2007) Individual choice in the definition of death. J Med Ethics 33:146–149
Belshaw C (2009) Annihilation: the sense and significance of death. Acumen, Stocksfield
Bernat J (1998) A defense of the whole-brain concept of death. Hast Cent Rep 28:14–23
Bernat J, Culver C, Bernard G (1981) On the definition and criterion of death. Ann Intern Med 94:389–394
Boniolo G (2007) Death and transplantation: let's try to get things methodologically straight. Bioethics 21:32–40
Canadian Council for Donation and Transplantation (2003) Executive summary: a review of the literature on the determination of brain death. Available at http://www.ccdt.ca/english/. Accessed 10 July 2014
Capron A (2001) Brain death – well settled yet still unresolved. N Engl J Med 344:1215–1221
Chiong W (2005) Brain death without definitions. Hast Cent Rep 35:20–30
DeGrazia D (1999a) Persons, organisms, and death: a philosophical critique of the higher-brain approach. South J Philos 37:419–440
DeGrazia D (1999b) Advance directives, dementia, and 'the someone else' problem. Bioethics 13:373–391
Feldman F (1992) Confrontations with the reaper: a philosophical study of the nature and value of death. Oxford University Press, New York
Fisher J (1999) Re-examining death: against a higher brain criterion. J Med Ethics 25:473–476
Fost N (2004) Reconsidering the dead donor rule: is it important that organ donors be dead? Kennedy Inst Ethics J 14:249–260
Green M, Wikler D (1980) Brain death and personal identity. Philos Public Aff 9:105–133
Halevy A, Brody B (1993) Brain death: reconciling definitions, criteria, and tests. Arch Intern Med 119:519–525
Hershenov D (2006) The death of a person. J Med Philos 31:107–120
Holland S (2010) On the ordinary concept of death. J Appl Philos 27:109–122
Holland S, Kitzinger C, Kitzinger J (2014) Treatment decisions and the permanent vegetative state: evidence from families and experts. Med Health Care Philos 17:413–423
Kamm F (2001) Brain death and spontaneous breathing. Philos Public Aff 30:297–320
Kellehear A (2008) Dying as a social relationship: a sociological review of debates on the determination of death. Soc Sci Med 66:1533–1544
Lamb D (1996) Organ transplants and ethics. Avebury, Aldershot
Lizza J (1993) Persons and death: what's metaphysically wrong with our current statutory definition of death? J Med Philos 18:351–374
Lizza J (2006) Persons, humanity, and the definition of death. John Hopkins University Press, Baltimore
McMahan J (1995) The metaphysics of brain death. Bioethics 9:91–126
Molina A, Rodriguez-Arias D, Youngner S (2008) Should individuals choose their definition of death? J Med Ethics 34:688–689
Owen A et al (2006) Detecting awareness in the vegetative state. Science 313:1402
Potts M, Evans D (2005) Does it matter that organ donors are not dead? Ethical and policy implications. J Med Ethics 31:406–409
President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (1981) Defining death: a report on the medical, legal, and ethical issues in the determination of death. US Government's Printing Office, Washington, DC

President's Council on Bioethics (2008) Controversies in the determination of death: a White Paper by the President's Council on Bioethics. Available at http://bioethics.georgetown.edu/pcbe/reports/death/. Accessed 10 July 2014
Rachels J (1986) The end of life: euthanasia and morality. Oxford University Press, Oxford
Rich B (1997) Postmodern personhood: a matter of consciousness. Bioethics 11:206–216
Shewmon D (1998) Chronic "brain death": meta-analysis and conceptual consequences. Neurology 51:1538–1545
Shewmon D (2001) The brain and somatic integration: insights into the standard biological rationale for equating "brain death" with "death". J Med Philos 26:457–478
Shrader D (1986) On dying more than one death. Hast Cent Rep 16:12–17
Singer P (1995a) Rethinking life and death: the collapse of our traditional ethics. Oxford University Press, Oxford
Singer P (1995b) Is the sanctity of life ethic terminally ill? Bioethics 9:307–343
Truog R, Robinson W (2003) Role of brain death and the dead-donor rule in the ethics of organ transplantation. Crit Care Med 31:2391–2396
Veatch R (1975) The whole-brain-oriented concept of death: an outmoded philosophical formulation. J Thanatol 3:13–30
Veatch R (2004) Abandon the dead donor rule or change the definition of death? Kennedy Inst Ethics J 14:217–234
Wikler D (1988) Not dead, not dying? Ethical categories and persistent vegetative state. Hastings Cent Rep 18:41–47
Youngner S, Arnold R, Schapiro R (eds) (1999) The definition of death: contemporary controversies. Johns Hopkins University Press, Baltimore

Suicide

15

Steven Edwards

Contents

Abstract

This chapter will begin by signaling the importance of defining suicide. It illustrates two main approaches to the definition of suicide, one which focuses on the deliberate nature of suicide and the other which focuses more narrowly on the intention of the person concerned. Following discussions of the instrumental nature of suicide, and the issue of rational suicide, it is shown why it is held that suicide need not involve self-killing, need not require the presence of a desire to die, nor even the death of the person who suicides. Discussions of problematic cases such as altruistic and coerced suicides are also included.

S. Edwards (✉)
Philosophy, History and Law, Swansea University, Swansea, Wales, UK
e-mail: s.d.edwards@swansea.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_12

Introduction

Suicide statistics are a cause for concern across most nations, and strategies for reducing the incidence of suicide are commonly presented and revised (e.g., Her Majesty's Government 2012). The World Health Organization (WHO) organizes an annual "World Suicide Prevention Day" (see http://www.iasp.info/wspd/) in addition to compiling international data on suicide statistics across nations: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/#

According to these statistics, there are 800,000 deaths per year by suicide globally, and it is the second leading cause of death for people in the age group 10–24. Further, more people die through suicide than the sum of those dying as a result of war and murder combined. As the WHO document also makes clear, due to the fact that in many countries there are strict taboos surrounding suicide, the statistics themselves may not give an accurate picture of the true scale of the problem. It is very likely that the figure of 800,000 under represents the true number.

The WHO define suicide as "the act of deliberately killing oneself" (http://www.who.int/topics/suicide/en/) and is representative of many similar definitions which require deliberate intent on the part of the person and also the death of the person (e.g., Beauchamp 1993; McMahon 2002; Jackson 2006).

Durkheim's famous definition is more specific: "The term suicide is applied to all cases of death resulting directly or indirectly from a positive or negative act of the victim himself which he knows will produce this result" (1897/1964, p. 82).

This is ostensibly more subtle than the WHO definition in the sense that it distinguishes suiciding by acting (a positive act) from suiciding by omitting to act (a negative act). So a person who lies waiting for a train to kill them dies as a consequence of a "negative act," that is, by deliberately omitting to get out of the way of the oncoming train.

This definition is thought to differ from definitions such as the WHO definition in another way. All Durkheim seems to require is that one deliberately acts (positively or negatively) in such a way which one believes will result in one's death – i.e., even if one does not desire to die. Critics complain that it wrongly classes as suicide instances such as those which occurred as a result of the attacks on the Twin Towers in September 2001 where people felt compelled to jump to their deaths or heroic acts in which a person behaves in a way which they know will result in their death even though they would prefer to live (Beauchamp 1993, p. 74; Holland 1971; Hill 2011).

These kinds of concerns have led to the production of even more specific definitions which try to exclude those kinds of cases from the definition of suicide. An example of such a definition comes from Hill: "A commits suicide by performing an act x if and only if A intends that he or she kill herself by performing x. . ., and this intention is fully satisfied" (2011, p. 192). This definition focuses on that which is *intended*, and as will be seen below, such definitions exploit a specific way of defining "intention" which permits a distinction between acts which are intended from those which are merely "deliberate." This is all explained below.

To give a further example of a type of concern which has been raised in the philosophical literature, consider cases where a person is killed by a third party. At first sight, it is tempting to think such killings cannot be suicides because definitions of suicide typically refer to it as self-killing (as is done in the three definitions given so far). But as, for example, Frey (1981) points out, very strictly speaking the person who lies on the train track and waits for the oncoming train is killed by the train, or even the train driver, not by themselves.

These disputes illustrate that there remains considerable scope for disagreement regarding how suicide should be defined, and the definition of suicide has been a topic of debate in the philosophical literature, with some querying whether due to its multifaceted nature it *can* be defined (Holland 1969, p. 32). So in addition to the philosophical literature regarding the *ethics* of suicide (e.g., Beauchamp 1993; Cholbi 2012a, b), a further discussion concentrates on its definition and this chapter focuses on that.

There is of course good reason to be concerned that any definition of suicide which is employed is indeed accurate. If it is not, then statistics such as those cited above will be unreliable in a deeper sense than being a possible underrepresentation of the true suicide statistics.

Defining Suicide

The Instrumental Nature of Suicide

The claim that suicide is instrumental in structure looks plausible. A person suicides in order to avoid some situation or state of affairs they abhor or the prospect of which fills them with so much dread they choose to die as opposed to enduring it. For example, a person might suicide in order to avoid a period of futile suffering, or life of poor quality, or life shorn of the love of someone dear to them, and so on. Hence according to Jackson's definition: "you only commit suicide if you kill yourself from a certain motive – out of a desire no longer to live an unhappy or painful life, or in order to avoid future misery" (2006, p. 139). So, suicide is here conceived of as a means to some goal such as the avoidance of futile suffering.

More complex cases may involve an importantly different kind of goal. Someone may kill themselves in order to avoid being tortured or to save the lives of their children or some other third party. In these kinds of cases, and in contrast to the kind of case referred to in Jackson's definition, the person would, all things considered, prefer to live. Were it not the case that one was about to be tortured, or were it not the case that one's children were facing some terrible fate, one would not opt to kill oneself. In spite of this important difference, though, and a dispute about whether they really do count as suicides (Beauchamp 1993; Hill 2011), they support the suggestion that suicide is essentially instrumental in character. As seen above, the definitions also regard it as something undertaken *intentionally*, but in the philosophical literature, this term is often employed in a technical way, so it is important to move on to consider that now.

Suicide as an Intentional Action or Omission

Suppose suicide does indeed stem from an intentional action (or omission). See, e.g., Cholbi: "Suicide is intentional self-killing: a person's act is suicidal if and only if the person believed that the act, or some causal consequence of that act, would make her death likely, and she engaged in the behaviour to intentionally bring about her death" (2012a, p. 21).

How should the term "intentional" be understood in this context? In the philosophical literature, it is common to hold that *intentions* are made up of beliefs and desires and that explanations of the actions of others refer to their intentions (see, e.g., Lennon 1990). To see this, consider that Smith may desire a glass of milk and believe she has a bottle in the fridge, so she walks to the fridge. If it is asked "why did Smith walk to the fridge?" an explanation of this is supplied by reference to the beliefs and desires just described. Also, it is presumed here that Smith is acting rationally; there is a coherence between what she believes and desires and the act she performs which makes it a rational action. In the context of suicide, a person might believe that life in the future offers nothing more than constant frustration and suffering, desire to escape such suffering, and so act to bring this goal about. Because of this combination of mental states, the intention to take their own life may be entertained. Here again, there is coherence between belief, desire, and action which enables one to explain the action of the person. It can be seen why, if the person really did hold or does hold those beliefs, and really did have or does have those desires, they acted as they did or might try to act in order to bring about their own death.

This understanding of rational action is formalized in the quote below from Nordenfelt:

> A wants [desires] to bring about P, A believes that he is in situation S, A believes that he must do F in order to bring about P, A believes that he can do F, A does not have any reason for abstaining from doing F. (Nordenfelt 2007, p. 91)

So, to consider the kind of case alluded to by Jackson, a person A desires to avoid the burden of enduring futile suffering (they want to bring about P). A believes that her situation is such that only through suicide can she achieve this goal: A believes she is in situation S and must do F – commit suicide – in order to bring about P. A believes she can commit suicide (perform action F), and as far as A is concerned, she can't see any reason for abstaining. So if suicide is thought of as a rational action stemming from the intentions of the person, it would follow that explanations of it are subsumed by the pattern given by Nordenfelt.

Rational Suicide

However, one key question has been whether or not there can be such a thing as a rational suicide (Bloch and Heyd 1981). It might be held that there cannot be such a thing because the very desire itself – the desire to commit suicide – is intrinsically

irrational. One might think this because in choosing to end one's life one chooses to foreclose the possibility of any future choices. So in a sense, one negates the very capacity through which one makes such a decision. In exercising one's autonomous will, one is acting so as to eradicate the possibility of the future exercise of autonomy, and so, it may be held there is something self-contradictory about such an act which therefore shows it cannot be a rational act (Kant 1785/1948, p. 85; Schramme 2013). If it is the case that suicidal intentions are intrinsically self-contradictory, it would follow that they are intrinsically irrational – since to contradict oneself is to hold a proposition to be both true and false at the same time and to contravene the most basic law of logic (the law of noncontradiction).

So, can one, after all, suicide on the basis of an irrational intention – an irrational combination of beliefs and desires? This question obviously presupposes the distinction between rational and irrational intentions. Suppose someone kills herself as a result of a belief which is caused by a serious mental health problem (Hewitt 2010). Perhaps the person believes they have been instructed to kill themselves, and must do as instructed, yet the instructions stem from auditory hallucinations. Strictly speaking although colloquially one might say this person acted irrationally, their beliefs, desires, and actions exhibit a pattern of coherence which would seem to be sufficient for their act to count as rational. To refer back to the characterization of rational acts given by Nordenfelt and quoted above, consider that A believes she must obey voice V; A desires to follow V's instructions; A believes that in order to obey V, she must suicide; A can suicide; A has no reasons for abstaining from suiciding.

Although one might think this is irrational due to the falsity of the beliefs (it is of course false that she has to obey V), the pattern of claims is logically coherent and not self-contradictory. Indeed this would be brought out in an explanation of the person's suicide. Given what they believed to be the case, one can understand why they acted as they did.

On another way of thinking about how to define rational action, however, such an act might be defined as irrational. Hence according to Culver and Gert's definition of irrational actions, they are acts which "consist of harming oneself without an adequate reason" (1982, p. 28). This of course leads to the determination of what indeed is to count as an "adequate reason." In the example just discussed, as far as the person is concerned, she has an adequate reason in that she has been instructed by a person she must obey to suicide. So the fact that a false belief leads one to suicide need not render the suicide irrational – even if one of the beliefs is false and is due to the presence of a serious mental health problem. The pattern of reasoning here is not illogical, and nor is it irrational in the sense of having an incoherent – self-contradictory – intentional structure. To bring this out further, an incoherent and thus *irrational* intentional structure looks as follows:

> Assume that A wants to have P. A also believes that having Q is a necessary condition for getting P. Then presumably A must also want to have Q (given that he or she has no independent reason for wanting to reject Q). If a person does not in fact want Q in such a situation, then he or she is irrational. (Nordenfelt 2007, p. 106)

If a hungry person wants to relieve their hunger (P), and believes that the food in front of them can relieve their hunger (Q), then the person must want the food (absent any other reason for rejecting the food such as being on hunger strike). But the person who believes they must do as instructed by the internal voice they hear, and so ends their own life as instructed, is not thereby acting in an irrational manner. There is no internal contradiction in their beliefs, desires, and actions. So they exhibit no irrational combination of reasons, nor illogical ones. An illogical one, on this approach to rationality, would require a person both desiring and not desiring to commit suicide (both desiring and not desiring Q). So as explained, the person's actions satisfy the criteria for being rational actions given above.

It can be seen then that although it has been held that suicidal desires are intrinsically irrational, they need not be unless they stem from contradictory beliefs and/or desires. Even those which stem from beliefs which are patently false need not be irrational, unless one agrees with a definition of irrationality such as that offered by Culver and Gert and which builds in reference to what is typically believed to be true in a particular cultural context. Having discussed the question of the intrinsic irrationality or otherwise of suicidal acts, it was seen above that suicide is regarded as self-killing, so it is necessary to say a little on this subject too.

Self

Suicide is usually thought of as some form of self-killing. There are references to "mass suicides" too of course such as that which occurred in Jonestown, Guyana, in 1978 when 918 people killed themselves while under the influence of the ideas of Jim Jones, an American religious figure (see Layton 1988). But a "mass suicide" is, rather, a collection of individual suicides. What makes such an event a "mass suicide" is that it is composed of multiple self-killings.

But what is a self? This question raises a large and difficult set of philosophical questions (e.g., Parfit 1984). For example, it may be argued that more than one self can inhabit the same body – e.g., in cases of multiple personality. So ending the life of one human body might at the same time comprise the ending of the lives of several selves (Nagel 1986). Even if one disregards such possibilities as being too philosophically exotic, problems arise regarding the possibility that selves and their bodies may be distinct in another way. Suppose a person with severe dementia bears no psychological resemblance to the person they were prior to the onset of the disease. It may be claimed that the self who previously "occupied" that body ceased to exist, even though the body itself continues to function (see De Grazia (2005) on the "someone else problem"). Should that later self proceed to end their own life, the self who kills themselves need not coincide with the self who inhabited the body prior to the onset of dementia it may be claimed.

The problem of the relationship between the identity of the self and the human body actually killed arises in a slightly different way in an example provided by Hill (2011). He conceives of a person JS who happens to have amnesia. He learns of the existence of a person of the same name and plots to kill him – unaware that he is the

said JS. He then arranges for JS to be given a lethal dosage of drugs, unaware that he will be the one who will receive the lethal dose. He self-kills and intentionally kills JS, but because he does not realize that *he* is JS, he does not realize that the death of JS is also his own death. This is why in Hill's definition, he builds in the condition that the person who intends to kill themselves does not make this kind of identity error ("A commits suicide by performing an act x if and only if A intends that he or she kill himself or herself by performing x (under the description "I kill myself"), and this intention is fully satisfied" 2011, p. 192).

Assuming that selves and bodies do coincide, which is after all true in typical cases of suicide, suicide is more than mere self-killing since not all self-killings constitute suicide. To drink a fatal amount of poison believing it was water would amount to a self-killing, but not a suicide; this is because there is no intention (no desire to die) to self-kill. So the suicide must stem from some intention entertained by the suicide and the question then arises regarding the nature of that intention.

Intentional Self-Killing: Intention and Outcome

One might suppose that the presence in a person of an intention which includes a desire to die is sufficient to clarify that a death is a suicide. But a complication arises concerning the extent to which the intent matches the outcome. Using the schema from Nordenfelt described above, suppose A desires to avoid future suffering, believes that drinking the poison in front of them will end their life, and so drinks the poison. But A has not taken enough of the poison, staggers into an oncoming car, and dies. Here the ultimate cause of the death is not the poison but the collision with the car.

So must one die in the way one had intended to die in order to have committed suicide? Beauchamp holds: "An act or omission is a suicide if person intentionally brings about his or her death, unless the death (a) is coerced or (b) is caused by conditions that are not specifically arranged by the agent for the purpose of bringing about the death" (1993, p. 79). And further it is confirmed that death must occur "in accordance with the final plan selected by the agent" (1993, p. 80). Much will depend upon how specific this final plan is. In the example of A above, since his final plan is to die through poisoning, it would seem this is not a suicide since dying through being hit by a car plays no part in his intention. Of such a case, it could be said they died while in the process of or having attempted suicide. Hill goes even further than Beauchamp in claiming that one's intention must be "fully satisfied" (2011, p. 192) in order for it to be the case that one suicides. So the person, A, hit by the car definitely does not suicide on that view.

Perhaps it is best to allow some scope for flexibility rather than requiring an exact match between plan and outcome. (Beauchamp acknowledges that his definition does not "eliminate all problems of imprecise boundaries" (1993, p. 83).) To see this, consider someone who intends to kill themselves by jumping in front of the 7.30 train from Watford Junction to London Euston – perhaps because this is the train

they caught on their journey to work for decades. They duly jump in front of the train they reasonably assume is the 7.30 Watford to Euston (they desire to be killed by that train, believe the train that is on its way is that train, and so lie on the track). But they are wrong; that train is late and the one they jumped in front of is the delayed 7.15 Watford to Euston. It would be strongly counterintuitive to claim their death was not suicide because the train they jumped in front of was not the train they intended. Even though their intention is not "fully satisfied" in the sense that the train which killed them was not the train they had intended would kill them, it still seems plausible to think the person has suicided and done so intentionally.

Yet, now think of a different "train" example. This time, the person intends to lie on the track and wait for the 7.30 train; as they lie there, they are struck by lightning and killed. Here the cause of death seems so different from the cause intended by the would-be suicider that it seems accurate to deny they in fact suicide. This seems correct since as Beauchamp suggests there is too much of a divergence between the "final plan" of the person and what transpires. Of such a case, one may say they died while in the process of attempting suicide, and so their intention was one which is so far from being "fully satisfied" it fails to qualify as being a suicide.

A different kind of problem arises when what one intends and the chosen means do actually match up, and a controversial way of defining suicide which deals with cases of this kind, defended by Fairbairn (1995), will be described later. For now, suicides which involve third parties will be discussed in more detail. As mentioned, strictly speaking the "train track" example referred to above shows that one can suicide without killing oneself: one can be killed by a third party, it seems, and yet be deemed to have killed oneself.

Involvement of Third Parties

It was seen above that strictly speaking one need not kill oneself in order to suicide. The paradigm example of this is the kind of "train track" case referred to above. But there are less clear cases. For example, Cholbi refers to cases in which a terminally ill patient requests that a third party kills her where in his view such a person has "intuitively committed suicide" (2012b, p. 3). But in so-called assisted suicide, although the physician may set up the intravenous drip containing the lethal liquid, the fact that it is the patient who performs the action which releases the liquid into her vein, causing the death, makes such a situation an instance of suicide. One might say that this in fact is what distinguishes assisted suicide from euthanasia (Hill 2011, p. 194). In the latter – paradigmatically in active voluntary euthanasia – the last act, say the administering of a lethal injection, is performed by a third party.

Even more complex cases include situations where the would-be suicider manufactures a situation in which he will be killed by someone else. For example, a person might deliberately, and with the intent to die, behave as if they intend to shoot a police officer whom they know is armed and will shoot them if so threatened (Frey 1981, pp. 198–201; Cholbi 2012a, p. 3). They duly do so and achieve the desired

result. So there is no doubt that as in the "train track" example referred to above, the would-be suicider creates a situation in which their death is almost inevitable.

However, there does seem enough of a difference between the train track example and the other two just described (suicide by armed police officer and the terminally ill patient case) to warrant claiming the first is clearly a suicide, while the latter two may not be. The causal chain in the first case is so direct that, barring a very remarkable event indeed, the death of the person is inevitable. This is much less obvious in the other two cases (those concerning the police officer and the terminally ill person) where the intentions of the police officer and the physician play a much more central role in the causal sequence leading to the death of the would-be suicider.

So although in the police officer and terminally ill person cases, their deaths have the tacit agreement of the individuals concerned, the fact that their deaths involve the intentions of others to a significant extent weakens the claim that these are suicides. The officer may decide to shoot to wound rather than kill; the doctor's nerve may fail her and she may refuse to go through with the administration of the lethal substance. There is no echo of such a gap in "train track" case. True, the engine driver may see the would-be suicider and try to stop the train, but the death is immune from anything under the control of the train driver. Frey is willing to accept in police-officer-type cases that if the officer does *not* intend to kill the would-be suicider who has deliberately engineered the situation in order to get killed, then the act *is* accurately classed as suicide. But he suggests that if there *is* intent to kill on the part of the officer, then there is more cause to doubt that the death was a suicide: rather it was unintentional killing by the police officer (unintentional because the police officer would not have shot the would-be suicider had she known the facts).

What Frey says here though aligns with the general conclusion that the more scope there is for the intent of others, the weaker the case for classifying the death as suicide, and conversely, the less scope for intrusion of the intentions of others, the stronger the case for classifying the death as suicide.

A parallel analysis can be given in cases where, for example, a patient refuses life-saving treatment. Here – assuming the treatment is easily accessible – the patient chooses a course of action which will inevitably lead to her death. Such cases are conceptually on a par with the "train track" case in which the person is killed by the train but intentionally causes their own death by omitting to move. Thus in order for an act or an omission to qualify as suicide, there has to be a very direct causal link between the person who suicides ("the suicider") and the outcome. In cases of "assisted suicide," although a third party may provide the means, e.g., provide a lethal substance installed for intravenous infusion, the last act, so to speak, must be performed by the suicider. If the person who provides the substance and sets up the infusion also initiates the flow of the lethal liquid into the body of the patient, then that person has ended the life of the patient, and so it is not a suicide.

Although the cases just discussed involve third parties, they are cases in which the would-be suicider intends and so desires their death. But a further problem concerns cases of deliberate self-killings but which some commentators deny as suicides because the person involved would actually prefer to live.

Intentional Self-Killings Which Are Not Suicides?

As mentioned in the introduction to this chapter, this possibility arises due to a more nuanced understanding of the term "intention" which is found in the philosophical literature. In ordinary language, one would typically think of the terms "deliberate" and "intentional" as equivalent in meaning, but this is not so according to those in the field of philosophy of psychology. Indeed the very possibility that there can be intentional self-killings which are not suicides arises at least partly due to the difference in meaning between "deliberate" and "intentional." The problem which ensues is in cases where a person acts deliberately, but it is claimed, not from a desire to die.

Deliberate actions are correctly regarded as intentional (as opposed to unintentional), and as seen, intentions are regarded in philosophical literature as being composed of beliefs and desires. But in situations where it is plausible to suppose that a person acts intentionally but does not desire to die, since what is "deliberate" and what is "intentional" seems separable, conceptual problems ensue: "the semantic similarity between "intention" and "intentional" can be taken too far and [in some cases] restricting "intentional" only to a person's intentions whether ends or means is wrongheaded" (Cholbi 2012a, p. 27; Searle 1983, p. 83).

The examples considered so far involve people who do in fact intend to die and would rather die than live any longer; such people act deliberately and desire to die. But the problem regarding the interpretation of the term "intention" is illustrated by those whom one might think of as having been led to self-kill by the circumstances in which they find themselves. All things being equal, they would prefer to live, and were they to survive their suicide attempt and be released from the conditions which led them to end their life, they would be pleased to have survived. Such a situation is likely to have arisen in the tragic events of September 11, 2001, for example. Those who jumped from the Twin Towers to avoid being burned to death would prefer to have lived. But given a choice between death from jumping off the Tower and exposing themselves to being burned, they opted for the former. Their death by self-killing is a means to avoid being fatally burned. So self-killing is a side effect it may be said of their escaping from a horrific form of death (Hill 2011). If by some unlikely event something were to have broken their fall, and they survive the jump from the tower, they would have been pleased.

If one thinks of the intentions (beliefs and desires) of the people in this terrible situation, they apply to cases in which a person intends to end their own life in the sense of acting deliberately, selects the means appropriate, and achieves the chosen end "successfully." Yet, there is legitimate concern about whether their deaths ought really to be classified as deaths from suicide (Beauchamp 1993, p. 74; Hill 2011). The people here desire to avoid being fatally burned, believe that jumping from the Tower will bring that about, and so they jump. Since – in contrast to the case of A discussed above – there is no explicit desire on the part of the person to end their life, it might reasonably be said it is inaccurate to say they suicide. Similar concerns might be raised regarding so-called suicide bombers, and indeed Hill (2011) argues for similar reasons that they are not correctly described as suiciding. Rather, in such

cases, it is said that the death of the person is a foreseen but not intended (because not desired) consequence of either avoiding the flames in the Twin Towers case or defeating the enemy in the "suicide bomber" cases.

Alternatively, it might be thought that this is a rather manufactured description of the psychological states of the people involved and couch the problem in terms of ranked preferences, say. Thus one may say the person prefers to live, but given the need to select between two options, each of which will have a fatal consequence, the person manifests a strong preference to avoid death by burning by jumping. Or, having calculated the options open to them, the person prefers to avoid death by burning, believes they can avoid this by jumping, knows this will result in their death, and so prefers to jump – even though, were they to survive, they would be pleased.

This analysis tallies with what Frey suggests about Captain Oates' death (1981, p. 195). Oates was concerned that his ill health was jeopardizing Scott's Antarctic expedition in 1912. So in order to not be a burden to his colleagues, he simply left his tent and walked into the frozen Antarctic wastes. Of this case, Frey writes, "it is. . . reasonable to argue that, having taken stock of his situation and Scott's, Oates wanted to die in order to allow Scott to carry on" (1981, p. 195). Thus in both cases, one finds a deliberate act stemming from a choice between two options – at least as those in the Twin Towers and Oates saw their respective predicaments – and a preference for one option over the other.

In Durkheim's definition, such cases (those involving the Twin Towers, the suicide bomber, and Captain Oates) are cases of suicide because their death results from "a positive or negative act of the victim. . . which he/she knows will produce this result" (op.cit., p. 44), but on the definition of Hill, and possibly the WHO definition, depending upon how the term "deliberately" is interpreted, they did not. If "deliberately" means "on purpose," i.e., intentionally, then there is agreement between the WHO definition and that of Durkheim. But if "deliberately" means "done with an intent to die" – and therefore with a desire to die – the two definitions will differ in how they regard this complex kind of case.

First- and Second-Order Desires?

To try to shed light on this apparent puzzle, a distinction between first- and second-order desires may be appealed to Frankfurt (1982). The desire to avoid death from burning leads the person to jump; the desire to avoid the burning can be regarded as a first-order desire. All things being equal, the person would have preferred to live. So at the second-order level, there is a desire to live. But due to the tragic events in which they have been caught up, they are faced with a tragic choice and take the decision to jump. Similarly with the suicide bomber, were their political or religious context sufficiently different, they would not choose to act in a way which results in their death.

By contrast, in what might be thought of as "uncomplicated" suicide, the first-order desire to die coheres with a second-order desire to die (cf. Jackson's definition

given above). In such cases, there is no conflict between the two levels of desire. The person desires to die and then selects the appropriate means. They may even leave evidence to confirm that this is their desire, explaining the reasons behind their action. By insisting on congruence between these two levels of desires, the claim that it would be inaccurate to classify those who jumped from the Twin Towers as suicides can be articulated, as can the view that suicide is more than simply deliberate self-killing: it must be self-killing where there is a desire to die at the two levels so to speak.

But the strategy to use this distinction to argue that some intentional self-killings are not suicides risks emptying the whole category of suicidal acts. To see this, consider again the proposal that the person who suicides but would rather carry on living is not suicide. Think of this in the context of a case in which a person takes their own life because they have an incurable chronic medical problem which causes them considerable suffering and disability. They would rather live without the chronic health condition, but cannot tolerate living with it. So just as it might be said that the unfortunate person on the burning tall building really wants to live, but in order to escape the flames they jump, so one might say the person with the chronic disease would prefer to live (without the condition) but takes their own life to avoid the only other option. To deny that the person who jumps commits suicide surely implies that the same is true of the person who takes their own life to escape the incurable chronic disease. And this risks emptying the concept of suicide of any content because for any set of conditions which drive a person to suicide, it can be imagined that, were things otherwise, they would not desire to kill themselves.

Alternatively, as seen, it can be argued that there is in fact no intent to suicide present in such cases (because of the absence of a desire to die). Rather, the death by jumping is a foreseen consequence of the desire to escape from the flames.

Altruistic Self-Killings

A further category of disputed self-killings is those undertaken for altruistic reasons, for the sake of others. Cholbi's discussion of "foxhole jumper" illustrates the kind of case under discussion here (2012a, pp. 15–38). These are cases where a soldier deliberately throws his own body to absorb the blast of an exploding shell in order to save the lives of his colleagues. Cholbi's definition of suicide, given above, also refers to its involving "intentional self-killing" (2012a, p. 21). And he says of such cases that their deaths have the "rational endorsement" of the soldier; this due to the fact that "It strains credulity to deny that the prospect of death does not. . . play a role in Foxhole jumper's deliberation about [her] choices" (ibid., p. 21). So Cholbi is willing to concede that there is no desire to die in such cases, but since he defines suicide as "intentional," he owes us an account of how this can be the case. His answer is to remind us of the kind of ambiguity regarding the terms "deliberate" and "intentional" that were described above as well as to exploit the notion of rational

endorsement. So since in the Twin Towers, suicide bomber, and Captain Oates' cases, the likelihood of the person dying is foreseen, since their acts have their rational endorsement, they are accurately characterized as suicides as Cholbi (2012a) sees these.

Searle draws a distinction which seems to fit well with Cholbi's analysis. Searle distinguishes between intentions which precede acts and those which don't (1983, p. 84). The person who plans their death on the railway track entertains intentions prior to the act and then acts, but some acts – for example, the kind of heroic act in the "foxhole jumper" cases – seem closer to reflex actions. For Searle, they manifest intention in action. Since such acts would be commended – the soldier might receive a posthumous medal for bravery - it does not look plausible to deny they are taken deliberately. Hence the foxhole jumper *is* acting intentionally, even if – with Cholbi – one allows that dying is not part of his intention in the sense that there is no desire to die. On Searle's analysis, such acts are intentional and manifest "intention in action."

Coerced Self-Killing

A further class of cases which present the kind of ambiguity found in the Twin Towers and altruistic self-killings cases are to be found in "coerced" self-killings (Beauchamp 1993, p. 79). Much will depend upon how the term "coerced" is defined, but usually it is defined in such a way that one person can only be coerced into doing something by another person (Wilkinson 2003, pp. 82–98; but see also Cholbi 2012b, p. 5 for a broader definition). For example, Beauchamp discusses a case in which a woman is faced with a choice between killing herself and her children being killed (1993, p. 81). Suppose she is presented with such a choice and opts to kill herself. Beauchamp denies her self-killing could correctly be described as suicide. This is because, although intentional (deliberate), the action is not autonomous he suggests. An autonomous action, he argues, is one which is not coerced. This is because autonomy is usually understood as involving the ability to be self-governing, that is, to be autonomous. When one is coerced, a person's autonomy is eroded by that of the coercer. Hence it can be seen why Beauchamp considers coerced acts not to be autonomous. But, to use an example of Cholbi's, consider that someone is coerced into singing a song (perhaps they are threatened with being shot unless they do so), a coerced singing is still a singing; and so, a coerced suicide is still a suicide it may be said (2012b). However, assuming that in order to act deliberately, it is necessary that one acts autonomously – i.e., one is "self-governing" – it is legitimate to query who is correct here, Beauchamp or Cholbi.

Outcome

The final element of the WHO definition given above concerns outcome. Suppose the person who intends to end their own life by shooting themselves misses or

survives. Did they commit suicide? Ordinarily this would be thought of as an attempted but not an actual suicide. But consider Fairbairn's definition of suicide:

> Suicide is an act, whether of commission or omission, and whether performed by himself or others, by means of which an individual autonomously intends and wishes to bring about his death because he wants to be dead or wants to die the death he enacts (1995, p. 84).

Unusually this definition does not require that a person actually dies after having tried to self-kill. According to Fairbairn, a serious intent and undertaking of the relevant action are sufficient – even if the person survives. The reason why he proposes such a definition is because he thinks the role of intention is central and not given sufficient weight. The main consideration, according to Fairbairn, is whether or not the person really intended to end their own life and whether the means they pursued to end it were indeed likely to succeed. So the person who takes enough of a drug to kill themselves, but who knows someone will appear and save them, or call an ambulance, does not commit suicide on Fairbairn's view – even if they do in fact die. But a person who intends to end their own life, and takes steps which they believe will bring about their own death, does commit suicide. If, by accident as it were, they are discovered and saved, they still committed suicide by Fairbairn's account. An example he gives is that of the writer Sylvia Plath. She apparently took enough sleeping pills to end her own life and hid herself away in a remote cellar, even hiding herself within the cellar itself. As things turned out, she was discovered and brought 'round. Again, on Fairbairn's account, she committed suicide. The implication of Fairbairn's view is that some currently living people who are said to have attempted suicide should, strictly speaking, be regarded as having actually committed suicide. Needless to add, this is an extremely radical view.

In summary, then, beginning with the WHO definition of suicide, it was made clear that providing a definition is an important but difficult task. Attention was drawn to the apparent instrumental nature of suicide – it being a means to some end and also to the question of whether suicide is intrinsically irrational. It was seen that three common intuitions regarding suicide can be called into question quite plausibly: the intuitions that suicide is self-killing, that it requires the desire to die, and that it requires the death of the person. Challenges to each of these were considered. It was shown that suicide certainly does not require self-killing in a strict sense of that term; it was also shown that the question of the requirement of a desire to die is extremely thorny, depending upon the role played by the intentions of the person. For some theorists (e.g., Durkheim), a person can accurately be described as having suicided even if they did not desire to die. But for other theorists (e.g., Hill, Beauchamp), the presence in the suicide of a desire to die is necessary; otherwise their death is a foreseen consequence of some other acts they intend. It was also seen that Cholbi's definition tries to straddle these two ways of thinking about suicide by arguing that the distinction between them is too crude and stems from a flawed way of thinking about the relationship between deliberate and intentional acts. In the final section of the chapter, the radical thesis that suicide need not require the death of the suicide was also described.

Definitions of Key Terms

Intentional To describe an action as intentional is to say either (a) that it is done deliberately or further (b) that it stems from specific beliefs and desires.

Summary Points

- Suicide is a means to some end.
- It is not intrinsically irrational.
- It need not require self-killing.
- It need not require the desire to die (according to some commentators).
- It need not require the death of the person (according to one well-known approach).
- Defining suicide is problematic partly due to the terms "intentional" and "deliberate."

References

Beauchamp TL (1993) Suicide. In: Regan T (ed) Matters of life and death: new introductory essays in moral philosophy. McGraw-Hill, New York, pp 69–120
Bloch S, Heyd D (1981) The ethics of suicide. In: Bloch S, Chodoff P (eds) Psychiatric ethics. Oxford University Press, Oxford
Cholbi M (2012a) Suicide, the philosophical dimensions. Broadview Press, Ontario
Cholbi M (2012b) Suicide. Available at http://plato.stanford.edu/entries/suicide/. Accessed 17 July 2014
Culver CM, Gert B (1982) Philosophy in medicine: conceptual and ethical issues in medicine and psychiatry. Oxford University Press, New York
De Grazia D (2005) Human identity and bioethics. Cambridge University Press, Cambridge
Durkheim E (1897/1964) Suicide. In: Jones RA (ed) Emile Durkheim: an introduction to four major works. Sage, Beverly Hills, pp 82–114
Fairbairn GJ (1995) Contemplating suicide: the language and ethics of self harm. Routledge, London
Frankfurt H (1982) Freedom of the will and the concept of a person. J Philos 68:5–20
Frey RG (1981) Suicide and self-inflicted death. Philosophy 52:193–202
Her Majesty's Government (2012) Preventing suicide in England. Her Majesty's Government, London
Hewitt J (2010) Rational suicide: philosophical perspectives on schizophrenia. Med Health Care Philos 13:25–31
Hill DJ (2011) What is it to commit suicide? Ratio 24:192–205
Holland RF (1969) Suicide. In: Vesey G (ed) Talk of god. Macmillan, London, pp 72–75
Holland RF (1971) Suicide. In: Rachels J (ed) Moral problems. Harper and Row, New York, pp 352–354
Jackson J (2006) Ethics in medicine. Polity Press, Cambridge
Kant I (1785/1948) Groundwork of the metaphysic of morals. In: Paton HJ (ed tr) The moral law. Hutchinson, London
Layton D (1988) Seductive poison. Anchor books, New York

Lennon K (1990) Explaining human action. Duckworth, London
McMahon J (2002) The ethics of killing. Oxford University Press, Oxford
Nagel T (1986) Mortal questions. Oxford University Press, Oxford
Nordenfelt L (2007) Rationality and compulsion: applying action theory to psychiatry. Oxford University Press, Oxford
Parfit D (1984) Reasons and persons. Oxford University Press, Oxford
Schramme T (2013) Rational suicide, assisted suicide, and indirect legal paternalism. Int J Law Psychiatry 36:477–484
Searle JR (1983) Intentionality: an essay in the philosophy of mind. Cambridge University Press, Cambridge
World Health Organisation (WHO) Suicide. Available http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/#
WHO Suicide prevention day. Available at http://www.iasp.info/wspd/. Accessed 17 Jan 2014
Wilkinson S (2003) Bodies for sale: ethics and exploitation in the human body trade. Routledge, London

Enhancing Human Abilities and Characteristics Beyond Normality

16

Andrew Bloodworth

Contents

Abstract

This chapter concerns the ethics of enhancing human abilities and characteristics beyond normality. It begins by exploring what enhancing in this context means and how it might be distinguished from other commonly accepted enhancement practices, such as training or education. It then moves to explore some key arguments that oppose such practices. The first (Kass LR (2003) Ageless bodies, happy souls: biotechnology and the pursuit of perfection. New Atlantis Spring:9–28. Available online http://www.thenewatlantis.com/publications/ageless-bodies-happy-souls. Accessed 29 Mar 2011) on the grounds that such methods threaten our agency as human beings; we as human beings would ultimately be less responsible for our achievements. The second (Sandel MJ (2007) The case against perfection. The Belknapp Press of Harvard University Press, Cambridge, MA/London) argues against certain practices as reflecting a problematic attempt to master our human nature. After exploring the limitations of such arguments, liberal arguments that support such enhancement practices are addressed. Such arguments often stipulate the need to ensure safety and the

A. Bloodworth (✉)
Department of Inter-Professional Studies, College of Human and Health Sciences, Swansea University, Swansea, UK
e-mail: a.j.bloodworth@swansea.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_60

fair distribution of enhancement technologies. Further arguments support such enhancements as reflective of rather than threatening our human nature. On such a view, what is important about us as human beings is our striving to be better and making choices to support this.

Introduction

This chapter is concerned with the arguments supporting the enhancement of human capabilities beyond normality and with those arguments that oppose such enhancements. It will begin by examining key terms operating in this high-profile debate. This will be followed by a description of key arguments both for and against such practices and common challenges to these arguments. Often those opposed to enhancement of human capabilities beyond a norm are referred to as bioconservatives. Here Sandel's (2007) *Case Against Perfection* and Kass' (2003) conservative stance will be considered. More liberal arguments such as positions expressed by Savulescu (2007) and Harris (1998) will then be examined.

Enhancement itself is a difficult term to pin down. In its loosest sense, in referring just to increases or improvements, there seems to be little to object to. Many medical treatments improve our well-being or function. Education itself, as Savulescu (2007) points out, is designed to improve capacities. Often those claiming that certain forms of enhancement are ethically dubious are referring to a subset of enhancements that utilize medical means to enhance performance beyond normality. There are two important and related claims evident here: first the suggestion that there is something about the method itself that is problematic and, second, that using this method elevates capacities beyond a *normal* level for that person and that this is ethically problematic.

Education, an example already referred to, may well elevate the capacities of the individual beyond previous norms achievable for that person. There has been interest, however, in how certain medications might also prove useful in enhancing cognitive capability, memory, for example, or concentration (Greely et al. 2008; Sahakian and Morein-Zamir 2007). Arguments suggesting that the use of medical means for enhancement purposes is ethically problematic emphasize the different levels of agency evident in each of the enhancements (Kass 2003). Education requires an active agent seeking to learn and develop. Merely taking a drug and reaping the cognitive reward appears to offer a shortcut of some kind.

Taking a tablet to return to normal levels of functioning appears less controversial. An academic struggling with illness who takes a drug in order to restore normal levels of functioning appears merely to be facilitating a return to work, not taking an illegitimate shortcut. Other examples, however, seem to test our intuitions in this field a little more. First other methods of enhancing beyond a norm in which the individual seems somewhat passive are often considered benign. An individual might take caffeine in the form of a cup of coffee to give them a boost; they might even lay off coffee for some time before an athletic event and then have a couple of

cups before the event in the hope that it reduces the fatigue levels experienced. This sort of example makes our objection to certain methods look somewhat arbitrary. Taken in a drink commonly consumed an enhancement is considered ethically permissible; in a tablet designed for medical purposes, the practice is argued to be ethically problematic.

It can also be difficult to ascertain whether the reference to normality can hold strong against certain examples. The claim might be described as follows. It is of course legitimate to utilize medical means to enhance our recovery, for example, physiotherapy to speed our recovery from a strained muscle or antidepressants to treat a medical condition and allow some sort of return to normal mood levels. It is unacceptable ethically, however, to use such means to enhance capabilities beyond normality. For a cyclist to take a drug to enhance their oxygen-carrying capabilities ultimately means their performance is improved, beyond the previous norm for the individual, or for a fully functioning academic to take a tablet that elevates their capacities. Again, however, other examples seem to jeopardize these initially solid-looking foundations. Is any speed of recovery permissible so long as ultimately the normal performance levels for that individual are not exceeded? Physiotherapy and massage might seem commonplace methods of speeding recovery from injuries for athletes. What about the use of platelet-rich plasma (PRP) therapy, where blood is removed and centrifuged and the platelet-rich part reinjected with the intention of speeding recovery (see Moraes et al. 2014)? If one is only concerned with restoring a norm, and can ascertain that this will not exceed it, it might be said that there is no ethical problem to discuss. Others might, however, claim that this is an illegitimate shortcut in recovery from injury. It would be very difficult to draw a line, however, that distinguished between legitimate and illegitimate methods of recovery. It would be a stretch, for example, to consider physiotherapy or massage as less artificial or unnatural than PRP therapy, which after all involves a reinjection of one's own blood.

Ascertaining just what constitutes normal function for an individual and what methods will restore but not exceed this normal function must be a difficult task. One such way would be to look for a deviation in the functioning of the parts of the individual. Where everything is working as expected, enhancing cannot be justified; where there appears to be a deviation in the functioning of parts of the organism, a disease, the therapeutic measure can be justified. In this approach, influenced by Christopher Boorse's (1975) approach to health, therapy is viewed as the use of medical means to treat a scientifically identifiable condition, a disease that is essentially a deviation from a statistically identifiable norm caused by a subfunctioning within the organism. There are challenges, however, to this conception of health and the over reliance on the notion of disease. Examples can be offered where the function of two individuals taken as wholes is equally impaired. But in one instance, the impairment cannot be tracked to a disease. If both individuals suffer the same level of impairment, however, one might argue that they are equally in need of medical intervention. One response to this problem would be to seek a notion of normal function that does not rely upon the notion of disease.

Arguments Against Enhancement

Having raised some difficulties in applying key terms, and some difficulties in distinguishing between different forms of enhancements, the key arguments will now be considered. First the bioconservative arguments opposing certain forms of enhancement will be addressed. Methods utilized to enhance human capabilities might be objected to on a number of grounds: that they are unsafe or that an unacceptable level of injustice might result from only certain people being able to access such enhancements. The arguments focused on here though are deeper objections, working upon the basis that even if these concerns over safety and fairness can be alleviated, these methods might threaten human freedom or indeed human flourishing (Sandel 2007) in some way.

A common objection already briefly referred to concerns the extent to which the individual is active in the process of enhancing. Kass (2003) opposes certain enhancements on the basis that they threaten human agency. Taking a drug to enhance levels of concentration beyond a norm means that the person is somehow less responsible for the resulting work. Likewise even in a (hypothetical) world in which all methods of enhancement were permitted in sporting competition, athletes utilizing such methods would be less responsible for their achievements however astounding. The argument is essentially that means matter (Kass 2003). It matters how human beings go about things; a good life for a human being suggests Kass is one in which the person is dedicated and strives to achieve their aspirations, not one in which they seek medical intervention to enable such achievements. Upholding a distinction between therapy and enhancement further supports this argument. Medical intervention is necessary in order to recover, but not permissible as a means of exceeding our human limitations. Once such methods to enhance have been utilized, we are no longer striving as human beings to achieve within our limitations, but are in some important ways less responsible for our medically enhanced performance.

Similar moves have been made in the context of genetic intervention prior to birth. Intervention to protect against certain illnesses or disabilities might be regarded in some quarters as therapy and justifiable. Interventions that intend to maximize certain characteristics, intelligence perhaps, or moral capacities, threaten the eventual freedom of the unborn child. Such characteristics are less a result of the choices made growing up and more the product of design over which the child had no option. Not only has the person doing the enhancing treated the embryo or unborn child as something to be manipulated; once the person discovers that he/she has been manipulated in this way, his/her own sense of agency and responsibility for their lives will be affected (Habermas 2003; Herissone-Kelly 2012).

There are some challenges to these arguments. First the claim that the achievements are no longer those of the persons who have been enhanced, that they are no longer responsible, has been subjected to some critique. In the context of interventions that work at a genetic level, the complex relationship between genes and, for example, sporting performance needs further explanation. To say that genetically enhanced sporting performance is no longer the responsibility of the athlete

overstates the role of the genetic intervention and understates the continued striving and training undertaken by the athlete. Murphy (2014, p. 243) makes this point, stating that "outcomes do not flow from capacities in any causally determinate way." This argument can be extended to other forms of enhancement. An academic who takes a supplement to enhance their ability to concentrate beyond a norm may still produce an excellent paper based upon years of reading and developed expertise. To say that they are no longer responsible for their achievement overstates the role of the concentration enhancer and understates the years of hard work that proceeded this. There is also a range of other ways in which an academic might be seen as not solely responsible for their position or achievement. Many might have been financially privileged, and this enabled attending college; they may have had support from our families to achieve their aims. It is difficult to distinguish why some forms of support are deemed ethically permissible and other medical means impermissible.

In the context of genetic intervention on an unborn child, the notion that this in some way affects the freedom of the child as they grow up can be challenged. Many things happen to children and indeed adults that are beyond control and affect the direction in which lives go, for example, enrolment in certain schools or meeting certain people. It is a myth to suggest that our future is entirely open and difficult to distinguish between restrictions that are commonplace and those that appear to be caused by genetic enhancements. Genetic enhancements may even be seen as a way of supporting or enhancing autonomy, rather than as restricting freedom, as they may open more doors than they close.

Sandel (2007) in his *Case Against Perfection* appears to accept criticisms of the claim that enhancements are wrong because they threaten human agency. For Sandel it is not so much that such enhancements threaten agency, but that they reflect a form of hyper-agency that makes them ethically problematic. Sandel's (2007) concern is that seeking to enhance our capabilities beyond our normal human limitations reflects a refusal to accept the giftedness of human nature. It is not that such enhancements make us less free, but that they reflect a desire for us to be responsible for every aspect of our lives, a striving for mastery that Sandel considers problematic.

Here Sandel can question on these grounds a range of interventions other than enhancements, for example, sex selection of children. But for the purposes of this paper, this questioning of the drive for mastery encompasses a range of enhancements, some of which have already been discussed, for example, athletes' use of performance-enhancing drugs or enhancers of mood, concentration, and memory. Such methods reflect a refusal on the part of the athlete or striving academic to accept the unbidden, to accept that part of the way they are is beyond their control, and that flourishing as human beings requires working within such limitations, rather than seeking to master them.

Sandel also extends to consider other ways in which such enhancements might be thought of as threatening human flourishing. In this aspect of his argument, he expresses concern that a failure to recognize the giftedness of talents and traits will reduce human sympathies. He refers to how an increasing use of enhancements

might see an erosion of humility, as human beings begin to consider ourselves as effectively self-made. That Sandel is concerned with this illustrates the significant differences between the nature of his argument and Kass' concern that such enhancements will essentially erode human agency. Here Sandel's concern is that human beings will consider themselves as more responsible for their achievements, because of the choices they have made to enhance in these ways. This might also lead, Sandel suggests, to criticism of those who choose not to enhance, for example, athletes being criticized for "playing naked" and not maximizing their attributes via medical means. Sandel is also concerned that the use of such enhancements will lead to a lack of solidarity. As the individuals perceive themselves as wholly in control of their destiny, they might be less sympathetic to those who appear to be struggling in some way. The claim is that individuals would be less likely to accept that someone has just been unlucky and more likely to reflect upon the ways in which an individual could have avoided their fate.

Sandel's criticism of the desire for mastery would not extend presumably to common medical interventions that seek to restore, not enhance, human functioning. So this challenge to certain methods, or methods used in certain contexts, requires the distinction between therapy and enhancement to be upheld. Certain criticisms of Sandel's line can involve picking at this distinction. Burnout is a good example. After a long career, an individual may find that they can no longer function in work at previous levels. They feel apathetic and struggle to engage in work. Their reduction in function may be in line with another condition with an easily verifiable biological origin. Can the use of enhancement technologies to help restore the individual to their previous levels of functioning be justified; were such means available? One response might be to say no. Unless the individual can demonstrate a disease, in Boorse's terms, the intervention would not constitute a therapy. This seems unnecessarily restrictive though. Burnout, along with a number of other conditions of which we are only learning about, might constitute a reduction in function and inability to realize vital goals (Nordenfelt 2007) comparable to that of other recognized diseases. The distinction between therapy and enhancement begins to look arbitrary on these grounds, if the ultimate aim is to relieve suffering. Another feasible response to the burnout worker is to say that the enhancement is permissible, and indeed might appropriately be labeled a therapy, if function is not enhanced to a level beyond the individual's normal level; however, that is decided upon. This begins to get tricky though; is it possible to distinguish between this and medicating for other deviations from normal function that are less drastic, for example, tiredness? In Sandel's terms this seems to mark a drift from an acceptance of the limitations of human nature to striving to master all that limits us.

A more straightforward criticism concerns the perfectionist nature of Sandel's arguments. Sandel appears to be supporting a particular notion of the good life, one in which central are values such as humility, and flourishing is confined to working within human limitations rather than seeking to extend beyond them. While it does not necessarily follow from Sandel's arguments that such enhancements should not be legal (the ethics of such methods may be questioned but overall it still be thought of as better to permit them), those with a different conception of the good life may

well have a different view as to the ethical justification of enhancements or otherwise. A common move from those bioliberals supporting such enhancements is to suggest that governments ought to remain neutral in considering how conceptions of the good life figure in decisions about public policy (Harris 1998; Roduit et al. 2013). Finally there are some challenges to Sandel's reference to solidarity being damaged by enhancement methods. Attitudes toward others need not be connected to whether or not they are considered responsible for their fate. Kamm (2005, p. 13) states that "In many cases arguments for the duty to aid others seem to have more to do with respect and concern for the value of other persons than with whether they have or have not gotten themselves into whatever situation they are in."

Arguments in Favor of Enhancement

Liberal arguments in favor of permitting enhancement technologies often draw upon some of the points made above. Interestingly, some proponents of this school of thought have argued that an obligation to enhance exists, should this maximize overall well-being, much in the way that there is an obligation to treat and prevent disease (Savulescu 2007). Some arguments extend to criticize the very notion of human flourishing that Sandel considers threatened. Savulescu et al. (2004), in advocating for permitting doping in sporting competition, consider these methods not to be counter the spirit of sport but to reflect it. "Performance enhancement is not against the spirit of sport; it is the spirit of sport. To choose to be better is to be human" (2004, p. 670). One could imagine this line extended beyond sport, to suggest that a truly flourishing life is one that extends beyond current limitations, not one that accepts them.

Savulescu (2007) considers a range of arguments that support enhancements as ethically justifiable. First he suggests that we would consider with opprobrium hypothetical parents who are neglectful and lazy and do not offer their children the opportunity to improve themselves. The start point here is to ask why might such attempts to enhance be deemed as ethically problematic? Harris (1998) describes a hypothetical school in which they have discovered a way of educating children to ensure everyone leaves the school more physically fit and more intelligent. Harris contends that "We ought to want this" (1998, p. 171). An obvious objection might be that the method itself is of ethical significance here. Education involves the student working at things too, not just utilizing medical means, a shortcut of some kind. Savulescu (2007) is less convinced. He contends that there is no moral difference between environmental (education, diet, training) enhancements and biological ones. The point made in response to Kass' claims regarding agency is relevant here. For those ways of enhancing that seem uncontroversial, it isn't the case that we're solely responsible for our improvements; we might have had a good teacher or coach, for example. For those ways that have been argued to threaten agency, where we are no longer responsible for our achievements, the individual will still likely have continued working hard. An athlete who takes a

performance-enhancing substance does not stop training if they want to be successful.

Savulescu (2007) also has a particular stance on the value of health that threatens the distinction between therapy and enhancement. Health is of instrumental value, argues Savulescu. It is valuable only to the extent that it enables a good life. This threatens distinctions tested earlier justifying the treatment of someone with an underlying condition over another person, suffering equally with the same characteristic (e.g., shyness) but without the biological evidence for their problem. Our concern should be to enhance well-being, and if enhancements achieve that we can justify them. Thus, Savulescu suggests some restrictive criteria (the safety or the enhancement, potential harm to others, and issues of distributive justice); we ought to permit such methods. In the case of genetic selection for the unborn child, Savulescu proposes additional criteria: that the parent's choices be based on plausible conception of well-being and good life for the child and that the enhancement itself is consistent with the preservation of a range of life plans. Talking more generally, Savulescu also suggests that an ethical enhancement will ensure the individual retains significant control over and responsibility for achievements.

What of the criticism of this position? Some have opposed arguments against enhancement on the basis that they espouse a version of the good life that might not be unanimously shared. Even these more liberal arguments referred to above, however, don't seem entirely neutral between conceptions of the good. Here a picture of the good life that emphasizes the value of choosing to strive for better is on offer. In considering enhancements of an embryo or unborn child, Savulescu (2007) moves to consider the plausibility of conceptions of the good life, drifting away from an approach that remains entirely neutral between these.

Some of those in favor of enhancement rely on arguments that mark the difficulty in distinguishing between practices currently thought of as acceptable and those where there seems to be some question. As the paper has progressed, cases that test the distinction between therapy and enhancement have been addressed. Whether it is possible to mark an ethical distinction between education, training, and medical enhancement has also been discussed. That it can be difficult to distinguish between such cases, however, need not mean that we do away with attempts to mark such distinctions altogether. The contention may be that an approach such as Sandel's describes a disposition so problematic that it warrants attention, even if it is difficult to operationalize. McNamee (2008) makes a similar point in the context of doping in sports. Some might object to this on the grounds of preserving natural sport performance. While it is difficult to get clear on what natural sports performance means, such a notion still might be able to do some work in opposing the permitting of performance-enhancing drugs in sport.

A more liberal approach to permitting enhancements might also be challenged utilizing slippery slope arguments. Were the enhancements we have discussed so far to be permitted where would this eventually leave us? It might be argued that this threatens to take us away from the key aspects of our being human. Ultimately, it might be speculated that enhancements could progress to such an extent that human beings no longer existed. Such arguments, however, need to demonstrate the

moral problem with progressing down this slippery slope. It would need to demonstrate why it is important to mark and indeed oppose a drifting away from particular values. Savulescu (2009) is critical of an approach that attaches significance to the biological fact of being a human. Surely it cannot just be species preference that protects against enhancements that threaten certain capacities. It would be better suggests Savulescu (2009) to consider what we value and why, as opposed to the biological facts of species membership. Mere species preference – it's better to be human than something else – is also an argument that requires a great deal more in terms of foundation to support it.

Conclusion

This chapter has offered a short overview of some of the key arguments concerning the ethics of human enhancement. It has also problematized some key distinctions. It has neither offered an exhaustive consideration of the key arguments within the field nor offered a particular resolution to these difficult matters. Some of the key arguments opposing human enhancement explored here concern a threat to human agency or view attempts to enhance as a reflection of an attempt to master all aspects of our human nature. These arguments offer a theoretical basis for many intuitive objections, such as those that concern an objection to taking a shortcut, or not being responsible for our achievements. They also describe concerns with a refusal to accept our limitations as human beings. Problems with such arguments have also been considered, however, including asking how we distinguish human enhancement via medical means from other commonly accepted forms of enhancement. These arguments have also asked whether preventing enhancements on such grounds can be justified, particularly if we can ensure (through considering issues of safety and justice) that such enhancements might maximize well-being for the population as a whole.

Definition of Key Terms

- Enhancing concerns the improvement of our capacities.
- Therapy means restoring human function to a normal level.
- Enhancement refers to processes that facilitate exceeding our normal human capacities.

Summary Points

- Those opposing enhancement are often referring to a subset of enhancement practices that utilize medical means to enhance our capabilities beyond our normal level.

- These arguments utilize a distinction between therapy and enhancement, to justify therapeutic practices, while opposing enhancements.
- There are problems in employing such a distinction and in distinguishing between commonly accepted practices and the use of medical means to improve our capacities.
- Arguments against enhancement suggest that we would be less responsible for our achievements and refer to a refusal to accept our limitations as human beings.

References

Boorse C (1975) On the distinction between disease and illness. Philos Public Aff 5(1):49–68
Greely H, Sahakian B, Harris J, Kessler RC, Gazzaniga M, Campbell P, Farah MJ (2008) Towards responsible use of cognitive-enhancing drugs by the healthy. Nature 456:702–705
Habermas J (2003) The future of human nature. Polity Press, Cambridge
Harris J (1998) Clones, genes and immortality. Oxford University Press, Oxford
Herissone-Kelly P (2012) Habermas, human agency, and human genetic enhancement. Camb Q Healthc Ethics 21:200–210
Kamm FM (2005) Is there a problem with enhancement? Am J Bioeth 5(3):5–14
Kass LR (2003) Ageless bodies, happy souls: biotechnology and the pursuit of perfection. New Atlantis Spring:9–28. Available online http://www.thenewatlantis.com/publications/ageless-bodies-happy-souls. Accessed 29 Mar 2011
McNamee MJ (2008) Sports, virtues and vices. Routledge, Oxon
Moraes VY, Lenza M, Tamakoi MJ, Faloppa F, Belloti JC (2014) Platelet-rich therapies for musculoskeletal soft tissue injuries. Cochrane Database Syst Rev (4). Art No.: CD010071. doi: 10.1002/14651858.CD010071.pub3
Murphy TF (2014) Genetic modifications for personal enhancement: a defence. J Med Ethics 40:242–245
Nordenfelt L (2007) Rationality and compulsion applying action theory to psychiatry. Oxford University Press, Oxford
Roduit JAR, Baumann H, Heilinger J (2013) Human enhancement and perfection. J Med Ethics 39:647–650
Sahakian B, Morein-Zamir S (2007) Professor's little helper. Nature 450:1157–1159
Sandel MJ (2007) The case against perfection. The Belknapp Press of Harvard University Press, Cambridge, MA/London
Savulescu J (2007) Genetic interventions and the ethics of enhancement of human beings. In: Steinbock B (ed) The Oxford handbook of bioethics. Oxford University Press, Oxford, pp 516–535
Savulescu J (2009) The human prejudice and the moral status of enhanced beings: what do we owe the gods? In: Bostrom N, Savulescu J (eds) Human enhancement. Oxford University Press, Oxford, pp 211–250
Savulescu J, Foddy B, Clayton M (2004) Why we should allow performance enhancing drugs in sport. Br J Sports Med 38:666–670

How Can Aging Be Thought of as Anything Other Than a Disease? 17

Arthur Caplan

Contents

Abstract

Unless one is so concerned about the social or economic consequences of doing so, it is hard to see why aging ought not be characterized as a disease. The changes associated with aging, unlike those associated with growth and sexual maturation, are manifestly dysfunctional. The causes of the dysfunctional changes that fuel senescence are clearly rooted in the loss, collapse, or deterioration of cellular functions. These in turn are caused by wear and tear over time on cells, genetic mutations, buildup of toxic substances, and programmed cell death. While these changes are universal and beset all humans, to not describe them as disease is simply to sugarcoat the many dysfunctions of aging as "natural." The fact that they occur for almost all people at advanced ages does not make them any less dysfunctional relative to the experience of the individual in terms of "symptoms" or the overall ability of the person beset by these changes to flourish and survive. Aging is a disease. The only interesting question is whether we choose to do anything to treat it.

A. Caplan (✉)
NYU Langone Medical Center and School of Medicine, New York, NY, USA
e-mail: arthur.caplan@nyumc.org

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_10

Introduction

Why shouldn't we treat aging as a disease? Scientists are beginning to do exactly that as the mechanisms and causes of aging become clearer and begin to appear open to manipulation and alteration. Manipulability is one characteristic that moves human traits from being accepted as ordinary or normal or natural to being classified as diseases (Caplan 2014). Why do we age? Why is it that humans, like all other complex biological life forms, undergo a gradual deterioration of our major system functions over time? The phenotypic changes – wrinkling, aches and pains, loss of hair, brittleness of bones, diminished reflexes, diverticulitis, cognitive decline, incontinence, loss of muscular strength (sarcopenia), hardening of the arteries, loss of visual acuity, skin aging, etc. – are the stuff of aging. This collapse of function over time is known as senescence. We know it to be a process independent of whether one is afflicted by various diseases. It is also clearly an independent process distinct from other diseases. Humans do not age at the same rate. Decline in function, as athletes know, appears in humans in their 30s. Some humans age at extraordinary rates. This is evident in the changes exhibited by persons suffering from progeroid syndromes such as Werner syndrome or Hutchinson-Gilford progeria syndromes, in which persons undergo accelerated aging due to genetic anomalies. Progeria is considered a disease, but when the same changes happen to an individual 80 years older, they are considered normal and unworthy of medical attention.

Before looking at the argument for why aging is best regarded as a disease, it is necessary to rebut a very bad argument against so viewing it. That argument holds that it is best not to label aging as a disease because it is immoral to interfere with aging, as the consequences of doing so will be terrible.

The Price of Calling Aging a Disease Should Stop Any Effort to Do So

Many fear that if aging is classified as a disease, we will inevitably set out to cure it. And the project to cure, stop, or reverse aging would be such a long shot, so costly, so socially disruptive, so unjust to the poor, so awful for the environment, so self-indulgent, and so inimical to the enjoyment of life that it cannot be allowed to happen. Many commentators including Daniel Callahan, Leon Kass, Francis Fukuyama, Gilbert Meilaender, Colin Blakemore (Callahan 2013; Kass 2002; Fukuyama 2002; Meilaender 2002; Blakemore 2012), and others (Pijnenburg and Leget 2007) hold to some version of this view.

But those who worry about the personal, social, and economic consequences of attempts at life extension or stopping senescence in its tracks are making assumptions that are not persuasive. They put a normative horse – curing aging will lead to awful consequences – ahead of the conceptual cart: whether aging in and of itself is reasonably viewed as a disease.

Those who see Armageddon behind any effort to triumph over aging must demonstrate that humans are not clever or flexible enough to learn how to cope with persons having a lot longer lives. One way to address this contention is to ask whether humankind has adjusted to life extension in the past. If one compares life for, say, the ancient Chinese, Hebrews, Greeks, and Romans with life for today's Chinese, Israelis, Greeks, and Italians, it would seem that more people living longer life spans have not obviously brought more misery in its wake for either those groups or the world overall (Scheidel 2001). Nor would many contemporary members of these groups volunteer to swap out their longer lives for the shorter ones of their ancestors. It is hard to credit the worry that it is morally fraught to even question whether aging is a disease given that it does not square with the corollary that an average life span of 25–35 years 2,500 years ago must be obviously preferable to the 75–80 years enjoyed today in the developed world. This is so even if many spend their final "extended" years frail, demented, or debilitated. And it would be hard to argue that, despite such very real problems as overpopulation and environmental damage, the quality of life for the average person has decreased so much from the time of our ancient shorter-lived forebears that we plainly live less happily today. Few, upon serious reflection, would trade their longer life spans for the much shorter lives lived by their ancestors thousands of years ago.

Callahan and others are right to wonder about the social and economic consequences of pursuing longer lives. But the empirical evidence does not support the belief that trying to live longer must necessarily bankrupt society or lead to lives full of pain and misery. We may need policies to ensure that a fair proportion of resources is devoted to the young, that seniority on the job does not become stasis in the workplace, and that we do not use medical technology aggressively once longer life has become a burden for the community or simply too painful for most to endure. We may also need to rethink career paths, the funding of social welfare programs, and even the definition of marriage and extended family if we live a lot longer lives. But there is no convincing empirical evidence, just hypothetical hand-wringing, to suppose that longer life cannot be made enjoyable, productive, just, endurable, and meaningful.

There's an even more compelling reason not to take seriously the view that the price to be paid for conceding that aging is a disease is too high: the obvious point that the issue of what the price we ought to pay follows from and does not precede a decision about what aging is. Aging happens. We need to describe the process. We need to understand why it happens. We need to explain the processes that drive what we see as the changes, almost all physically negative, that are associated with aging. That enterprise will allow us to seek consensus on what aging is – a disease, a non-disease, or something else. But the debate over what aging is comes prior to a decision as to what anyone ought to try to do about it. If we can agree that it is a disease, then certain barriers to interfering with it, neglecting it, or accepting it fall away. There is still plenty of normative room left to debate what if anything ought be done about aging even if we recognize it as a disease.

Is Aging Natural and Does It Matter?

As I have argued previously (Caplan 2005), the perception of biological events or processes as "natural" or "unnatural" is often decisive in determining whether physicians or the public treat states or processes as diseases. One need only think of the controversies that surround the biological "naturalness" of homosexuality, some forms of mental illness, osteoporosis (WHO 1994), menopause, or impotence to see that this is so (Socarides 1970; Illich 1974; Goldberg 1975; Heiss, et al. 2008; Caplan 2014).

Why do we think of aging as a natural process? The reason that comes immediately to mind is that aging is a universal process. It occurs with a statistical frequency of 100 %. The obvious question then is whether commonality, familiarity, and inevitability are sufficient conditions for labeling certain biological states as "natural" and therefore outside the realm of disease (Hausman 1975).

Universality and inevitability, the hallmarks of naturalness, do not rule out labeling some biological processes as diseases. Coronary atherosclerosis, neoplasms, high blood pressure, sore throats, colds, tooth decay, occasional insomnia, gingivitis, diverticulitis, and depression are all universal in their distribution among humans. All seem to be inevitable. Yet their universality does not take them out of the disease camp. The inevitability of infectious disease does not cause the physician to dismiss infections as natural occurrences of no particular medical interest even if they are characteristic of the entire membership of a group or species. Gum disease strikes us all, but dentists, floss in hand, battle on against it.

If naturalness is a poor indication of disease status, are any other criteria available that can be used to drive a wedge between aging and disease? Some argue that disease refers to states of a species that are dysfunctional. Disease is a deviation from normal functions that serve species-typical goals such as survival and reproduction. Thus, Christopher Boorse, the main proponent of this view (1975, 1977), contends,

> Normal functioning in a member of the reference class is the performance by each part of all its statistically typical functions with at least statistically typical efficiency, i.e., at efficiency levels within or above some chosen central region of their population distribution. (Boorse 1977)

Thomas Schramme maintains that, on a Boorsian account of disease, aging cannot be seen as a disease since

> [a]ging is not a disease because it is not a dysfunction. Functions are established relative to age groups, so processes that are functional earlier in life are not necessarily functional later in life. For instance, the ability of cells to grow quickly ceases to be a function because it is not a standard contribution to individual survival and reproduction in old age. On the contrary, it is statistically normal for cells to grow slower when the organism becomes older. And even if we were to accept certain mechanisms as functions during the whole adult life of human beings, typical processes in old age would still not count as dysfunctions because they are statistically normal. (Schramme 2013)

I am sympathetic to the idea that the key feature of disease is dysfunction. But I think that talk of statistical normality in defining disease is just another way to introduce the objection from naturalness.

A clear case can be made that senescence – the aging process – is dysfunctional and thus ought to be viewed as a disease. First, the aging process in itself serves no function – it is not an aspect of human species design. *Homo sapiens* is not "meant" to age. There is nothing inherently goal directed or teleological about aging. Indeed, we age at different rates, as do our bodily systems, suggesting that the idea of senescence as typical of a time period in life depends on a very fine-tuned definition of "time period."

Second, the changes associated with aging – unlike those associated with growth and sexual maturation – are manifestly dysfunctional. The causes of the dysfunctional changes that fuel senescence are clearly rooted in the loss, collapse, or deterioration of cellular functions (Lopez-Otin et al. 2013). These in turn are caused by wear and tear, genetic mutations, buildup of toxic substances, and programmed cell death. While these changes are universal and beset all humans, as Schramme notes in suggesting they be dismissed as "age-typical" (2013), this is simply another way to try to recast the manifest dysfunctions of aging as "natural." The fact that they occur for most people in their old age does not make them any less dysfunctional relative to the experience of the individual in terms of "symptoms" or the overall ability of the person beset by these changes to flourish and survive.

Aging Serves No Purpose

Worse still for the view that aging, while dysfunctional, ought not count as a disease because it is species typical is the fact that aging is an accident of evolution. Some see aging's role in evolution as purposeful – the way the old make way for the new. The German cytologist and evolutionary biologist August Weismann first advanced this interpretation of aging at the turn of the twentieth century (Weismann 1891). Weismann argued that aging benefits the population as a whole by removing the superannuated and allowing evolutionary change to occur. It is from this quasi-Aristotelian attribution of a design that aging is often thought to serve a grand function in the scheme of evolution.

The determination of the function of aging depends on why it exists (Caplan 1976, 2005). The explanation of aging as serving an evolutionary role as the great cleansing device that permits change rests on faulty evolutionary analysis. It assumes that biological processes exist to directly benefit or advance the evolutionary success of a species or population – a kind of group selection that makes little sense with respect to senescence. Evolutionary selection rarely acts on entire species or population. Selection mainly acts on individual organisms, the genes they carry, and their phenotypic traits and properties. If some traits or properties confer advantages in certain environments, it increases the likelihood that the organisms having these genes will pass them on to future generations.

Given that selective forces act on individuals and their genotypes and rarely species, it makes no sense to speak of aging as serving an evolutionary function or purpose that benefits the species. Evolutionary biologists explain the existence of aging (Williams 1966; Ghiselin 1974) by noting that almost all features, traits, or properties in individual organisms will be selected for if they confer a relative reproductive advantage on the individual, or his or her close kin. Any variation that increases reproductive fitness has a very high probability of being selected and maintained in the gene pool of a species. Selection, however, cannot foresee the possible consequences of favoring certain traits. The environment selects for those traits that give an immediate reproductive return. An increased metabolic rate, for example, may prove advantageous early in life, in that it may provide more energy for seeking mates and avoiding predators. A high rate of metabolism may also result in early deterioration of the organism due to an increased accumulation of toxic wastes in the body (Herndon et al. 2002). Natural selection can neither predict nor accommodate such delayed debilitating consequences. Nor does it care if they occur. Natural selection is only interested in short-term gain. Aging exists solely as a consequence of evolution's lack of "foresight." It is simply a by-product of selective forces that work to maximize the organism's immediate chances of reproductive success. Senescence has no function. It is simply the inadvertent subversion of organic function, later in life, that results from favoring reproductive advantage early in life.

The common belief that aging serves a function or purpose is mistaken. And, if this is so, the common belief that aging is a natural process with a purpose or goal is also mistaken. The fact that it is typical of the human species or most other living species is, without some functional underpinning, no reason to accept it or to deny its dysfunctionality.

Still, even if senescence has no point, that does not make it a disease. What it does suggest is that it has no point and it is clearly dysfunctional. Aging has plenty of undesirable symptoms, is manifest in the loss of numerous functional capabilities, and can be explained by a series of noxious changes in cellular functioning whose causes remain poorly understood but the reality of which is all too clear. Surely a visitor from another planet or a time traveler from our own who had conquered senescence would take pity on our enfeebled elders and do what they could to fix, remediate, or cure them if they could.

Conclusion

It is true that certain biological processes, such as puberty and pregnancy, have been the subject of heated debates over their standing as possible disease states. But all of these states have functions even if they are accompanied by unwanted experiences or an increased risk of death. The beliefs that it is natural and normal for only men and women to have sexual intercourse or for women to undergo menopause or for the infertile not to have children have been successfully challenged. The process of aging can be classified as typical of humans, yet that does not make it any less

dysfunctional or utterly pointless in terms of why it occurs. It is likely that as medicine moves toward a better understanding of the causes of senescence (Faragher et al. 2009), society will open the door for its reclassification as a disease and, thus, a proper subject for medical attention, concern, and control (Faragher et al. 2009; Cabreiro and Gems 2011; Fulop et al. 2010). It is true that we need not devote any resources to curing or slowing senescence. But the decision to leave aging alone should rest only on considerations of cost, justice, and the ability to intervene with the process. There should be little doubt that a collection of changes that disable us and then kill us for no reason other than evolutionary indifference constitutes a disease.

Summary Points

- The decision to call aging a disease does not mean society needs to seek to cure it.
- Decisions as to whether aging is a disease ought not hinge on the social impact of such a classification.
- There is nothing natural about aging.
- Aging serves no purpose.
- The changes associated with aging are manifestly dysfunctional.
- Persons age at different rates – the process is not universal.

References

Blakemore C (2012) Life extension: "Moral obligation" or "a disaster for humanity and the planet? http://frontierpsychiatrist.co.uk/life-extension/

Boorse C (1975) On the distinction between illness and disease. Phil Public Affairs 5:49–68

Boorse C (1977) Health as a theoretical concept. Philos Sci 46:557–561

Cabreiro F, Gems D (2011) Treating aging: progress toward dietary restriction mimetics. Biol Rep 2:76

Callahan D (2013) On dying after your time. NY Times

Caplan AL (1976) Ethics, evolution, and the milk of human kindness. Hastings Cent Rep 6:20–25

Caplan AL (2005) Death as an unnatural process. EMBO Rep 6(1):72–75

Caplan AL (2014) Ethical implications of drugs for erectile dysfunction. JAMA Virtual Mentor 16:928–931

Faragher R, Sheerin A, Ostler E (2009) Can we intervene in human ageing? Expert Rev Mol Med 11:e27. doi:10.1017/S1462399409001197

Fukuyama F (2002) Our posthuman future. Farrar, Straus & Giroux, New York

Fulop T, Larbi A, Witkowski JM, McElhaney J, Loeb M, Mitnitski A et al (2010) Aging, frailty and age-related diseases. Biogerontology 11:547–563

Ghiselin MT (1974) The economy of nature and the evolution of sex. University of California Press, Berkeley

Goldberg S (1975) What is "normal"? Logical aspects of the question of homosexual behavior. Psychiatry 38:227–243

Hausman DB (1975) What is natural? Perspect Biol Med 19:92–101

Heiss G, Wallace R, Anderson GL et al (2008) Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin. JAMA 299(9):1036–1045
Herndon LA, Schmeissner PJ, Dudaronek JM, Brown PA, Listner KM, Sakano Y, Paupard MC, Hall DH, Driscoll M (2002) Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. Nature 419:808–814
Illich I (1974) The political uses of natural death. Studies Hast Center 2:3–20
Kass LR (2002) Life, liberty and the defense of dignity. Encounter Books, San Francisco
Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. Cell 153(6):1194–1217
Meilaender G (2002) Genes as resources. Hedgehog Rev, Fall, www.virginia.edu/iasc/hedgehog.html
Pijnenburg MAM, Leget C (2007) Who wants to live forever? J Med Ethics 33(10):585–587
Scheidel W (2001) Debating roman demography. Brill, Leiden
Schramme T (2013) 'I hope that I get old before I die': ageing and the concept of disease. Theor Med Bioeth 34:171–187
Socarides CW (1970) Homosexuality and medicine. JAMA 212:1199–1202
Weismann A (1891) Essays upon heredity and kindred biological problems. Clarendon, Oxford, UK
WHO (1994) WHO assessment of fracture risk and its application to screening for postmenopausal osteoporosis. WHO Technical Report Series 843. WHO, Geneva, Switzerland, 1994
Williams GC (1966) Adaptation and natural selection. Princeton University Press, Princeton

Part II

Organisms

Human Organisms from an Evolutionary Perspective: Its Significance for Medicine

18

Mahesh Ananth

Contents

Abstract

Defenders of *evolutionary medicine* claim that medical professionals and public health officials would do well to consider the role of evolutionary biology with respect to the teaching, research, and judgments pertaining to medical theory and practice. An integral part of their argument is that the human body should be understood as a bundle of evolutionary compromises. Such an appreciation,

M. Ananth (✉)
Department of Philosophy, Indiana University South Bend, South Bend, IN, USA
e-mail: mananth@iusb.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_62

which includes a proper understanding of biological function and physiological homeostasis, would provide a crucial perspective regarding the understanding and securing of human health needs currently lacking in the medical arena.

Introduction

Geneticist and evolutionary biologist Theodosius Dobzhansky (1973, p. 125) famously decreed, "Nothing in biology makes sense except in the light of evolution." This near-oracular pronouncement has echoed throughout the discipline of biology, even extending to other non-biology-related subject matters (Mayr 2000). Still, the full effect of evolutionary thinking with respect to medicine is waiting to be felt. One straightforward reason for this lag is a time horizon difference. On the one hand, medical practitioners are concerned with caring for the human body based on an individual/personal history time frame (ontogeny), and public health officials focus their efforts around producing the social infrastructure that both mitigates the spread of disease and secures health benefits that pertain to individuals, populations, and communities in the short run (see APHA 2016, url, and Hartig et al. 2014). On the other hand, evolutionary biologists examine species/bodies/traits under the guise that their current existence owe to advantages gained via a more or less geologic timescale perspective (phylogeny) (Ewald 1994, p. 8). Still, granting that humans are the product of a long history of evolutionary processes, it is reasonable to explore the significance of humans as one among the animals for medicine (i.e., evolutionary/Darwinian medicine) in a way that can *supplement* the perspective of medical practitioners and public health officials. Indeed, there is no hubris here on the part of the defenders of evolutionary medicine. Rather, proponents of evolutionary medicine acknowledge that the activities and achievements of medicine reveal a history of stellar insights, which have drawn upon solid scientific practices and reflection. They also counsel that one more accoutrement to the already impressive "medical toolkit" could very well engender unanticipated boons (World Health Organization, 2007, url).

This chapter offers an analysis of some of the key philosophical and biological issues pertaining to the role that evolutionary biology can play within the theoretical and practical aspects of medicine. To this end, section "What is Evolutionary Medicine?" will address the nature and scope of evolutionary medicine. Section "The Human Body as an Evolutionary Entity: The Compromised Physical Self" will explicate the idea of the body as a bundle of evolutionary compromises and the relevance of such a notion to medicine with special attention given to four prominent concepts of biological function and two components of homeostasis. Section "The Evolved Bundled-Homeostatic Body: Influence on the Practice of Medicine and Public Health" will draw upon the biological function and homeostatic aspects of the human body and propose how such a perspective can help make sense of the mechanism of fever, sickle-cell anemia, and allergic responses.

What Is Evolutionary Medicine?

Evolutionary medicine [hereafter **EM**] is concerned with both biological and philosophical issues pertaining to the field of medicine and public health (WHO, url). Although it may seem unambiguously obvious, on the biological side, medicine is inexorably linked to biology. In general, this is primarily made manifest in the physical health judgments [hereafter **PHJs**] tendered by healthcare professionals. Inescapably, then, concepts that are germane to biology will be correspondingly relevant to **PHJs**. On the assumption that both medicine and public health are concerned with understanding, establishing, and improving human health, and simultaneously mitigating and controlling the effects of disease states, **EM** can contribute insights that may be valuable to the practices of medicine and public health on the health, disease, and public policy fronts; this includes evolutionary explanations and medical prescriptions pertaining to "defenses, infection, novel environments, genes, design compromises, and evolutionary legacies" (Nesse and Williams 1994, 1997). It is this evolutionary component that is not patently discernible, argue defenders of **EM**, in the biologically based **PHJs** made by medical practitioners and reveals why the link between biology and medicine requires further investigation. On the philosophical side, the issues range over establishing a proper understanding of the human body. This is accomplished by giving special attention to the biological concepts of function, homeostasis, and natural selection. Ultimately, the aim of **EM** is to provide both conceptual tools and practical insights for medical and public health professionals in their quest for improving the human condition as it relates to health and disease. Although this **EM** perspective is beginning to gain traction in the scholarly literature (Sprouffke et al. 2012; Alexander et al. 2014), **EM** advocates have been strongly urging that medical practitioners and public health officials would do well to consider more diligently evolutionary reflections as they pertain to both **PHJs** and the nature and modification of health and disease states in human populations.

The Human Body as an Evolutionary Entity: The Compromised Physical Self

Since maintaining and improving physical human health are integral parts of the primary goals of medicine and public health, it is crucial to have an accurate understanding of the human body from an evolutionary perspective and how this relates to the concept of health. What, then, is the human body from a Darwinian perspective? *The answer is that the human body is* (*1*) *a bundle of adaptive compromises*, (*2*) *which are evolved functions that* (*3*) *have an overarching goal of concomitantly securing both internal and organism-to-environment homeostasis* (*4*) *for the sake of survival and reproductive success* (Ananth 2008 and see Goldstein

1963, for physical holism discussion). What this means is that many of the features of the human body are the product of a long geologic history of give-and-take as our species adapted to a range of environmental perturbations. From this vantage point, not only were certain features "selected for" survival and reproductive success with respect to local environments, but these very features *qua* whole organism were modifying the local environments creating a constant dynamic playing field (Odling-Smee et al. 2003). Thus, not only were traits *functioning* to help secure survival and reproductive success in local environments, but they were also doing so while maintaining a kind of dual internal compromise *equilibrium* and an organism-environment-compromise equilibrium in the presence of varying degrees of environmental dynamism.

What are the implications of the human body as an evolved internal-external-equilibrium "bundle of compromises"? Nesse and Williams (1994, p. 19) succinctly provide the following answer:

> Like any engineer, evolution must constantly compromise. . .If something works well enough that its deficiencies do not constitute a selective force, there is no way natural selection can improve it. Thus, while every part of the body has some reserve capacity to deal with occasionally encountered extreme circumstances, every part is also vulnerable when its reserve capacity is exceeded. There is nothing in the body that never goes wrong.

What can be gleaned from Nesse and Williams is that an evolutionary perspective on the body includes acknowledging that less than optimal features are an integral part of understanding the human body as a bundle of compromises. The reason for this is that *adaptations are adaptations to local environments*. If a feature functions suboptimally, but still performs well enough in the light of how well other features are functioning in a local environment, then its suboptimal functioning does not necessarily represent a selective disadvantage. Rather, it may very well be the case that no group occupies the fitness landscape wherein optimal fitness is possible (Gilchrist and Kingsolver 2001).

Relevance to Medical Teaching and Practice

The importance of anatomy to medicine cannot be overemphasized (Older 2004 and Papa and Vaccarezza 2013). Over the past 20 years, however, the teaching and understanding of anatomy have come under fire. On the academic side, the somewhat slow transition from lecture and cadaver dissection training/learning to technology-based training/learning has given this discipline a somewhat dull image, resulting in subpar dissemination of knowledge at both undergraduate and medical school venues. On the economic side, labs and instructors are costly resulting in a lack of infrastructure upkeep (Turney 2007 and Bergman et al. 2014). Despite the many in-house battles regarding pedagogy and funding, there appears to be some consensus that a balance must be achieved between traditional learning styles and technology-driven approaches in the face of scarce

resources. What isn't mentioned in much of this review literature is the relevance of evolutionary thinking with respect to the teaching of anatomy and medicine in general. As it currently stands, anatomy texts tend to focus on the systems approach; that is, students are taught about the detailed workings of various systems in the body (e.g., nervous system, digestive system, endocrine system, etc.) and their *mechanical part interactions* with very little pedagogy regarding how these systems are really bundles of *evolved interacting compromises*. This supposedly stark oversight in medicine has not been lost on defenders of **EM**:

> Evolutionary biology is so firmly integrated with the rest of biology that it is not possible to mark a boundary between them. But modern medicine has been a peninsula...From secondary school through medical school, the fundamental relevance of evolution to all of human life often has been ignored or even suppressed. (Ewald 1994, p. 7)

Both the suppression of and indifference to the role of evolutionary thinking in medicine may very well be the product of religious and political worries, but its almost-glaring omission from the field needs serious reconsideration in the eyes of most proponents of **EM**. To illustrate, as a way of mitigating the putative "banausic" effect of anatomy on medical students, Turney (2007) has insisted that the natural wonder associated with the workings of the human body should be given its due. Yet, much like others analyzing this recent decline in the interest in anatomy, Turney does not suggest that a "Darwinian twist" might be one of the missing elements needed to ignite a much looked-for blend of enthusiasm and creative curiosity. Defenders of **EM** believe this can be achieved by showing the practical implications that **EM** can have on **PHJs** and *a fortiori* the practices of everyday medicine. As Shanahan (1999, lecture) has suggested, "At present Darwinian medicine exists mainly at the level of theory, but the insights of Darwinian medicine could work a profound transformation in *practice* of medicine in the next century." Nesse and Williams (1994, p. xi) also offer their respectful recommendation to medical mavens: "We are not urging an alternative to modern medical practice but rather an additional perspective from a well-established body of scientific knowledge that has been largely neglected by the medical profession." If this transformation and/or added perspective suggested by advocates of **EM** happens sooner rather than later, then understanding the human body as a bundle of evolutionary compromises is a grand place to start from the **EM** perspective. Correspondingly, the concepts of biological function and homeostasis, which are discussed in the next two sections, also require explication in order to paint an accurate picture of the human body *qua* evolving bundle of compromises.

Biological Function

Systemic Functionalism

Beyond an evolutionary bundle of compromises, the previous section introduced both "biological function" and "bodily homeostasis" as integral parts of the concept

of the human body as an evolved entity. This section provides a discussion of the nature of biological function, while the next section addresses the nature of bodily homeostasis. In the philosophical literature, there is a robust discussion, not only on the nature of biological function (Ananth 2017; Nissan 1997 and Wouters 2003a and 2003b) but also on the variety of evolutionary accounts of function with respect to health and disease (see Ananth 2008, for citations). The result is that four accounts of biological function have emerged as front runners: (1) Systemic Functionalism, (2) Etiological Evolutionary Functionalism, (3) Propensity Evolutionary Functionalism, and (4) Mixed Evolutionary Functionalism. Although debates persist on which account most plausibly reconciles philosophical worries and biological realities (Ariew et al. 2002), the "systems" account provides a point of departure. According to this view, if a component **X** causally contributes to system **S**'**s** performance of **Z**, then **X**'**s** contribution constitutes a function of **S**'**s** performance of **Z** (Cummins and Roth 2010 and Boorse 2002). The heart, for example, is a feature of the circulatory system. This system, which includes the lungs and blood vessels, works to keep about 5 l of blood continuously moving through the body in a constant exchange of waste-filled blood for oxygen-rich blood. So, the heart has the function of pumping and exchanging good and bad blood through the body because it causally contributes to the circulatory system's overall task of moving blood throughout the body. We can formally make sense of Systemic Functionalism as follows:

Systemic Functionalism: A feature **X** has a function in system **S** if and only if activity **Y** of **X** causally contributes to **S**'**s** overall capacity/performance of **Z**.

In terms of health and disease, Boorse argues that Systemic Functionalism [hereafter **SF**] should be favored. His rationale is that physical health and disease are restricted to the subfield of pathology, which is concerned with *the proper function of parts of organisms within specific subsystems of the body* (Boorse 1987; my italics). From this perspective, the thyroid is healthy when it actualizes the function of producing thyroxin in the endocrine system because it causally contributes to the endocrine system's overall capacity/performance of metabolic regulation.

It is worth noting that, much like the systems approach taken in terms of understanding anatomy, **SF** is the version of biological function endorsed by much of the medical/clinical community (see Tyreman 2001). What this means is that emphasis is given to understanding how the parts of organisms function within particular subsystems with less attention given to the subtle evolutionary give-and-take with respect to these systems and the body as a whole. For so long as the thyroid produces the appropriate amount of thyroxin in terms of its contribution to metabolic regulation, the medical community would endorse such production as indicative of a healthy functioning thyroid (Tyreman 2001, adroitly elaborates on this point in his attempt to distinguish the varieties of function endorsed within osteopathy as distinct from the part-functionalism predominantly embraced by the medical community).

Although there are a number of objections to **SF** (Ananth 2017; Perlman 2010), it is worth noting that it says very little about biological features *as evolved features* – the same omission in most **PHJs**. Given this omission, the pressing objection from an **EM** perspective is that **SF** is unable to distinguish genuine functions from accidental side effects. For instance, the heart's thumping sound is a capacity of the heart, but the system to which the heart contributes, namely, the circulatory system, performs the task or has the capacity for blood circulation. The heart-thumping sound does not contribute to this overall performance, so this capacity should not be considered a function. Yet, the systemic functionalists *appear* to be forced into accepting *the thumping sound* as a genuine function of the heart.

Additionally, and more pressing to the discussion at hand, the part-functionalism that underwrites **SF** obscures a proper understanding of the body as an integrated evolved entity. This includes understanding the human body as an evolved ecologically oriented entity, something not well integrated into current **PHJs**. Depew (1998, p. 31) expresses this point when he reminds Darwinians that biological systems *qua* bodies are "bounded, informed, autocatalytic dissipative systems [and] are by definition parts of ecological communities, and that the information which they store and use is subject to dynamics that are inseparably both competitive and cooperative... For natural selection can play the deep, essential, and above all creative roles suggested by their theories only when organisms are treated ecologically." The upshot is that the field of medicine should be cautious about incorporating solely **SF** into its understanding of the human body.

Evolutionary Functionalism

In an attempt to distinguish genuine functions from accidental side effects and be more sensitive to the sense of body *qua* ecological entity suggested by Depew above, Darwinians have insisted that an evolutionary selectionist account must be incorporated into the concept of function. Drawing from the work of Mayr (1974) and Wright (1976), *Evolutionary Functionalism* (hereafter **EF**) developed as follows:

Evolutionary Functionalism: A feature **X** has a function in an organism **O** if and only if activity **Y** of **X** produces effect **E** because **Y** and **E** were naturally selected (over some other causes and effects) to bring about the goals **G** of survival and reproductive success of **O**.

For example, the liver has a function of blood detoxification in mammals, because the activity of converting ammonia into the less toxic compound urea produces the effect of detoxified blood. Moreover, this activity and effect were naturally selected for the sake of survival and reproduction. It is this selectionist account that allows for distinguishing genuine functions from side effects; that is, the thumping sound of the heart is a side effect of the naturally selected effect of blood pumping, or the far less noticeable sounds of detoxified blood moving out of the liver are side effects of the naturally selected effect of blood detoxification.

Reply to Evolutionary Functionalism

At first glance, **EF** as an alternative or supplement to **SF** seems plausible because it is able to distinguish genuine functions from side effects and appears to be sensitive to the body as an evolved ecologically based unit, but it has had its detractors within the Darwinian camp. Specifically, there are more subtleties to **EF** than are captured in the above definition. For instance, how to understand "selection" in the above definition is not made clear. Are features currently selected? Were they only selected in the past? Will they be selected in the future? Is it really possible to make sense of health and disease by way of **EF**? These questions suggest that greater specificity is required in order to put forth an evolutionary concept of function that is a worthy alternative to **SF**. Specifically, three philosophical approaches to understanding **EF** – as they pertain to health and disease – will be distinguished: (i) the backward-looking *Etiological Evolutionary Functionalism* and (ii) the forward-looking *Propensity Evolutionary Functionalism*. Additionally, this section will argue that neither of these versions of biological function is adequate, but that a worthy alternative, (iii) *Mixed Evolutionary Functionalism*, is available.

Etiological Evolutionary Functionalism

According to Etiological Evolutionary Functionalism (hereafter **EEF**], a feature performs a function in a system of an organism if and only if (1) the feature's presence in a system was useful with respect to the organism's reproductive success in previous generations and that (2) it is the result of evolutionary selection forces. From this perspective, a trait is functional because its presence is due to its ability to produce, in a self-sustaining fashion, a beneficial difference that related traits were unable to produce. The result is that, in contrast to side effects or lucky features, what counts as a functional trait is one that can recycle itself as a result of *delivering a reproductive advantage to the organisms* of which it is a part (Millikan 1993, p. 38 and Hardcastle 1999, p. 32; my italics). The formal definition of **EEF** looks like this:

Etiological Evolutionary Functionalism: A feature **X** currently has a function in an organism **O** if and only if activity **Y** of **X** produces effect **E** because **Y** and **E** were naturally selected (over some other causes and effects) to bring about the goals **G** of survival and reproductive success of **O**.

With this general description in place, a specific example is in order. In humans, iduronate sulfatase is the lysosomal enzyme that is designed to breakdown mucopolysaccharides (a gel-like substance found in the body of cells). For example, connective tissue outside of cells needs to be replaced on occasion. When this replacement occurs, iduronate sulfatase metabolizes the old connective tissue. On occasion, in males only, a genetic error occurs such that not enough iduronate sulfatase is present to break down the mucopolysaccharides that build up from the remaining old connective tissue. The result of the buildup of mucopolysaccharides (in lysosome cells) is the following multisystem collapse: hyperactivity, aggressive

behavior, coarse facial features, enlargement of internal organs, dwarfism, stiffening of joints, progressive deafness, and severe mental retardation. This genetic disease is known as Hunter syndrome (Ananth 2008).

From the **EEF** perspective, the function of these iduronate sulfatase enzymes is to metabolize mucopolysaccharides, because, ancestrally, there was selection pressure in favor of them doing just this to ensure survival and reproductive success. Notice that this understanding of iduronate sulfatase enzymes not only makes sense of the specific function of metabolizing mucopolysaccharides but also gives consideration to the organism as a whole in terms of survival and reproductive success. Additionally, how well iduronate sulfatase enzymes function is a product of how these enzymes function in relation to other neighboring organelles in the cell; its evolved functions are understood as adaptive compromises with respect to other related organelle functions. In severe cases, human males who either lack iduronate sulfatase enzymes or do not produce enough of them have multisystem dysfunction, rendering them physically unfit and thus unhealthy (see Bechtel 1985 for further discussion). Again, such fitness assessments can complement **PHJs** when thinking about the body as an evolving ecologically conceived bundle of compromises.

Replies to Etiological Evolutionary Functionalism

A quick glance at **EEF** might move one to consider this version of **EF** credible. For, as part of its content, it appears to include both the necessary and sufficient conditions for what it means for **X** to have a function and is respectful of the body as an evolving entity. As Bigelow and Pargetter affirm, "The big plus for the etiological theory is that it makes biological functions genuinely explanatory, and explanatory in a way most comfortable with the modern biological sciences" (Bigelow and Pargetter 1987, p. 187). Moreover, it provides a general framework for distinguishing genuine functions from mere accidents, because a feature of a biological system is a function if and only if it is the product of natural selection.

By relying solely on evolution, however, **EEF** must ascribe functions to those features that no longer have functions. As Nissen correctly remarks, "Since history is forever, if functions are determined by their history, functions are forever. New functions can be added, but old ones never die. This means that vestigial organs still have their original functions" (Nissen 1997, p. 185). Thus, **EEF** is triumphant in distinguishing genuine functions from accidents because of its reliance on evolutionary causal history, but such an achievement proves to be a somewhat pyrrhic victory in that vestigial organs, like the appendix, human tailbone, and human male nipples, do not lose their evolved functions – even if those functions can no longer be actualized.

There is one additional concern that needs to be noted. Some argue that **EEF** is committed to defining functions in terms of actual reproductive success. As it is described in this section, it is ambiguous whether or not such a commitment is entailed. Still, Bigelow and Pargetter note that some scholars argue that **EEF** assumes that

> fitness can be judged only retrospectively: that it is only after we have seen which creatures survived that we can judge which were the fittest; moreover, it assumes that the fact that certain creatures have survived, whereas others did not, is what constitutes their being fit. (Bigelow and Pargetter 1987, p. 190)

Clearly, this may be a serious problem with versions of **EEF** that are committed to the view that **Y** is a function of **X** if and only if **Y** contributes to *the actual* reproductive success of **O**. The concern is that it is absurd to confer a function on an organism or take away a function from an organism based on lucky or unlucky anomalous environmental perturbations, which are not part of the normal environment in which the organism has evolved.

It is this sort of worry that leads Bigelow and Pargetter to claim that the "etiological theory is mistaken in defining functions purely retrospectively, in terms of actual survival" (Bigelow and Pargetter 1987, p. 191). So, if it is the case that **EEF** is committed to actual reproductive success, then this criticism is quite relevant and should influence **PHJs** accordingly; that is, for all of its pluses, **EEF** should not be solely employed as the sense of function anchoring the concept of biological function for **PHJs**.

Propensity Evolutionary Functionalism

The point that can be taken away from the above criticism is that it is better to say that **Y** is a function of **X** if and only if **Y** has *the capacity* its ancestors had to contribute to the reproductive success of **O**. By substituting "capacity" for "actuality," the problem of environmental anomalies disappears. This capacity view of biological function is known as the propensity interpretation. For example, the function of iduronate sulfatase enzymes is to metabolize mucopolysaccharides and not some other substance found in the body of cells, because creatures whose mucopolysaccharides are broken down by iduronate sulfatase enzymes have a greater disposition of surviving and reproducing than creatures whose mucopolysaccharides cannot be metabolized. Propensity Evolutionary Functionalism (hereafter **PEF**) can be defined as follows:

Propensity Evolutionary Functionalism: A feature **X** has a function in an organism **O** by performing activity **Y** if and only if **Y** produces effect **E** because **Y** and **E** confer and will continue to confer a propensity **P** (within a certain range of environmental pressures) to bring about the goals **G** of survival and reproductive success of **O**.

In thinking about a healthy biological system, Bechtel endorses the analysis offered by Bigelow and Pargetter as follows:

> There is [a] conceptually intermediate position that has been developed in philosophical reflections on evolutionary theory – a propensity interpretation of fitness. What the propensity interpretation of fitness does is define fitness in terms of propensity to reproduce, not reproductive success itself. This is all that is required for our purposes, *for we can now define*

> *something as functional if it increases the propensity of its bearer to reproduce . . .*(Bechtel 1985, p. 151; my italics). The approach I am exploring directs one to engage in an engineering analysis to identify how the physiological organization of the system equips it to deal with the selection forces working upon it. A healthy state of the system is one in which it makes best use of its physiological endowments in responding to selection pressures. (Bechtel 1985, p. 154)

Bechtel argues that the concept of function that should be relevant to the concept of a healthy state of a system is **PEF**. From this perspective, **PHJs** can endorse a notion of a healthy body by understanding how features contribute to the propensity or likelihood that said body will survive and reproduce.

Replies to Propensity Evolutionary Functionalism

The difficulty with **PEF** is that it takes function as ontologically prior to selection – a move that begs the question of how **PEF** is able to determine what is and is not a function. As Bigelow and Pargetter note (1987, p. 192; my italics), "On our theory, *the character already has the function*, and by bad luck it might not survive, but with luck it may survive, and it may survive *because* it has a function." Similarly, in response to the etiological account, Bechtel claims (1985, p. 150; my italics), "The correct order is to claim that *those things that are functional will evolve*, rather than to claim that those things that evolve are functional."

The above claims by Bigelow, Pargetter, and Bechtel are problematic. The obvious problem with giving priority to function over selection is that it begs the question of *why* it is the case that **X** has the function in the first place. **X** has the function to do **Y**, because **X** yields a survival-enhancing propensity on **O**. But why does **X** have the function *qua* survival-enhancing propensity that it has? Clearly, they cannot rely on propensity here, because propensity and function are one and the same once selection is no longer part of the concept of function. That is, if Bigelow, Pargetter, and Bechtel presume (as they do) that character **X** "already has the function" prior to selection, then this means that **X** already has a propensity prior to selection. If not selection, then what confers "having a propensity" that makes it the case that **X** is a function? They could respond by claiming that a propensity is a property or capacity of a trait to do **X**. Yet, this leads to a regress problem. For now it can be asked, how is it the case that a property or capacity "already" exists in a creature without introducing some sort of causal history to account for the capacity? As it stands, Bigelow, Pargetter, and Bechtel have no answer, because a capacity or property is an unexplained metaphysical element of their analyses. Natural selection, on the other hand, is a physical force or process (like gravity). Are propensities thought to be the same? This seems unlikely, because Bigelow and Pargetter have already ruled out the possibility that propensity relies on contingent natural phenomena. The upshot of this overall objection is that Bigelow, Pargetter, and Bechtel have not offered a persuasive account of what a function is. Thus, it is not at all clear that giving priority to propensity over selection is preferable.

Bigelow, Pargetter, and Bechtel might object to the above criticism by claiming that "What is a function?" is different from the question "What causes functions to be present?" For example, if Bigelow and Pargetter find a watch, they can claim to know it's a watch first and then ask who made it. Obviously, they know that they have to give a causal history to account for its function, but that is different from defining what a function is. That is, they could argue that they do not have to answer the second question in order to answer the first. This reply, however, is vulnerable to the next criticism.

Setting aside the ontological priority worry, it is not clear that the **PEF** can distinguish genuine functions from mere side effects. For example, imagine that the metabolic activities of iduronate sulfatase enzymes not only break down mucopolysaccharides, but they also have the accidental benefit of improving the sense of smell. On the propensity interpretation, both the metabolic activities and the improved sense of smell would have to be considered genuine functions because the former (directly) confers a survival-enhancing propensity and the latter (accidentally) confers a survival-enhancing propensity. Peter McLaughlin voices a somewhat similar concern as follows:

> 'The function of [**X**] is **Y**' is true, not only when **X** does **Y** due to its propensity, but also when it has a strong propensity to do **Y**, but happens to do it by accident and not due to its propensity... Furthermore, if low probabilities were to count as low propensities, then it would seem that even accidents occur on account of a propensity. (McLaughlin 2001, p. 126)

The implication of the above account leads McLaughlin to conclude correctly that those who embrace the history-free propensity interpretation of fitness are "forced to attribute a function to more or less everything" (McLaughlin 2001, p. 126). Indeed, this criticism reveals why they must address the question "What causes functions to be present?" Thus, the **PEF** should be rejected by medical professionals as being an integral part of the concept of a healthy evolving body on the grounds that it not only gives ontological priority of place to function over natural selection but also because it cannot distinguish genuine functions from mere side effects.

Mixed Evolutionary Functionalism

Thus far, the general conclusion is that **EF**, **EEF**, and **PEF** are not able to emerge unscathed in route to their respective concept of biological function. **EEF** suffers from focusing on actual reproductive success and not being able to allow an entity to lose its function, whereas **PEF** cannot distinguish genuine functions from fortuitous accidents and it cannot justify giving priority to function over selection.

The more defensible alternative combines these three accounts. **PEF** has the advantage that it is not committed to the actual reproductive success of a trait, but only to the disposition of such a trait to enhance reproductive success. The advantage of both **EF** and **EEF** is that they are able to distinguish genuine functions from mere side effects. Moreover, **EEF** gives priority to selection over propensity in order to determine what is a genuine function because it takes causal history into account. In the spirit of unification, the appropriate account of function will give priority to natural selection, but claim that selection ranges over propensities to survive and

reproduce. In full, a feature of an organism is a function if and only if it confers a propensity to produce a specific set of activities/effects and corresponding specific benefits *and* to enhance the goal of survival and reproductive success on an organism *and* that such a propensity set is established through natural selection. The formal characterization of this dual-functional mixed account (hereafter **MEF**) is as follows:

Mixed Evolutionary Functionalism: A kind of **EF** explanation maintains that a feature **X** has a function in an organism **O** by performing an activity **Y** if and only if **Y** produces effect **E**, and both **Y** and **E** confer a survival-enhancing propensity **P** on **O** (within a certain range of environmental pressures) and will continue to confer **P** on **O** (so long as a certain range of environmental pressures is present). And, moreover, **P** is currently present, because, ancestrally, there was natural selection in favor of retaining **P** to bring about the goals **G** of survival and reproduction.

A return to the enzyme example will help to explain the above account. Recall that iduronate sulfatase is the lysosomal enzyme that is designed to break down mucopolysaccharides. With respect to the mixed account above, iduronate sulfatase is a function of the human organism, because of its ability to produce the specific effect/activity of metabolizing mucopolysaccharides; and this effect correspondingly confers a survival-enhancing propensity on the human organism. Importantly, the reason why it currently confers such a propensity is because, ancestrally, there was natural selection in favor of retaining such a propensity (over a range of environmental pressures) for the sake of survival and reproduction (Ananth 2008).

Benefits to Medicine

Early in this chapter, it was stressed that the body as a bundle of evolutionary compromises captures the principal sense of how the body should be understood within the framework of **EM**. Two key aspects of understanding the body in this way is that "the bundle" must be understood in terms of proper functioning and homeostasis. Although medicine currently presumes almost exclusively something like **SF**, this analysis has suggested that **EF** should supplement or complement it. Specifically, **MEF** is the version of **EF** that seems most reasonable. What this means is that the evolutionary compromises of the body should be understood as compromises that manifest their activities as suggested by **MEF**. So, if it is correct that **SF** is an integral part of how medicine understands the human body, then **MEF** should correspondingly be included in this understanding.

Evolutionary Dual Homeostasis

Intercellular Homeostasis

If it is acknowledged that the body is best understood as an evolutionary functioning bundle of compromises, then it should be no surprise that there is an intricate

balancing act that has unfolded within the body (Bernard 1957/1865 and Cannon 1963) and between the body and its environment (Goldstein 1963 and Depew 1998). Moving beyond the efforts of Bernard and Cannon (Cooper 2008), the research regarding bodily homeostasis continues to this day on both the biomedical front and the philosophical head. For instance, on the biology side, Keesey and Powley (2008, p. 445; my italics) have stressed that "though the early workers failed to include *body energy* among the conditions of the body subject to homeostatic regulation, a sound foundation, based upon the work of the past several decades, appears now to be in place for its inclusion." Additionally, on the philosophical flank, as part of his defense of a naturalistic concept of health, Boorse (1997, pp. 78–79; bracketed addition mine) acknowledges that his account assumes homeostasis as a necessary condition. He elaborates as follows:

> Though I did not stress the dynamism [i.e., the process of homeostasis] of normal physiology in presenting [my naturalistic concept of health], I always assumed it . . . Obviously, no fact is more pervasive than what is often called 'dynamic equilibrium' of normal physiology: the normal functional variation within organisms acting and reacting to their environment. The normal level of almost all part-functions varies with what an organism is doing, what other part-functions are being performed, and the environment . . . A common pattern is that environmental stress evokes short-term compensatory functions that maintain homeostasis up to a point, but beyond that point the coping mechanisms break down and a discontinuity, a discrete state of illness, results.

It is now appropriate to sketch the framework of a homeostatic system. The general idea is that an individual organism's internal environment is in homeostasis if it responds appropriately to various stimuli from the internal and external environment. Specifically, on the physiological level, the following internal physical states must be kept stable for the intercellular fluid to be in homeostasis (Roberts 1986, p. 201):

(i) The chemical composition of the intercellular fluid (e.g., constant level of glucose in the bloodstream)
(ii) The osmotic pressure of the intercellular fluid (determined by the relative amounts of water and solutes)
(iii) The level of carbon dioxide in the intercellular fluid
(iv) The temperature of the intercellular fluid
(v) The elimination of waste from the intercellular fluid

If the above five states of the intercellular fluid are held fairly constant, then the internal environment is considered to be in homeostasis. Recall that homeostasis is maintained by both positive and negative feedback. *Many organs and organ systems of the body are designed by natural selection to secure intercellular homeostasis* (Basanta, et al. 2008). For example, it is crucial that intercellular temperature be within a certain range so that metabolic processes can occur. To this end, overall body temperature must remain at a certain level to ensure that intercellular homeostasis is maintained. This example is a glimpse into the interconnected and

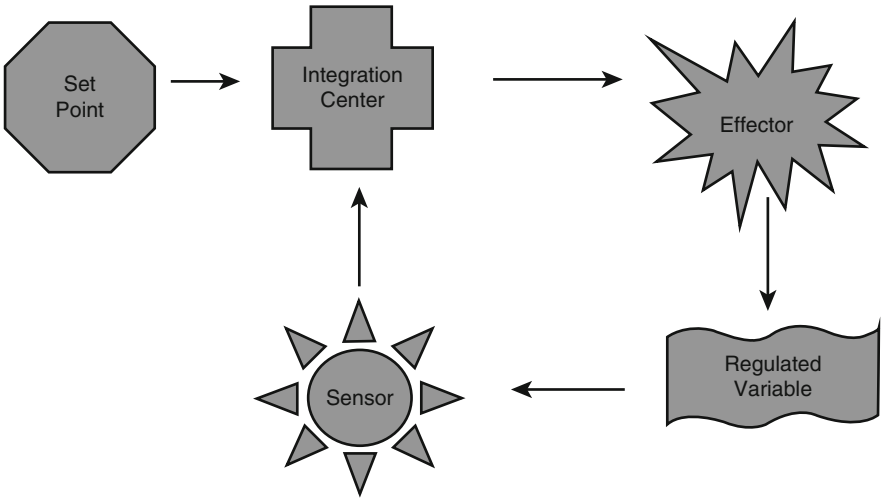


Fig. 1 Standard homeostatic system (With permission from McGraw-Hill Education)

hierarchical nature of the human body. As a way of elaborating on the discussion, Fig. 1 from Seidel (2002, p. 3) offers a pictorial look at the five elements of a standard homeostatic system.

Through the example of body temperature, the above five components can be summarized as follows (Roberts 1986, pp. 205–10):

(i) **Regulated variable** is a variable that is kept constant. For example, the following are regulated variables: body temperature, blood pressure, and the blood content of glucose, oxygen, and potassium ions. (Note that heart rate, cardiac resistance, urine output, and breathing rate are not regulated outputs. Rather, they are usually understood as *effectors*, which are designed to maintain set point levels.)

(ii) **Set point** is a quantitative value for the regulated variable. For example, 98°F is the approximate temperature of the interior of the human body.

(iii) **Sensor(s)** assesses the current status of the regulated variable. The anterior hypothalamus and the skin are the temperature-sensor organs of the human body.

(iv) **Integration center** compares current conditions with the set point. The anterior hypothalamus is the organ that acts as the integration center for the human body. It receives the information about surface body temperature from skin nerve endings. It compares this information with the set point.

(v) **Effector** brings current status of the regulated variable into line with the set point. With respect to body temperature, this feedback process is also initiated by the hypothalamus and is an effector along with the anterior and posterior hypothalamus. If the body temperature is above the set point (i.e., overheated), then sweat production is initiated, and the shivering center is inhibited in order

> to return the body's temperature to its set point. If the body temperature is below the set point (i.e., under-heated), then cellular metabolism is increased through the anterior hypothalamus, and shivering is increased through the posterior hypothalamus in order to return the body's temperature to its set point. If there is a foreign invader (e.g., bacteria), body temperature can rise in an attempt to destroy it. In the case of a fever, the set point itself increases as well.

The above elements of homeostasis are relevant to the various organs and organ systems of the body. The result is a feedback loop between the intercellular fluid and many of the other structures of the body.

Organism Homeostasis

What about the physiological functions of the body as a whole? Boorse hints at this concern in the above quotation when he defines dynamic equilibrium as "the normal functional variation within organisms acting and reacting to their environment." Then he goes on to claim that "the normal level of all part-functions varies with *what the organism is doing*" (my italics). Boorse's use of "within organism" suggests that he is concerned with how the internal part-functions of organisms react to their environment, but his use of "what the organism is doing" suggests that internal part-functions maintain their dynamic equilibrium *qua* homeostasis with respect to the physiological activities of the organism. For example, eating, waste removal, sleeping, running, walking, etc. are functions of the body (not merely any particular part) as a whole that help to sustain intercellular fluid. Where this analysis departs from Boorse is that intercellular homeostasis, which is associated with the integrated internal activities of the body, is distinct from the external behaviors of the body. These external behaviors are distinct evolved patterns that are not only being influenced by intercellular fluid, but are also influencing intercellular homeostasis (Ananth 2008). Therefore, the idea is that there is a dual-homeostatic interaction between behavioral activities of organisms as a whole and their intercellular activities.

For example, sleeping, eating, and waste removal are necessary for regaining lost energy. Energy restoration is crucial not only for carrying out intercellular processes but also for the organism as it contends with daily environmental disturbances. Importantly, these physiological activities are *coordinated activities* of the organism as a whole as it interacts with its environment – they cannot be understood in terms of parts alone. What part of the body pumps blood? What part of the body has the function of walking? What part of the body has the function of sleeping? The answer to the first question is the heart. The remaining questions do not have such a straightforward answer. The reason is that walking, sleeping, eating, reproducing, swimming, etc. are evolved coordinated activities of an organism as a whole in relation to its environment. It is the body as a whole that walks. In general, the legs, the arms, and torso coordinate to create a pattern of activity called walking. Similarly, it is the body as a whole that reproduces. It is the body as a whole that sleeps. It

is these sorts of physical activities of the whole body that cannot be captured by only a part-functionalism or intercellular homeostasis, but is relevant to what an organism is doing as a whole entity. It is to these sorts of activities that "organism homeostasis" refers.

In defense of this holistic notion of body movement, Goldstein (1963, pp. 229–230) says that when humans make a certain movement

> we do not innervate individual muscles or muscle groups, but a change in the present state of *innervation* of all the body muscles takes place. Thus, a *pattern* of innervation results, in which one definite single contraction, namely, the one which is intended, stands in the foreground. For the appropriate contraction of one muscle group, i.e. for that contraction by which a definite effect results, a certain state of innervation of the remaining body muscles is requisite. To be sure, we do not notice this state of innervation, because it seems to be insignificant for the intention of that movement. But it is not at all insignificant, it rather *enables the organism* to execute the movement correctly.

As Goldstein makes clear, specific body movements require that all (he probably means most) body muscles (in addition to the specific muscles of a particular movement) be coordinated or stimulated to ensure that the specific body movement is accomplished. The pattern of movement that emerges is in conjunction with the pressures from the external environment. Drawing from Goldstein's account, it is this pattern of movement in response to environmental stress that is the product of natural selection. For example, the overall patterns of swimming motion of fishes or flying patterns in birds are evolved patterns that are crucial to survival and reproductive success. It is these sorts of behavioral patterns that allow organisms to interact in an energetically balanced way with their environments. It is the energy balance created by these coordinated behaviors that is here being called organism homeostasis. On this view, organisms share a close relationship with their environment such that energy balance is part of understanding organisms as ecologically oriented creatures (Depew 1998 and Keesey and Powley 2008).

Thus, along with intercellular homeostasis, organism homeostasis must be included in the discussion of homeostasis. That is, since it is the overall organism that directly contends with the environment, intercellular homeostasis can be viewed as a necessary condition for overall physiological homeostasis, which is crucial to both intercellular homeostasis and survival and reproduction. For instance, walking, running, sleeping, jumping, grasping, and other evolved behavioral activities can be viewed as effectors that are crucial to maintaining an organism's life (i.e., energy balance maintenance) under a certain range of environmental influences.

In his defense of a naturalistic concept of health, Bechtel (1985, p. 149) hints at this sense of "organism homeostasis":

> The idea that living organisms incorporate a complex organization that makes them homeostatic systems provides an important element needed in a satisfactory physiological concept of health. In terms of it, one can define a healthy system as one that is at or near its designed equilibrium state. Significant deviations, especially those in which some external agency is required to restore the system to the equilibrium state, are disease states.

According to Bechtel, the complexity of the human body is sufficient to understand it as a homeostatic system. Notice that this claim is distinct from the idea that the human body is composed of homeostatic systems. The first claim refers to what is here being called organism homeostasis, while the second claim refers to internal homeostasis. To this end, the physical activities of the brain (or specific parts) can be seen as the integration centers that assist in fight-flight responses, resting responses, bathing responses, etc. These overall physical functions of the body are not easily captured by a strict part-functionalist account. Rather, this requirement of dual homeostasis reveals that there are functions that can be attributed to the organism as a whole. The upshot is that, once the discussion on health includes homeostasis, organic functional holism is compatible with part-functionalism.

This section has offered a brief glimpse into the dual nature of the concept of homeostasis with respect to the evolved body. It included a discussion of both intercellular homeostasis and organism homeostasis and a general explanation of the different elements that comprise a homeostatic system. The general conclusion that should be drawn from this section is that homeostasis should include not only the internal intercellular balances of the body, but the many behavioral activities of the body designed to contend directly with the environment.

Bringing the Bundle Together

This chapter began with a plea from the **EM** camp that medical practitioners and educators ought to incorporate evolutionary thinking into both their **PHJs** and overall pedagogy. This plea requires accepting a set of philosophical and physical concepts that make adequate sense of the body as an evolving system. Endorsing and implementing such a request also require that the body be taken seriously as a bundle of evolutionary compromises. Moving in this direction further requires endorsing a specific concept of biological function, namely, **MEF**. Additionally, the body as a dual-homeostatic system completes such an account of the body as an ecologically bound system. Synthesizing all this reveals that, from the perspective of **EM**, *the body is best understood as an ecologically oriented and evolutionary dynamic dual homeostatic properly functioning holistic system*.

The Evolved Bundled-Homeostatic Body: Influence on the Practice of Medicine and Public Health

Given the predominantly theoretical analyses of the previous sections, it is reasonable to offer a number of examples that reveal the usefulness to medical practitioners of including an **EM** perspective in their conceptual tool kit. The examples included below are (1) defense mechanism of fever, (2) genetic disease of sickle-cell anemia, and (3) the immunological response of allergies.

Defense Mechanism: The Fever

The body has many defense mechanisms of which fever is a classic example. In general, a fever is indicative of the body attacking a pathogen. In mammals, the internal heat created by fever is an evolved immunological feature of the body that has the function (i.e., **MEF**) to destroy and/or mitigate the proliferation of a foreign invader. All of this is clear to the medical community, so what can **EM** contribute to this case? For a start, it is important to keep in mind that pathogens need to be understood in terms of a pathogen-host-environment relationship. As noted before, *since adaptations are adaptations to local environments*, it is eminently reasonable that there are pathogens that cannot be controlled by way of a febrile response because the class of pathogen has either evolved to withstand the fever response or it is a species of pathogen that is foreign to the range of environmental dynamism with which the human body has dealt. This means at least one of two things. One, pathogens are able to produce variations (as a result of their rapid reproductive rates) that can withstand the febrile defense response – human evolutionary adaptations cannot keep pace with the swift evolutionary changes exhibited by some kinds of rapid reproducing organisms. Two, the foreign pathogen is a novel invader that is either unencumbered by the febrile response or benefits from such a pyretic environment. Although all of these possibilities could be hypothesized without necessarily employing the tenets of **EM**, an evolutionary take could very well assist in rendering the realities of such possibilities, if missed from the traditional part-functionalist/mechanical medical model, and could guide future experiments regarding the presence of these sorts of unanticipated variations.

Indeed, artificial attempts to control a fever, either by allowing it to persist or stopping it (e.g., the use of aspirin or acetaminophen), run the risk of controlling certain bacteria while simultaneously allowing for the proliferation of other viruses. For example, not lowering a fever can reduce the availability of iron to the body and this can mitigate the growth of some bacteria. (*Again*, *this should not be a surprise when thinking of the body as a bundle of compromises in the attempt to maintain energy homeostasis – heat increase compromises quantity of iron in the body.*) Simultaneously, this reduction in iron can cause, for example, *E. coli* and *Vibrio cholera* to produce toxins that actually exacerbate illnesses. In contrast, there is some evidence that artificially hindering a chicken pox-induced fever in children resulted in slower recovery than those who only took a placebo. Also, in another study, patients who took aspirin or acetaminophen for a common cold had a poorer antibody response and greater nasal stuffiness than those who were on a placebo (Nesse and Williams 1994, pp. 27–28). As Nesse and Williams stress (1994, p. 29), "The important point, with respect to the adaptive significance of fever, is that we need to know what we are doing before we interfere with it...We hope that medical research will soon produce the evidence to help doctors and patients decide when fever is and is not useful." Yet, this should not be perceived as a trivial task given the nature of the human body. As Ewald (1994, p. 19) cautiously points out, "fever could be a weapon that backfires, causing worse disease than would be present without

fever." Thus, **EM** proponents are stressing that distinguishing defenses from other expressions of infection requires in part to "respect the evolved wisdom of the body" (Williams and Nesse 1994, p. 31).

This is a reminder that the human body as an ecologically oriented bundle of evolutionary compromises must contend with other organisms that have their unique compromised bundles that are subject to selection. For example, researchers (Bishai et al. 1996) treated a patient with rifampin (an antibiotic) in an attempt to destroy the tuberculosis bacteria, and it appeared that this antibiotic was effective; indeed, doctors could not culture any tuberculosis bacteria in the lungs of the patient to whom rifampin was given. Unfortunately, the patient did succumb, and upon autopsy and DNA sequencing, it was determined that a mutated version of the tuberculosis bacteria was able to withstand the rifampin. As Freeman and Herron (2004, p. 511) point out, "the data are consistent with the hypothesis that bacterial populations evolve in response to selection imposed by antibiotics." So, from an **EM** perspective, this sort of example reveals that not only is the human body a bundle of evolutionary compromises, but that there can be organisms and cells within the body that are themselves evolving in response to human intervention. At the very least, the **EM** perspective can be employed in supplementing the view that the judicious use of antibiotics is paramount when bundles of evolutionary compromises are competing with and within one another.

Additionally, the use of aspirin suppresses fever, pain, and inflammation, while acetaminophen suppresses only fever and pain. This is important because inflammation can reduce the proliferation of some viruses by allowing additional defense responses at the point of inflammation and reducing the flow of blood and thus movement of the viruses from the infection area (Ewald 1994, p. 21). For instance, Ewald (1994, p. 22) tells us that certain infections in mice are exacerbated as a result of inflammation suppression via anti-inflammatory drugs. What this means is that an **EM** perspective regarding variation in viruses could determine whether or not to use aspirin or acetaminophen to treat a particular illness and to be on the lookout for such variations. Drawing upon similar **EM** insights with respect to cholera and dysentery, it is clear why Ewald (1994, p. 19) tenders the following conclusion: "Evolutionary and biochemical principles therefore suggest that the overall net effect of fever may be positive or negative. . .Because these alternative evolutionary scenarios have not been generally recognized, key experiments to distinguish between them have not been done." Thus, the fever example illustrates how **EM** can supplement the efforts and directions of biomedical research as well as decisions made by doctors caring for patients.

Genetic Disease and PHJs: Sickle-Cell Anemia

Sickle-cell anemia is a blood disorder caused by a gene that is also beneficial. This disorder occurs mostly in people from Africa (and some parts of India), where malaria is a major cause of death. To understand sickle-cell anemia, a few definitions related to genetics need to be made clear. First, alleles are alternative forms of a

particular gene that affect a specific trait in different ways. For example, consider eye color. Assume that brown eyes are "dominant" over blue eyes. "B" refers to a dominant allele. "b" refers to a recessive allele. The gene for brown-colored eyes includes the following set of alleles: **BB** and **Bb**. The alleles for blue eyes are **bb. BB** is a condition known as *homozygous dominant*. This means that so long as **BB** alleles are present, brown-colored eyes will always be present over any other colored eyes. **Bb** is the condition known as *heterozygous*. In this case, a person has both a dominant and a recessive gene. In heterozygous cases, the dominant allele swamps the effects of the recessive allele. So, **Bb** will produce brown eyes, even though a recessive gene is present. Finally, **bb** refers to the condition known as *homozygous recessive*. In this case, a person has two recessive alleles. With respect to eye color, **bb** will produce blue eyes.

In principle, sickle-cell anemia occurs in a similar way, but the effect under consideration is red blood cell modification with respect to malaria parasites. Assume that **RR** is the homozygous dominant condition, **Rr** is the heterozygous condition, and **rr** is the homozygous recessive condition. Genetically, in a simplified rendering, the three conditions produce the following effects:

1. Homozygous Dominant: These people carry two of the same forms of the gene (alleles) and are not able to modify the shape of their red blood cells. Although there are no detrimental side effects, these people are unable to defend against malaria parasites.
2. Homozygous Recessive: These people carry two recessive alleles that are able to modify the shape of their red blood cells. However, as result of this modification, they also suffer crippling side effects. This group of people is said to have *sickle-cell disease*.
3. Heterozygous: These people carry one dominant and one recessive allele. The recessive gene **r** is able to modify the shape of the red blood cells. Moreover, the combination of **Rr** defends against malaria without any serious crippling effects in certain environments.

Nesse and Williams (1994, p. 99) report on people with the homozygous recessive condition as follows: "Their red blood cells twist into a crescent or sickle shape that cannot circulate normally, thus causing bleeding, shortness of breath, and pain in bones, muscles, and the abdomen." Again, these people are said to have the *sickle-cell disease*. Those who are homozygous dominant for this gene have normal red blood corpuscles, but are unable to defend themselves against malaria. However, those who are heterozygous for the gene have the *sickle-cell trait*. These people have a hemoglobin structure that is able to remove the infected malaria parasites before they cause serious damage to the body (for further details, see Salthe 1998, p. 15 and Kark 2000, url). Much like Down's syndrome, sickle-cell anemia (the homozygous recessive condition) is a genetic disease and can only be acquired through the genes of parents.

How are we to make sense of **PHJs** with respect to sickle-cell anemia? From the **EM** perspective, with respect to those who have the homozygous recessive genes, they are deemed unhealthy (Ananth 2008). Both intercellular fluid and organism

homeostasis are greatly disrupted, rendering these people very unhealthy. In contrast, people who are homozygous dominant become unhealthy only if they contract malaria. In the case of those who are homozygous dominant, it is clear that the role of the environment is crucial to their health. So long as these people do not live in malaria-infested areas, they will have no health concerns with respect to their genetic condition. In contrast, the homozygous recessive condition will render a person very unhealthy in just about any environmental condition, because the deformation of the hemoglobin is an inevitable consequence of being homozygous recessive.

Now, the heterozygous condition must be assessed. At first glance, it appears that this condition is healthy, since malaria can be destroyed. From an **EM** perspective, Nesse and Williams (1994, p. 99; bracketed additions mine) offer the following summary judgment:

> The sickle-cell gene thus illustrates *heterozygote advantage*. Because of their resistance to malaria, heterozygotes are favored over both kinds of homozygotes: Homozygotes [who are recessive] for the sickle-cell allele have low fitness resulting from sickle-cell disease, while homozygotes [who are dominant] for the normal allele have low fitness resulting from their vulnerability to malaria.

One additional point needs to be made explicit concerning the above description. Specifically, these fitness claims by Nesse and Williams must be qualified with respect to the environment, because *adaptation* means *adaptation to local environments*. So, the fitness advantage that heterozygotes have over both sets of homozygotes is relative to the low-altitude environment in which malaria is present. If, however, people with each of these conditions were placed in an environment where no malaria existed and the altitude was very high, then the fitness advantages would change. Although the heterozygotes and the homozygote dominant people would still have a fitness advantage over the homozygote recessive people, the heterozygote people would no longer have a fitness advantage over the homozygote dominant people. The reason is that the heterozygote condition in high altitudes does not confer its propensity advantage in such places. That is, the high altitude causes the red blood cells to be modified. The result of this modification is hypoxia, which can produce fainting spells and other physically harmful conditions. In such a scenario, the homozygous dominant people would have a fitness advantage over the heterozygote people.

In general, heterozygote women are more prone to urinary tract infections than homozygote dominant women. Note that the extent to which the heterozygote condition is physically harmful with respect to exercise and other scenarios is unclear or controversial from the data collected and studied. (For the experiments and other physical ailments associated with the heterozygote condition, see Kark 2000 url.) Such details may further support the view that, from an evolutionary perspective, being "healthy enough" tells against strict optimality views regarding fitness – being ecologically bound bundles of compromises results in unwanted harmful side effects because selection can only work with variation that is available. Of course, if dual homeostasis is disrupted enough, then ascriptions of poor health are warranted in terms of **PHJs**.

Restated, in the low-altitude/malaria environment, the heterozygote people are healthier than both sets of homozygote people, because the evolved survival/reproduction propensity is present. Moreover, if it is true that there are side effects from the heterozygote condition even in this environment, then these people may still be somewhat unhealthy, but healthier than both homozygous people. The point is that adaptive traits need to be understood in relation to local environments – organisms, including humans, are ecologically bound entities – in making **PHJs**. The **EM** camp is sensitive to such nuances created by natural selection in ways that may not be emphasized or incorporated into the thinking of healthcare practitioners.

Allergic Immunity

Allergies are another set of conditions to consider with respect to **EM** (Ananth 2008). There are many types of allergies. They can be partially categorized as follows: (1) injected allergies (drugs, venom), (2) ingested allergies (foods), (3) inhaled allergies (pollen and animal dander), and (4) skin allergies (plants) (Profet 1991 and Barnes et. al. 1999). Allergies, which occur in varying degrees, are responses by the immune systems. In some cases, a minor allergic reaction can result in itchy eyes, mild sneezing, or slight inflammation of the tongue, having little or no serious effect on the dispositional properties of physiological homeostasis. The result is that an organism with a mild allergy is a relatively healthy organism. For example, some people are mildly or severely allergic to cat hair. In an environment where cats are present, people will be considered unhealthy or healthy to some degree, depending upon how their systems react to cats. Indeed, some people have no allergic reaction to cats. Of course, severe allergic reactions (e.g., bee stings in some people) can result in acute respiratory and pulmonary distress. In these sorts of cases, the organism is extremely unhealthy, because the dispositional properties of physiological homeostasis have been greatly reduced in the particular environment. Different sorts of allergies reveal that health is a state that not only admits of degree but may also admit of duration and vary with local environmental conditions.

A further qualification about allergies is needed. Many allergic reactions are immune system responses governed by the immunoglobulin-E (IgE) system. Some have argued that allergy is a vestigial system that is beneficial in other species, but simply damages tissue in humans and should be viewed as an immune-response error. Thus, much like the appendix, the IgE system can cause physiological problems, but has no present function. In response, Profet (1991, pp. 24–25) has argued that the IgE system is a specialized evolved backup system to remove toxins from the body. As she notes, "The evolutionary persistence of the allergic capability, despite its physiological costs, implies the existence of an adaptive benefit for this capability that outweighs the costs; this undermines the view that allergy is an immunological error." The idea is that the body does have various toxin-fixing antibodies and enzymes that can decompose various sorts of chemical toxins. Yet, there are some toxins (e.g., venom, industrial pollutants, phenolic acids, and alkaloids) that are able to bypass these defenses. Profet argues that the IgE system is a

second round of defense designed to eliminate these sorts of toxins that have evaded initial detection. In her own words (Profet 1991, p. 27), "[A]llergy is designed to be a last line of defense against toxins; that is, the allergic response is triggered when individual's primary antitoxin defense mechanisms have proven on a previous occasion to be insufficient in preventing a specific toxin from persisting in the bloodstream and damaging cells." Randolph Nesse and George Williams (1994, p. 163; bracketed addition mine) offer the following summary of Profet's theory about certain allergic reactions:

> [An allergy] gets toxins out of you in a hurry. Shedding tears gets them [i.e., toxins] out of the eyes. Mucous secretions and sneezing and coughing get them out of the respiratory tract. Vomiting gets them out of the stomach. Allergic reactions act quickly to expel offending materials. This fits the rapidity with which toxins can cause harm. A few mouthfuls of those beautiful foxgloves in your garden can kill you a lot faster than a phone call can summon aid. Appropriately for Profet's theory, the only part of our immunological system that seems to be in a great hurry is that which mediates allergy. Other aspects of allergy that she mentions in support of her theory include the propensity to be triggered by venoms and by toxins that bind permanently to body tissue, the release of anticoagulants during allergic inflammation to counteract coagulant venoms, and the apparently erratic distribution of allergies to specific substances.

With respect to **EM**, allergies may seem problematic. For the IgE system has an evolved propensity to fight off certain toxins, it does so by disrupting homeostasis. That is, the IgE system is a biological function that can render an organism unhealthy.

The reply to the above difficulty is a reminder that evolutionary systems are not perfect systems. In an attempt to resolve one problem, biological systems can have disrupting side effects. This is simply the result of the body as a bundle of evolutionary compromises. With respect to homeostasis, the IgE system is an effector that has the evolved propensity of maintaining intercellular homeostasis. In order to do this, it must (to some degree) disrupt organism homeostasis at times. (A fever is an effector in much the same way.) According to **EM**, allergic responses (and fevers) can render an organism mildly or severely unhealthy (depending upon their degree of disturbance to intercellular and organism homeostasis) in the short run, so that both intercellular and organism homeostasis can be secured in the long run. This example stands as reminder that biological features that appear to reduce the health status of people may have evolutionary functions that are not obvious. These sorts of examples require that the **PHJs** that are made be qualified for short-term and long-term benefits. The **EM** perspective is respectful and alert to such scenarios. So, rather than telling against **EM**, allergies (i.e., biological functions that can cause harm) validate it.

Concluding Remarks

As *The Origin of Species* comes to its finish, Darwin (1859/1958, p. 458) renders the following energetic proclamation:

> In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation. . .Much light will be thrown on the origin of man and his history.

It is interesting that Darwin chose to emphasize psychology as the "new" field of choice in anticipation of the fecundity he saw in his nascent ideas. For better or worse, evolutionary psychology has exploded, as Darwin envisioned, despite its arguably more delicate foundation (Buller 1999). Yet, despite its rather solid scientific underpinnings, **EM** is still awaiting for its Darwinian detonation to occur – much to the chagrin of its proponents. On the modest appeal that **EM** can augment the reflection, practice, and training of medical practitioners and their cohorts, **EM** champions would, in the spirit of Darwin's affirmations, enthusiastically bellow that embracing the body as a bundle of evolutionary compromises could very well complement and point toward promising areas of research within medicine.

Definition of Key Terms

Evolutionary medicine	discipline concerned with both biological and philosophical issues pertaining to the field of medicine and public health.
Evolutionary body	a bundle of adaptive compromises best understood as an ecologically oriented and evolutionary dynamic dual homeostatic properly functioning holistic system.
Physical health judgments	health and disease judgments pertaining to the physical human body made by medical practitioners and public health officials.
Systemic Functionalism	a feature **X** has a function in system **S** if and only if activity **Y** of **X** causally contributes to **S's** overall capacity/performance of **Z**.
Etiological Evolutionary Functionalism	a feature **X** currently has a function in an organism **O** if and only if activity **Y** of **X** produces effect **E** because **Y** and **E** were naturally selected (over some other causes and effects) to bring about the goals **G** of survival and reproductive success of **O**.
Propensity Evolutionary Functionalism	a feature **X** has a function in an organism **O** by performing activity **Y** if and only if **Y** produces effect **E** because **Y** and **E** confer and will continue to confer a

	propensity **P** (within a certain range of environmental pressures) to bring about the goals **G** of survival and reproductive success of **O**.
Mixed Evolutionary Functionalism	a feature **X** has a function in an organism **O** by performing an activity **Y** if and only if **Y** produces effect **E**, and both **Y** and **E** confer a survival-enhancing propensity **P** on **O** (within a certain range of environmental pressures) and will continue to confer **P** on **O** (so long as a certain range of environmental pressures is present). And, moreover, **P** is currently present, because, ancestrally, there was natural selection in favor of retaining **P** to bring about the goals **G** of survival and reproduction.
Intercellular homeostasis	the internal state of an organism in which the stability of chemical composition, osmotic pressure, carbon dioxide, temperature, and quantity of waste with respect to intercellular remains constant.
Organism homeostasis	the overall state of an organism with respect to its behavioral and ecologically oriented activities in conjunction with intercellular homeostasis.

Summary Points

- *Section "Introduction" summary point*: As a supplementary perspective to medical judgments, evolutionary medicine [**EM**] claims that it can assist in the theoretical, educational, and everyday practices of medical practitioners and public health officials.
- *Section "What is Evolutionary Medicine?" summary point*: **EM** explores the philosophical, theoretical, and practical implications of incorporating an evolutionary perspective to the existing field of medicine.
- *Section "The Human Body as an Evolutionary Entity: The Compromised Physical Self" summary point*: **EM** explores the importance of viewing the body as a "bundle of evolutionary compromises." From this perspective, medical practitioners are urged to embrace an evolutionary understanding of biological function and homeostasis as crucial to the understanding of the body as an ecologically oriented entity.

- *Section "The Evolved Bundled-Homeostatic Body: Influence on the Practice of Medicine and Public Health" summary point*: Some of the implications and overall efficacy of the **EM** conception of the human body are shown by an examination of some of the theoretical and practical approaches to (1) defense mechanism of fever, (2) genetic disease of sickle-cell anemia, and (3) the immunological response of allergies.
- *Section "Concluding Remarks" summary*: Charles Darwin was confident that an evolutionary perspective would open new areas of study with respect to the human animal. Although he specifies psychology as the prominent discipline that would blossom, contemporary advocates of **EM** would be delighted if he had included the field of medicine as well.

References

Alexander HK, Martin G, Martin OY, Bonhoeffer S (2014) Evolutionary rescue: linking theory for conservation and medicine. Evol Appl 7:1161–1179

Ananth M (2008) In defense of an evolutionary concept of health: nature, norms, and the human biology. Ashgate Press, Aldershot

Ananth M (2017) Bringing biology to life: an introduction to the philosophy of biology. Broadview Press, Calgary

APHA (American Public Health Association) 2016. https://www.apha.org/what-is-public-health

Ariew A, Cummins R, Perlman M (eds) (2002) Functions: new essays in the philosophy of psychology and biology. Oxford University Press, Oxford

Barnes KC, Armelagos GJ, Morreale SC (1999) Darwinian medicine and the emergence of allergy. In: Trevathan WR, Smith EO, McKenna JJ (eds) Evolutionary medicine. Oxford University Press, New York, pp 209–243

Basanta D, Miodownik M, Baum B (2008) The evolution of robust development and homeostasis in artificial organisms. Comput Biol 4:1–10. doi:10.1371/journal.pcbi.1000030

Bechtel W (1985) In defense of a naturalistic concept of health. In: Humber JM, Almeder RF (eds) Biomedical ethics review 1985. Humana Press, Clifton, pp 131–170

Bergman EM, Verheijen IWH, Van der Vlueten CPM, de Bruin ABH et al (2014) Influences on anatomical knowledge: the complete arguments. Clin Anat 3:296–303

Bernard C (1957) An introduction to the study of experimental medicine. Dover Publication, New York

Bigelow J, Pargetter R (1987) Functions. J Philos 84:181–196

Bishai WR, Graham NMH et al (1996) Rifampin-resistant tuberculosis in a patient receiving rifabutin prophylaxis. N Engl J of Med 334:1573–1576

Boorse C (1987) Concepts of health. In: Van DeVeer D, Regan T (eds) Health care ethics: an introduction. Temple University Press, Philadelphia, pp 359–393

Boorse C (1997) A rebuttal on health. In: Humber JM, Almeder RF (eds) What is disease? Humana Press, Totowa, pp 1–134

Boorse C (2002) A rebuttal on functions. In: Ariew A, Cummins R, Perlman M (eds) Functions: new essays in the philosophy of psychology and biology. Oxford University Press, Oxford, pp 63–112

Buller DJ (1999) Adapting minds: evolutionary psychology and the persistent quest for human nature. MIT Press, Cambridge

Cannon WB (1963) The wisdom of the body, revised and enlarged edn. W. W. Norton & Company, New York

Cooper SJ (2008) From claude bernard to walter cannon. emergence of the concept of homeostasis. Appetite 51:419–427

Cummins R, Roth M (2010) Traits have not evolved to function the way they do because of a past advantage. In: Ayala FJ, Arp R (eds) Contemporary debates in the philosophy of biology. Wiley-Blackwell, West Sussex, pp 72–85
Darwin C (1859/1958) The origin of species by means of natural selection or the preservation of the favoured in the struggle for life, mentor edn. Mentor/Penguin Books, New York
Depew D (1998) Darwinism and developmentalism: prospects for convergence. In: van de Dijver G, Salthe SN, Delpos M (eds) Evolutionary systems: biological and epistemological perspectives on selection and self-organization. Kluwer, Dordrecht, pp 21–32
Dobzhansky T (1973) Nothing in biology makes sense except in the light of evolution. Am Biol Teach 35:125–129
Ewald PW (1994) Evolution of infectious disease. Oxford University Press, Oxford
Freeman S, Herron JC (2004) Evolutionary analysis, 3rd edn. Pearson/Prentice Hall, Upper Saddle River
Gilchrist GW, Joel G, Kingsolver JG (2001) Is optimality over the hill? In: Orzack SH, Sober E (eds) Adaptationism and optimality. Cambridge University Press, Cambridge, pp 219–241
Goldstein K (1963) The organism, a holistic approach to biology derived from pathological data in man. Beacon, Boston
Hardcastle V (1999) Understanding functions: a pragmatic approach. In: Hardcastle V (ed) Where biology meets psychology. The MIT Press, Cambridge, pp 27–43
Hartig T, Mitchell R, de Vries S, Frumkin H (2014) Nature and health. Annu Rev Public Health 35:207–228
Kark J (2000) Sickle cell trait. http://sickle.bwh.harvard.edu/sickle_trait.html
Keeney RE, Powley TL (2008) Body energy homeostasis. Appetite 53:442–445
McLaughlin P (2001) What functions explain. Cambridge University Press, Cambridge
Mayr E (1974) Teleological and teleonomic, a new analysis. Boston Stud Philos Sci 14:91–117
Mayr E (2000) Darwin's influence on modern thought. Sci Am 238:78–83
Millikan R (1993) White queen psychology and other essays for alice. MIT Press, Cambridge
Nesse R, Williamd GC (1994) Why we get sick: the new science of darwinian medicine. Times Book/Random House, New York
Nesse RM, Williams GC (1997) Evolutionary biology in the medical curriculum: what every physician should know. Bioscience 47:664–666
Nissen L (1997) Teleological language in the life sciences. Rowman & Littlefield Publishers, Lanham
Odling-Smee JF, Laland KN, Feldman NW (2003) Niche construction: the neglected process in evolution. Princeton University Press, New Jersey
Older J (2004) Anatomy: a must for teaching the next generation. Surgeon 2:79–90
Papa V, Vaccarezza M (2013) Teaching anatomy in the XXI century: new aspects and pitfalls. Scientific World Journal 7281:1–5
Perlman M (2010) Traits have evolved because of a past advantage. In: Ayala FJ, Arp R (eds) Contemporary debates in the philosophy of biology. Wiley-Blackwell, West Sussex, pp 53–71
Profet M (1991) The function of allergy: immunological defense against toxins. Quart Rev Biol 66:23–54
Roberts MBV (1986) Biology: a functional approach, 4th edn. ELBS with Nelson, UK
Salthe SN (1998) The role of natural selection theory in understanding evolutionary systems. In: van de Dijver G, Stanley N, Salthe SN, Delpos M (eds) Evolutionary systems: biological and epistemological perspectives on selection and self-organization. Kluwer, Dordrecht, pp 13–20
Seidel C (2002) Basic concepts in physiology. McGraw Hill, New York
Shanahan T (1999) Darwinian medicine. Lecture given to the biology department at loyola marymount university. http://philpapers.org/rec/SHADM
Sprouffke K, Merlo LMF, Garrish PJ, Maley CC, Sniegowski PD (2012) Cancer in light of experimental evolution. Curr Biol 22:762–771
Turney B (2007) Anatomy in a modern medical curriculum. Ann R Coll Surg Engl 89:104–107

Tyreman S J (2001) The concept of function in osteopathy and conventional medicine: a comparative study. Doctoral Dissertation, British School of Osteopath

World Health Organization (2007) Darwinian medicine: basic theory with practical uses for public health. http://www.who.int/global_health_histories/seminars/2005-2007/en/

Wouters A (2003a) Philosophers on function. Acta Biotheor 51:223–235

Wouters A (2003b) Four notions of biological function. Stud Hist Philos Biol Biomed Sci 34:633–668

Wright L (1976) Teleological explanations: an etiological analysis of goals and functions. University of California Press, Berkeley

Human Nature as Normative Concept: Relevance for Health Care

19

Nicanor Pier Giorgio Austriaco O.P.

Contents

Abstract

Appeals to human nature as a normative principle for practical reasoning are often made in medical and health-care ethics. Such moral arguments presuppose an essentialist theory of human nature that posits the existence of an intrinsic human nature or essence that explains the necessary properties and causal powers that distinguish human beings from other animals. With an essentialist theory in hand, moral philosophers who advocate a metaethical theory called ethical naturalism propose that moral principles to guide human action, especially the identification of the virtues, can be discerned from and justified by considerations of human nature. Three areas in medicine and health care where human nature has been proposed as a normative principle to guide moral decision-making include attempts to describe the virtuous patient and health-care provider, debates over the legitimacy of biotechnological enhancements of human capacities, and arguments against the moral acceptability of reproductive cloning and other efforts to radically change human procreation.

N.P.G. Austriaco O.P. (✉)
Department of Biology, Providence College, Providence, RI, USA
e-mail: naustria@providence.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_78

Introduction

Appeals to human nature as a normative principle for practical reasoning are often made in medical and health-care ethics. For one influential ethical tradition that traces its roots to Aristotle, human actions, including all medical interventions, are good if they promote, perfect, or are in harmony with the ends that are perfective of human nature and are bad if they diminish, frustrate, or are in discord with those ends. Here, human nature remains a moral benchmark that cannot be altered without undermining the human good. For other traditions, however, ethical judgments are not determined by and should not be dependent upon human nature. For critics of the view that human nature is a normative guide for human action, there may even be scenarios where it is a moral imperative to change, improve, or "reinvent" human nature using medical technology or genetic engineering in the pursuit of some other perfection or ideal.

To clarify the normative role that human nature has played in discussions surrounding health care, this chapter begins with an overview of the two broad categories of theories of human nature that are prevalent in contemporary discourse. Essentialist theories posit the existence of an intrinsic human nature or essence that explains the necessary properties and causal powers that distinguish human beings from other beings. These properties and causal powers would be an objective description of the human being that is valid at all times and places, designating the human individual as belonging to a unique natural kind. In contrast, non-essentialist theories deny that there is an underlying nature that designates human begins as members of a natural kind. Instead, for non-essentialists, human beings share overlapping clusters of properties and powers, none of which all human beings must necessarily possess. Indeed, in the view of one kind of non-essentialism, called anti-essentialism by some, human nature is best described as an ad hoc clustering of properties and powers that is constructed by the particular cultures and societies in which human beings live. As such, it is ever changing and ever new.

This chapter then moves to a summary investigation of ethical naturalism, a tradition that proposes that the basic principles of morality can be discerned from a rational inquiry of the natural world. Not surprisingly, ethical naturalism presupposes an essentialist account of human nature. According to this metaethical view, human nature can be a normative guide for human action where good acts are those that help human beings to attain their species-specific ends well. An influential strand of this ethical tradition is constituted by contemporary philosophers who have been inspired by Aristotle. This is the account of morality that proposes that virtues, and not moral laws or social policies in themselves, are the proper objects of practical reasoning.

Finally, this chapter closes with a summary discussion of three areas in medicine and health care where human nature can be and has been used as a normative principle to guide moral decision-making. First, philosophers have reflected upon the ends of human nature to describe the virtuous patient and the virtuous health-care provider, to provide practitioners with a moral framework to guide virtuous human activity within the context of the provider-patient relationship. Next, appeals to

human nature have been used by critics of proposals to use biotechnology and genetic engineering to permanently enhance human capacities. Finally, references to human nature have also figured prominently in several arguments against the moral legitimacy of reproductive cloning and other efforts to radically alter human procreation.

Theories of Human Nature

What are human beings? Do they have shared properties and causal powers that distinguish them from nonhuman beings? And if they do, what are these properties and causal powers, and how do they relate to our biological and psychological constitution? To answer these and related questions, philosophers, religious thinkers, poets, and scientists throughout history have proposed theories of human nature that try not only to explain human beings but also to contextualize their fears, hopes, and aspirations. Though there appears to be broad agreement on what the term "human" means today, especially if it is understood as indicating membership in a biological species category called *Homo sapiens*, there is little agreement on what "nature" means. Nonetheless, theories of human nature can be broadly categorized into two kinds, those that are essentialist and those that are non-essentialist in character (Pojman 2006; Kronfeldner et al. 2014).

Essentialist theories of nature posit the existence of an underlying and intrinsic essence or core nature in a particular thing of a natural kind that explains the properties and causal powers it shares with other individuals of the same kind (Ellis 2001; [2002] 2014; Oderberg 2007; Bird and Tobin 2015). For example, an essentialist theory of chemical nature would explain the properties and causal powers of elemental gold, i.e., it is a bright, slightly reddish yellow, dense, soft, malleable, and ductile metal, which is relatively nonreactive, by appealing to gold having 79 protons in its nucleus, a received view called microstructural essentialism (Hendry 2006). Every atom of gold would have this essence and every other elemental atom would not. In the same way, essentialist theories of human nature posit the existence of an underlying human essence that explains the properties and causal powers that are together associated with human beings. Proponents of such an essentialist account include, among others, Greek philosopher and biologist Aristotle (384–322 BC), Christian theologian and Catholic saint Thomas Aquinas (1225–1274), French philosopher Rene Descartes (1596–1650), and contemporary Harvard scientist and sociobiologist Edward O. Wilson. The folk conception that human beings have an intrinsic underlying nature is also intuitively appealing and is widespread, even among children (Keil 1989).

Unlike the commonly held view that chemical natures are linked to their atomic numbers, there is no consensus among essentialists on what necessarily and sufficiently constitutes human nature and the properties and causal powers that emerge from it. For Aristotle and Aquinas, human beings are rational animals, matter-form unities that have the distinctive powers of thinking and desiring (Pasnau 2002). For Descartes, we are thinking extended things, body-soul composites of two substances

that can be conceived clearly and distinctly apart from each other (Skirry 2005). For Wilson, we are members of a species distinguished by a genetic makeup that has been shaped by an evolutionary process that has selected for shared adaptive traits and behaviors (Wilson 2004). Nonetheless, despite the lack of consensus on how to explain them, it is clear that anthropologists have identified human universals that are shared by human beings in different social contexts and historical periods, whether one calls them God-given characteristics or evolved adaptive traits (Brown 1991).

In contrast, non-essentialist theories deny that there is an intrinsic underlying core human nature. Proponents include, among others, English philosopher John Locke (1632–1704), Scottish philosopher David Hume (1711–1776), and the majority of contemporary philosophers working within the modern and postmodern traditions. These theories trace their roots to the nominalism of William of Ockham (c. 1287–1347) who proposed that universals do not exist, only individuals do (Loux 1974). According to the nominalists, universals are social conventions that name similarities shared by things that do not have any underlying shared reality. This claim certainly holds for artifacts. Whether a timepiece is a clock or a watch varies from culture to culture and from age to age. However, non-essentialism extends this observation from artifacts to natural kinds. Each individual thing, whether it is a watch, a gold atom, or a human being, is unique. There are no common natures.

Non-essentialists who deny the reality of biological natures, including human nature, often argue that the sort of typological categorization espoused by essentialists has been made obsolete and untenable by evolutionary theory for at least three reasons (Sober 1980; Okasha 2002). First, they point out that biological species come into and go out of existence as distinct populations of individuals are shaped and reshaped by natural selection. In other words, biological species are dynamic realities that are unlike the static essences that purportedly define a natural kind. Next, they argue that biological species as they are understood today are defined not by their intrinsic traits but by their genealogical relationships. Individuals of a species are similar not because they share a common underlying nature but because they share a common ancestor. Finally, non-essentialists claim that the variability and diversity among individuals that belong to the same species rule out the existence of biological natures. Each organism is different. It is significant for non-essentialists that there is nothing in biology comparable to an atomic number in chemistry that is shared by members of a chemical kind.

Instead, non-essentialists propose that human nature, like all biological natures, can be conceived of as covarying clusters of relational properties, capacities, or causal powers that are typically, but not necessarily, shared by individuals that belong to a population descended from a common ancestor (Boyd 1999; Griffiths 1999, 2011). Indeed, in the view of one kind of non-essentialism, called anti-essentialism by some, human nature is best described as an ad hoc clustering of properties and powers that is constructed by the particular cultures and societies in which human beings live, often to protect the interests of the powerful (Peterson 2001, Chap. 3). This account emerges from a postmodern worldview that contends

that all knowledge of reality is fabricated by the human beings who together inhabit and are influenced by their particular social, cultural, and historical contexts (Berger and Luckmann 1966). For these non-essentialists, human nature is whatever a particular community or society proposes it to be.

In response to their critics, essentialists raise three counterarguments (Devitt 2008). First, they note that the discovery of chemical transmutation where one element is changed into another element has not weakened the consensus that chemical kinds exist. In the same way, essentialists argue that the observation that one species can be transformed into another by natural selection does not necessarily rule out biological kinds. Next, they affirm that biological species are defined by their genealogical relationships. However, they also propose that a genealogical definition for a biological species does not rule out a structural definition that seeks to explain why individuals descended from a common ancestor actually have the similar properties that they do. To illustrate this point, consider three children. One could argue that they are siblings who belong to the same family because they share the same parents. This would be a genealogical explanation. However, one could also propose that they are siblings because they share certain familial traits, say, a broad forehead and a hooked nose. This would be a structural explanation. Structural explanations explain why all tigers are striped, all jaguars are black, and all kangaroos jump. They do so because they share common properties and causal powers, i.e., a common nature that distinguishes them from other individuals of other species. Finally, essentialists point out that individual atoms of the same chemical element with the same atomic number can vary because they have different atomic mass numbers. Variability and diversity do not necessarily rule out biological essentialism as long as one can find an underlying shared reality among the individuals that belong to the same biological kind. Recent work in systems biology has concluded that individual organisms of a biological species share a common molecular state space that comprises all the possible molecular states of their gene regulatory network (Auyang 1998, pp. 101–102). This molecular state space underlies the genotype and phenotype landscapes that are also uniquely associated with each biological species. A species' state space would specify a biological kind in the same way that an atomic number specifies a chemical kind.

The Normative Use of Human Nature in Practical Reasoning

Appeals to human nature as a normative principle in practical reasoning are made by advocates of moral or ethical naturalism, a metaethical theory that claims that moral principles can be justified by appealing to the architecture of human nature (Lenman 2014). Not surprisingly, ethical naturalists hold to an essentialist theory of human nature that can be a substantive guide to practical reasoning. To illustrate the basic contours of ethical naturalism, consider the influential account proposed by Rosalind Hursthouse, a contemporary moral philosopher working in the Aristotelian tradition, that links ethical naturalism to an account of the virtues (Hursthouse 1999).

Other prominent contemporary philosophers who would hold similar views include Philippa Foot (2001), Alasdair MacIntyre (1999), and Martha Nussbaum (1995).

Hursthouse begins her ethical analysis by exploring the way that we evaluate plants and animals (Hursthouse 1999, Chap. 9). She notes that it is not uncommon for us to conclude that an individual plant is a good or bad specimen of its natural kind by determining if two of its aspects, i.e., its parts and its operations, are good or bad in light of two ends, i.e., whether they are contributing, in the way characteristic of a member of its natural kind, first to the survival of the individual and then to the ongoing survival of the species. A good plant would be an individual with good roots and good leaves that facilitate its survival and that of its species. A bad plant would be a sickly plant that is unable to survive or produce progeny.

Moving up the ladder of life, Hursthouse explains that an animal is a good or bad specimen according to its natural kind not only according to whether it has good parts and good operations but also according to whether it acts well in a way characteristic of its species. Moreover, in her view, animals have an additional end beyond individual and species survival, and that is whether they are free from pain and, where appropriate, are enjoying the pleasure that is characteristic of their natural kind. Thus, a good owl would be an individual that can see well in the dark so that it can hunt well at night to nourish itself and its offspring. However, an owl that could hunt well at night but is experiencing pain in any way would not be a good owl.

Hursthouse then extends her analysis to human beings where the "criteria of goodness in human beings must be related to what human beings are and/or do as such" (Hursthouse 1999, p. 206). She proposes that as rational and social animals, human beings have four aspects, namely, their parts, their operations, their actions, and now their emotions/desires, that have to be evaluated with respect to four ends, namely, their individual survival, their species survival, their freedom from pain and enjoyment of pleasure, and additionally the good functioning of their social group. These four evaluative aspects and ends emerge from a reasoned interrogation of human nature revealed in the natural history of the human species.

Thus, for the most part, according to Hursthouse, a good human being would be an individual who is healthy, who feels and acts well, and who has good relations with others, including his family, his relatives, his friends, and his colleagues, relations that allow him and the rest of the group to live well. In her view, he would have the good character traits, which she and the Aristotelian tradition call the virtues, that allow him to live well, one of the many forms of life that are in accord with the ends that are distinctive of the human natural kind. To illustrate her point, Hursthouse proposes that human beings are good, and more specifically, they possess the virtue of courage, when they "defend themselves, and their young, and each other, and risk life and limb to defend and preserve worthwhile things in and about their group, thereby fostering their individual survival, the continuance of the species, their own and others' enjoyment of various good things, and the good functioning of the social group" (Hursthouse 1999, p. 209).

Finally, Hursthouse explains that ethical naturalism, as she and the Aristotelian tradition understand it, provides criteria for a particular character trait being a virtue,

i.e., it being conducive to a human being living a good life characteristic of his natural kind, and not criteria for right or wrong action, except indirectly. Therefore, to determine if a particular action were right or wrong, one would have to ask whether this is an action that would be undertaken by a virtuous individual in these particular circumstances, here and now, and not whether or not it is in accordance with some moral principle. Or to put it another way, one would have to ask whether this particular action would promote or undermine the virtue of the agent who is contemplating this particular action here and now. For ethical naturalists, the virtues, and not moral laws or social policies in themselves, are the proper objects of practical reasoning.

The Normative Use of Human Nature in Medicine and Health Care

Human nature has been used as a normative principle to guide moral decision-making in medicine in different contexts. Three will be summarized here to illustrate how appeals to human nature have played a diversity of roles in ethical debates in health care.

First, human nature can be used to describe the virtuous patient and the virtuous health-care provider in a manner analogous to the way that it has been used to paint the portrait of the virtuous human being (Pellegrino and Thomasma 1993; Austriaco 2011, pp. 113–119). These descriptions can then serve as a moral framework to guide virtuous human activity within the context of the provider-patient relationship.

To illustrate this approach, a virtuous patient would be described as an individual who acts well to promote her own well-being, the well-being of her family, and the well-being of her community, in the particular context of one who is sick and in need of health care. For herself, the patient would be called to be disciplined, temperate, and courageous so that she can successfully complete her medical regimen, no matter how difficult and inconvenient it may be for her. For her family and her community, she would be motivated to be charitable, long-suffering, self-forgetful, and cooperative, ordering her relationships with others so that everyone is challenged to seek her healing together in light of the common good. These character traits would help her to make prudent decisions regarding her health care, both medical and moral, which are integral to human flourishing.

Similarly, the virtuous health-care provider would be the individual who acts well to promote the four ends perfective of his nature, this time in the context of providing medical care to those in need. He would make sure that he cares for himself and his own well-being. This involves studying diligently, thinking clearly, and resting appropriately, among other things, so that he is at his professional best when he encounters his patients. But this self-care has to be offset with a genuine concern for the care of the other. Therefore, he too is called to be charitable, long-suffering, self-forgetful, and cooperative, so that he can work with his patients, their families, and the other members of his health-care team, even when it is inconvenient, to provide patients with the care that they need to heal. In all things, the virtuous health-care

provider would be challenged to consider the genuine good of everyone involved, his own included, when he makes the health-care decisions that promote the healing of his patients.

Next, human nature has also been used as a normative principle in the ethical debate, called the debate over enhancement ethics by some, over whether or not medical interventions should be used to permanently enhance human capacities (Buchanan 2009, 2011, Chap. 4; Giubilini and Sanyal 2015; Kaebnick 2011; Peterson 2010; Sharon 2014). Critics of efforts to use biotechnology or genetic engineering for human enhancement appeal to human nature in at least three ways.

First, critics argue that society should not permanently enhance human nature because doing so would alter our common understanding of human excellence and flourishing in ways that would undermine our social practices. For instance, in its moral analysis of proposals to enhance the performance of human beings, whether this is the performance of the vocalist, the student, or the athlete, with biotechnology, the President's Council on Bioethics highlights the excellence of the striving that we acknowledge when we honor the achievement of superior performance (President's Council on Bioethics 2003, pp. 101–157). The council then wonders whether the enhancement of the human capacities used in these performances would move us to lose sight "of why excellence is worth seeking at all, and hence what excellence really is, and how we pursue it as human beings, not as artifacts" (President's Council on Bioethics 2003, p. 156). Thomas Murray would be equally cautious about biotechnological enhancements for athletic prowess. He proposes that enhancing human capacity in sport would radically alter the meaning of sport as a human activity because "what we look for in athletes is a combination of *natural talents* and the *virtuous perfection* of these talents" (Murray 2007, p. 513). Permitting athletes to be enhanced with technological manipulation rather than through virtuous effort and practice would transform the competition of sport. It would now be an occasion to honor not athletic achievement but technological innovation.

Second, critics oppose enhancement because they argue that human nature is such a complex reality that enhancing one aspect of it could undermine the excellence of the whole. The President's Council on Bioethics, for example, is concerned that there is a danger in enhancement technologies "that we will become better in some area of life by diminishing ourselves in others, or that we will achieve superior results only be compromising our humanity" (President's Council on Bioethics 2003, p. 295). Likewise, Francis Fukuyama argues, "we want to protect the full range of our complex, evolved natures against attempts at self-modification. We do not want to disrupt either the unity or the continuity of human nature" (Fukuyama 2002, p. 172). For critics, even enhancements that seek to eliminate select limitations in our human nature would be problematic because in reality, they would be purging the human condition of opportunities for virtuous behavior and human excellence. Thus, Erik Parens challenges society to consider whether reducing our fragility, which he defines as our being subject to change and to chance, by biotechnological enhancement could in fact diminish our humanity by decreasing our appreciation for beauty, benevolence, and vulnerability (Parens 1995). Similarly, Michael Sandel is concerned that enhancement and genetic engineering could make us lose sight of the giftedness of the human

condition, because these technologies represent "a kind of hyperagency, a Promethean aspiration to remake nature, including human nature, to serve our purposes and satisfy our desires. The problem is not the drift to mechanism but the drive to mastery. And what the drive to mastery misses and may even destroy is an appreciation of the gifted character of human powers and achievements" (Sandel 2007, pp. 26–27).

Lastly, critics are concerned that biotechnical or genetic enhancement would so alter the nature of an individual such that it would, in the eyes of many of his peers, put him outside of the human species. It would make him "post-human." The Vatican is worried that genetic enhancements of this kind "would promote a eugenic mentality and would lead to indirect social stigma with regard to people who lack certain qualities, while privileging qualities that happen to be appreciated by a certain culture or society" (Congregation for the Doctrine of the Faith 2008, §27). Alternatively, Francis Fukuyama is concerned that this type of intervention would undermine the individual's dignity by depriving him of the human nature that is the source of his dignity and moral status (Fukuyama 2002, Chaps. 7–9). At a minimum, technological enhancements that are perceived to place individuals outside our species would weaken our common sense of humanity that grounds our conception of human rights as we know it.

Finally, human nature has been invoked as a normative principle in the debates over human reproductive cloning and other means of altering human procreation. To take one prominent example, the President's Council on Bioethics has argued that cloning-to-produce-children "would represent a challenge to the nature of human procreation and child-rearing" (President's Council on Bioethics 2002, p. 99). In reproductive cloning, "researchers would be transforming a sexual system into an asexual one, a change that requires major and 'unnatural' reprogramming of donor DNA if there is to be any chance of success" (President's Council on Bioethics 2002, p. 94). In the view of the Council, this "unnatural" form of reproduction would radically alter the meaning of human procreation transforming it from a begetting into a making: "The likely impact of cloning on identity suggests an additional moral and social concern: the transformation of human procreation into human manufacture" (President's Council 2002, p. 104). The Vatican has made a similar argument, condemning reproductive cloning because "[i]t represents a radical manipulation of the constitutive relationality and complementarity which is at the origin of human procreation in both its biological and strictly personal aspects. . . The difference should again be pointed out between the conception of life as a gift of love and the view of the human being as an industrial product" (Pontifical Academy for Life 1997, §3).

Objections to the Normative Use of Human Nature in Medicine and Health Care

Proponents of human enhancement and/or reproductive cloning respond to arguments against these medical interventions that appeal to human nature in at least three ways. First, they deny the existence of an intrinsic core human nature that could serve as a benchmark for moral inquiry. This is the objection of the non-essentialists and has already been considered above.

Second, they maintain that it is not clear why human nature must be normative in moral analysis. For example, Allen Buchanan is deeply critical of the Council's claim that human sexuality is sexual by nature, because he thinks that it does not follow that human procreation between same sex partners using biotechnology to combine their DNA, even though it is not sexual, is "less than human, incompatible with the fundamental dignity of humanity" (Buchanan 2009, p. 146). Thus, in his view, it is not clear why the nature of human sexuality as it is understood today should rule out procreative acts that are post-sexual. Indeed, for Buchanan, there are scenarios where it can be a moral imperative to change, improve, or "reinvent" human nature, including human sexuality, using medical technology or genetic engineering in the pursuit of some other perfection or ideal.

Finally, proponents of enhancement point out that human nature includes evolved adaptations that constrain our good in unfortunate ways. To put it another way, human nature has both good and bad aspects to it. Buchanan lists our limited altruism as one such limitation (Buchanan 2009). Would it not be reasonable, he proposes, to enhance this limitation, making human beings more generous and more sociable? This would change our human nature but in ways that improve rather than diminish it.

To briefly respond to the two final objections, critics of reproductive cloning who claim that this technological intervention would undermine the human dignity of the cloned person are arguing that manufacturing a human person is inimical to her dignity in the same way that buying and selling a human person, even with her consent, is inimical to her dignity. To affirm that a human being has dignity is to affirm that there is something worthwhile about each and every human being such that certain things ought not to be done to any human being and that certain other things ought to be done for every human being. According to opponents of reproductive cloning, among those certain things that ought not to be done to a human being because of her dignity includes selling, buying, and manufacturing her.

Finally, ethical naturalists would counter the final criticism by noting that at least the Neo-Aristotelian account grounds moral inquiry on a reasoned inquiry of the overall architecture of human nature rather than on particular evolved adaptations, hence the four aspects and four ends of human nature proposed by Hursthouse. Within this framework, biotechnological interventions that could help persons to better attain the ends of their human nature by minimizing the limiting effects of certain evolved adaptations would not be altering human nature as such.

Definitions of Key Terms

Human nature	The underlying essence or core nature in a human being that explains the properties and causal powers he or she shares with other individuals of the same kind.
Enhancement ethics	The subfield of ethics that deals with proposals to enhance human capacities with biotechnology and genetic engineering beyond what would be considered a therapeutic intervention.

Essentialism	A philosophical theory that posits the existence of an intrinsic nature in a particular thing of a natural kind that explains the properties and causal powers it shares with other individuals of the same kind.
Ethical naturalism	A metaethical theory that claims that moral principles can be justified by appealing to the architecture of human nature.
Non-essentialism	A philosophical theory that denies the existence of an intrinsic nature that explains the properties and causal powers of individuals that are thought to belong to a natural kind.
State space	The set of all possible molecular configurations that can be occupied by an organism over its developmental lifetime. Since all the individual organisms of a biological species share a common molecular state space, then a state space can be said to ground a biological kind in the same way that the atomic number of an atom grounds a chemical kind.
Virtue	Good character traits that enable human beings to act quickly, spontaneously, and happily, so that they may better attain the ends that are perfective of their nature.

Summary Points

- Appeals to human nature as a normative principle for practical reasoning are often made in medical and health-care ethics.
- Essentialist theories of human nature posit the existence of an underlying and intrinsic essence or core nature in human beings that explains the properties and causal powers they share with other individuals of the same kind.
- In contrast, non-essentialist theories deny that there is an intrinsic underlying core human nature.
- Non-essentialists who deny the reality of biological natures, including human nature, often argue that the sort of typological categorization espoused by essentialists has been made obsolete and untenable by evolutionary theory.
- Instead, non-essentialists propose that human nature, like all biological natures, can be conceived of as covarying clusters of relational properties, capacities, or causal powers that are typically, but not necessarily, shared by individuals that belong to a population descended from a common ancestor.
- Appeals to human nature as a normative principle in practical reasoning are made by advocates of moral or ethical naturalism, a metaethical theory that claims that moral principles can be justified by appealing to the architecture of human nature.
- As rational and social animals, human beings have four aspects, their parts, their operations, their actions, and now their emotions/desires, that have to be evaluated with respect to four ends, their individual survival, their species survival, their freedom from pain and enjoyment of pleasure, and additionally the good functioning of their social group.

- Ethical naturalism provides criteria for a particular character trait being a virtue, i.e., its being conducive to a human being's living a good live characteristic of his natural kind, and not criteria for right or wrong action, except indirectly.
- Human nature can be used to describe the virtuous patient and the virtuous health-care provider in a manner analogous to the way that it has been used as to paint the portrait of the virtuous human being.
- Human nature has also been used as a normative principle in the ethical debate, called the debate over enhancement ethics by some, over whether or not medical interventions should be used to permanently enhance human capacities.
- Human nature has been invoked as a normative principle in the debates over human reproductive cloning and other means of altering human procreation.

References

Austriaco N (2011) Biomedicine and beatitude: an introduction to Catholic bioethics. The Catholic University of America Press, Washington, DC

Auyang SY (1998) Foundations of complex-system theories in economics, evolutionary biology, and statistical physics. Cambridge University Press, Cambridge

Berger PL, Luckmann T (1966) The social construction of reality. Doubleday, Garden City

Bird A, Tobin E (2015) Natural kinds. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, spring 2015 edn. http://plato.stanford.edu/archives/spr2015/entries/natural-kinds. Accessed 28 June 2015

Boyd R (1999) Homeostasis, species, and higher taxa. In: Wilson RA (ed) Species: new interdisciplinary essays. MIT Press, Cambridge, MA, pp 141–185

Brown D (1991) Human universals. McGraw Hill, Boston

Buchanan A (2009) Human nature and enhancement. Bioethics 23:141–150

Buchanan A (2011) Beyond humanity? Oxford University Press, Oxford

Congregation for the Doctrine of the Faith (2008) Dignitas personae: instruction on certain bioethical questions. Libreria Editrice Vaticana, Vatican City

Devitt M (2008) Resurrecting biological essentialism. Philos Sci 75:344–382

Ellis B (2001) Scientific essentialism. Cambridge University Press, Cambridge

Ellis B ([2002] 2014) The philosophy of nature: a guide to the new essentialism. Routledge, Oxford.

Foot P (2001) Natural goodness. Clarendon, Oxford

Fukuyama F (2002) Our posthuman future: consequences of the biotechnology revolution. Farrar, Straus & Giroux, New York

Giubilini A, Sanyal S (2015) The ethics of human enhancement. Philos Compass 10:233–243

Griffiths PE (1999) Squaring the circle: natural kinds with historical essences. In: Wilson RA (ed) Species: new interdisciplinary essays. MIT Press, Cambridge, MA, pp 209–228

Griffiths PE (2011) Our plastic nature. In: Gissis S, Jablonka E (eds) Transformations of Lamarckism: from subtle fluids to molecular biology. MIT Press, Cambridge, MA, pp 319–330

Hendry RF (2006) Elements, compounds and other chemical kinds. Philos Sci 73:864–875

Hursthouse R (1999) On virtue ethics. Oxford University Press, Oxford, UK

Kaebnick GE (2011) Human nature without theory. In: Kaebnick GE (ed) The ideal of nature. The Johns Hopkins University Press, Baltimore, pp 49–70

Keil FC (1989) Concepts, kinds, and cognitive development. MIT Press, Cambridge, MA

Kronfeldner M, Roughley N, Toepfer G (2014) Recent work on human nature: beyond traditional essences. Philos Compass 9:642–652

Lenman J (2014) Moral naturalism. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, spring 2014 edn. http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral. Accessed 28 June 2015

Loux MJ (1974) Ockham's theory of terms: part I of Ockham's summa logicae. University of Notre Dame Press, Notre Dame
MacIntyre A (1999) Dependent rational animals. Open Court Publishing, Peru
Murray TH (2007) Enhancement. In: Steinbock B (ed) The oxford handbook of bioethics. Oxford University Press, Oxford, pp 491–515
Nussbaum M (1995) Aristotle on human nature and the foundations of ethics. In: Altham JEJ, Harrison R (eds) World, mind, and ethics. Cambridge University Press, Cambridge, pp 86–131
Oderberg D (2007) Real essentialism. Routledge, New York
Okasha S (2002) Darwinian metaphysics: species and the question of essentialism. Synthese 131:191–213
Parens E (1995) The goodness of fragility: on the prospect of genetic technologies aimed at the enhancement of human capacities. Kennedy Inst Ethics J 5:141–153
Pasnau R (2002) Thomas Aquinas on human nature. Cambridge University Press, Cambridge, UK
Pellegrino ED, Thomasma DC (1993) Virtues in medical practice. Oxford University Press, Oxford
Peterson A (2001) Being human: ethics, environment, and our place in the world. University of California Press, Berkeley
Peterson JC (2010) Changing human nature: ecology, ethics, genes, and god. William B. Eerdmans Publishing, Grand Rapids
Pojman L (2006) Who are we? Theories of human nature. Oxford University Press, Oxford, UK
Pontifical Academy for Life (1997) Reflections on cloning. Libreria Editrice Vaticana, Vatican City
President's Council on Bioethics (2002) Human cloning and human dignity: an ethical inquiry. President's Council on Bioethics, Washington, DC
President's Council on Bioethics (2003) Beyond therapy: biotechnology and the pursuit of happiness. President's Council on Bioethics, Washington, DC
Sandel M (2007) The case against perfection: ethics in an age of genetic engineering. Belknap Press of Harvard University Press, Cambridge, MA
Sharon T (2014) Human nature in an age of biotechnology: the case for mediated posthumanism. Springer, Dordrecht
Skirry J (2005) Descartes and the metaphysics of human nature. Continuum Press, London
Sober E (1980) Evolution, population thinking, and essentialism. Philos Sci 47:350–383
Wilson EO (2004) On human nature. Harvard University Press, Cambridge, MA

Conceptions of Health and Disease in Plants and Animals

20

Henrik Lerner

Contents

Abstract

This chapter analyzes theoretical conceptions of health and disease for plants and animals. Compared to human health, the discussion of these concepts is sparse, although animal health has received more treatment than plant health. The concept of disease seems to be taken for granted, and few attempts have been made to make a classification of specific types of diseases. The main emphasis in this chapter will be on the contemporary debate on animal health definitions. Although the theoretical discussion on these issues has so far been sparse, a

H. Lerner (✉)
Department of Health Care Sciences, Ersta Sköndal University College, Stockholm, Sweden
e-mail: henrik.lerner@esh.se

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_19

categorization of different kinds of health definitions has been presented at least for animals. The six animal health categories are health as homeostasis, as normal biological function, as productivity including reproduction, as well-being, as mental and physical control, and as ability to realize goals. Similarities and differences between health concepts for humans, animals, and plants are also discussed to see whether there is a possibility to have a universal definition for plants, animals, and humans.

Introduction

Like in humans, plants and animals are said to have diseases as well as be healthy. Here, in this chapter, theoretical conceptions of health and disease for animals and plants will be analyzed. In the introduction, there will first be a short biological–philosophical summary on relevant differences and similarities between animals, plants, and humans. Second, the state of art for definitions of plant and animal health is considered and commented on.

After the introduction, this text will analyze classifications of diseases in animals and plants which are not at all as fully developed as in human medicine. Thereafter, the focus will be changed to concepts of animal health at an individual level. This part starts with a comprehensive list of six categories of health definitions and is then followed by descriptions and a discussion of these six categories. After this follows a brief discussion of the concept of plant health and an analysis of both animal and plant health on the population level. Finally, there will be a presentation of proposals for universal definitions applicable to plants, animals, and humans.

Relevant Differences Between Animals and Plants

Animals, including humans is one of several biological kingdoms. Two other are plants and fungi. Through an analysis of biological and medical dictionaries, Lerner (2008, pp. 21–23) pointed out that the criteria distinguishing animals from plants and fungi were:

- Lack of photosynthesis (chlorophyll)
- Requirement of organic nutrients
- Requirement of oxygen
- Lack of rigid cell walls, such as plants and fungi have
- Capability of (voluntary) movement
- Presence of some form of nervous system

Plants and fungi are characterized by rigid cell walls and a lack of nervous system. Plants have photosynthesis. In this text, no specific distinction will be made between the kingdom of plants and the kingdom of fungi. Instead, “plants”

will be used as the term for both fungi and plants. The relevant differences between animals and the other two kingdoms, when it comes to important aspects of health, are the capability of voluntary movement and presence of the nervous system in animals. The latter aspect has implications for which species one can ascribe mental health.

Historically, philosophers and biologists have distinguished between plants, animals, and humans. For example, the classical philosopher and biologist Aristotle defined *telos* differently for each of these three groups. The essential faculty of plants was nutrition, of animals perception and appearance, and of humans rational thought (Irwin 2003). Both animals and humans were ascribed mental faculty in Aristotle's thoughts, but only humans were regarded as being rational. Since Charles Darwin presented his theory of evolution, biologists have questioned the separation of humans from the rest of the animal kingdom. Recent philosophers, especially in animal and environmental ethics, have emphasized the similarities between humans and at least closely related animals rather than differences. Although one normally distinguishes between the concepts of animal and human health, there are some who, in the modern discussion, suggest the unification of the discussions between animal, human, and even plant health (Nordenfelt 2006, 2007; Kůdela 2011; Döring et al. 2012, 2015). There is also a new research area called One Health, trying to bring biology, veterinary medicine, and human medicine together, where an all-embracing or universal concept is called for (Lerner and Berg 2015).

Introduction to the Subject Area

Compared to the philosophical discussion of human health, the philosophical discussion of animal and plant health is underdeveloped. Nordenfelt (2006, p. 138) maintains, with regard to animal health, that there are "not many theories but one can discern several different views in textbooks and encyclopedias." Concerning both animal and plant health, few academic journal articles deal with conceptual matters, and few books discuss the concepts at length. Often when a definition is proposed, there is a focus on how to operationalize the definition. Few textbooks for students bother defining the concept of health (Gunnarsson 2006; Döring et al. 2012).

Another important reason for the lack of ideas regarding animal health is that some researchers favor the concept of animal welfare instead. Within the research area that Lerner calls "the science of animal health and welfare" (which consists of researchers from veterinary medicine, ethology, animal science, psychology, and philosophy), the concept of animal welfare has attracted much attention and covers aspects of well-being, welfare, and health for animals (for an introduction and history of the concept of animal welfare; see Lerner 2008). Some proponents claim that health is only a part of the concept of welfare (Broom 2011). Others acknowledge health as an important concept in its own right (Gunnarsson 2006), and an extreme position holds to discard the concept of welfare in favor of health (McGlone 1993).

Proposals for definitions of animal health have been made from several different disciplines including veterinary medicine, biology, philosophy, and psychology. Scholars from Northern Europe have been particularly influential, especially Great Britain, Denmark, and Sweden. Researchers from Northern Europe are also the forerunners regarding legislation for animals.

There are only few influences from the discussion on human health. The WHO definition, claiming that "health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (WHO 1948) has, however, gained some influence. Additionally, Lennart Nordenfelt has transferred both Christopher Boorse's biostatistical theory of health as the absence of disease and his own ideas of health as ability in the animal health discussion (Nordenfelt 2006).

Most ideas of health in animals concern somatic health. The idea of mental health in animals is rather recent. Behaviorism guided the first ethologists who were studying animal behavior (Konrad Lorenz, Niko Tinbergen, and Karl von Frisch), and all behavior was explained in terms of stimulus–response patterns. During the 1960s–1970s an interest grew in studying the possibility for animals to think. In the science of animal health and welfare, a landmark was the book by Marian Stamp Dawkins called *Animal Suffering*: *the Science of Animal Welfare* (Dawkins 1980; Rollin 1998), which was a scientific account that supported the view that animals can experience other states besides pure physical pain. During the 1990s, welfare definitions based on animals' subjective (inner) states were presented and gained importance. Still, some questioned the ability of animals other than apes and dolphins to suffer (Bermond 1997), despite the increasing evidence that several species of mammals and birds do suffer (Rollin 1998). During the 2000s whole books devoted to the analysis of aspects of animal mental health appeared (McMillan 2005).

Compared to animal health, plant health has attracted much less consideration, although recently there has been some development (Nordenfelt 2007; Kůdela 2011; Döring et al. 2012). Plant pathology in the modern sense was developed in the mid-nineteenth century when a classification of diseases of plants based on causal factors was proposed, and definitions of plant health have existed since the 1940s–1950s (Kůdela 2011). One of the most promising attempts in the modern discussion borrows definitions from modern theories of human health. Döring et al. (2012) suggest that there should be a guiding framework of criteria for matters of definition.

Classification of Diseases

One popular way of defining health is to define it as the absence of disease. This is a common standpoint in veterinary medicine (Gunnarsson 2006) and such a definition is often taken for granted within the science of animal health and welfare. Therefore, it is important to deal with the concept of disease when discussing health. Without a clear definition of animal disease, one may start by analyzing classifications of

specific types of diseases. Three kinds of classification criteria seem to be present within veterinary medicine. Diseases are either classified by the kind of tissue or organ that is affected, by the type of tissue change occurring, or by the causal factors behind the disease (Broom and Kirkden 2004).

For animals there is no complete international classification of diseases like the *International Statistical Classification of Diseases and Related Health Problems* (ICD) for humans. The world organization for animal health, *Office International des Epizooties* (OIE), has developed a list of diseases that should be reported internationally. These are mainly epizootic diseases, highly transmissible, or zoonotic, transmissible to humans (OIE 2015).

Another attempt to create an international classification is through the *Internationalized Systematized Nomenclature of Human and Veterinary Medicine* (SNOMED). It was originally developed for human conditions but includes also terms and concepts from veterinary medicine, and the nomenclature is easy to expand (Zimmerman et al. 2005). The database is being administered by the *International Health Terminology Standards Development Organisation* (IHTSDO).

The classification of diseases is not fixed in SNOMED. The SNOMED system is based on an Aristotelian classification method of species and genera, but in SNOMED one species can belong to several genera, which results in a non-fixed list. SNOMED therefore differs from the ICD classification, although there are several expert groups that work to connect these classifications (Nordenfelt 2013). For veterinary medicine and animal health, SNOMED might be a more complete classification of diseases than the OIE list. Still, work has to be done in adjusting the information in the system to better match veterinary medicine. Zimmerman et al. (2005, p. 7) found in an evaluation that the "SNOMED representation of veterinary clinical pathology content was limited, missing and problem concepts were confined to a relatively small area of terminology."

Similarly, for plant diseases, there is no complete classification of diseases although there are attempts made based on causal factors for the diseases (Agrios 2005; Kůdela 2011). For both animal and plant diseases, the classification efforts have been too limited to result in a deeper philosophical discussion on how to classify diseases and to demarcate a concept of disease.

Categories of Animal Health Definitions

Animal health has in recent years gained such interest that it is possible to present a list of categories of definitions of animal health. This following list is a result of scrutinizing the works of Gunnarsson (2006), Nordenfelt (2006), and Lerner (2008). Lerner and Berg (2015) argue that one should distinguish between definitions concerning the individual level and the population level of health when analyzing concepts of health. All categories presented below belong to the individual level of health. The population level of health will be dealt with later (in the section Population Level of Health).

Although the idea that health is the absence of disease is common, this negatively defined concept is excluded from the following list. Only positive, monistic definitions of health are presented, that is, definitions that actually define what health is and define this with some specific characteristic (see Tengland 2006). In those analyses of animal health definitions mentioned above, definitions that include more than one characteristic are called combined or conglomerate. In this chapter these two terms are regarded as synonymous and "conglomerate" is the term used for these. Conglomerate definitions are frequent within the science of animal health and welfare. But combining two or more characteristics might cause inconsistency within the concept (Lerner 2008). The list of categories for monistic definitions of animal health is the following:

1. Health as homeostasis
2. Health as normal biological function
3. Health as productivity including reproduction
4. Health as well-being
5. Health as mental and physical control
6. Health as ability to realize goals

Health as Homeostasis

A homeostasis view of animal health exists in homeopathic textbooks (Day 1995) and in the field of ecological farming (Vaarst and Alrøe 2012). This view characterizes health as a body and mind in equilibrium. An explicit definition of health from a homeopathic textbook states that

> "the word 'health' implies the concept of a *mind and body together in harmony with the environment*. When the organism, comprising the mind and body, is out of harmony within itself or with its environment, then we have the state of disease (literally dis-ease)" (Day 1995, p. 9, original emphasis).

The term "harmony" is not explicitly defined but is regarded as an internal equilibrium, and diseases are things that threaten this internal equilibrium (see the section "Health as Normal Biological Function" for further discussion on the term "harmony").

In ecological farming, the international organization IFOAM (International Federation of Organic Agricultural Movements) has developed four ethical principles. One of these is a principle of health. According to the principle, health is the same for soil, plants, animals, humans, and the planet. The principle itself lacks a definition of health. Vaarst and Alrøe explained this principle in terms of homeostasis stating that:

> A healthy organism is an organism or a system in homeostasis, meaning that it has the ability to withstand shocks and adjust or react to changing environments (Vaarst and Alrøe 2012, p. 339)

But other definitions of health could also fit the principle. Marley et al. (2010) interpret the principle regarding health in terms of the WHO definition for humans (see the section "Health as Well-Being" below).

Health as Normal Biological Function

One influential definition in the group of theories that define health in terms of normal biological function proposes that health is

> the state when the organs and the organ systems function in harmony with each other and with the environment. Disease is consequently a state when this harmony no longer exists. (Ekesbo 2011, p. 192)

Although very similar to a definition of health as homeostasis, the Ekesbo definition includes important differences. The foundation of this definition is not homeostasis but adaptation as one aspect of normal biological functioning. Ekesbo aims to use robust biological characteristics, namely, that the animal is functioning in a normal or natural way. Unfortunately, the vague term "harmony" was used as a key term in the definition. Although preferred by some theorists within animal health and welfare science, this term has not been explicitly defined (Lerner 2008).

Another definition referring to biological function belongs to Donald Broom:

> Health refers to what is happening in body systems, including those in the brain, which combat pathogens, tissue damage or physiological disorder. Health is the state of an individual as regards its attempts to cope with pathology. (Broom 2011, p. 133)

In an earlier elaboration on animal health, pathology is defined as "changes that are detrimental to the organism" (Broom and Kirkden 2004, p. 342). Changes may occur from the molecular level up to the tissue level and also include changes in the function of the organism. Health in animals is mainly within the physical sphere, and feelings are not covered by the concept of health. Although Broom mentions the brain in his definition, it is clear that the entire focus is on physiological disorders. The reason to exclude mental states or behavior is based on scientific tradition within the field of veterinary medicine, a discipline "primarily concerned with physical abnormalities" (Broom and Kirkden 2004, p. 344). Mental states and behavior are covered by other important concepts in Broom's theoretical framework. Mental problems, such as behavioral disturbances, fall under the wider concept of animal welfare rather than animal health.

An elaboration of Broom's definition of health gives two possible interpretations:

1. "An animal is healthy if, and only if, it is immune and, in general, reparative systems are in order.

. . .

2. An animal is healthy if, and only if, the immune and reparative systems actually succeed in eliminating pathology, i.e., if the animal as a result is without disease

(induced by pathogens), tissue damage or physiological disorder" (Nordenfelt 2006, p. 49).

In the first interpretation the animal is healthy as long as the immune and reparative system works, although an infection might be present. In the second interpretation, the animal is healthy as a result of the immune and reparative process. According to Nordenfelt, the second interpretation, equating health with the absence of disease, makes it similar to Christopher Boorse's definition of human (and animal) health as normal functioning (see Boorse 1997).

Health as Productivity Including Reproduction

In this category definitions based on the animal's production or reproduction are clustered. The ideas of productivity and reproduction include increased biomass, production of milk, or number of offspring. All these three factors are important for the animal itself but could also be important for humans having the animals in custody. In farming an increase in all these three aspects yields a higher income for the farm. This category of definition had therefore its strongest impact when farming became industrialized. An example of a definition of health as productivity is health as "a state of maximum economic production" (Boden 1998, p. 243). Such definitions are today discarded. Industrialized farm settings that aim for maximum of economic production tend to have diseases originating from the high production rate. A strong normative claim within the science of animal health and welfare is that the definition of health should apply to the animal itself and not to human interests (Lerner 2008). Still, production and reproduction could be used in conglomerate definitions.

Health as Well-Being

When it comes to health as well-being, there has been a strong influence from the WHO definition of health for humans, stating that health is "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (WHO 1948). Kelly (2000) defines animal health in similar words but excludes mental well-being. The lack of mental well-being in this definition is striking and might be an effect of the behavioristic heritage within the field. Unfortunately, the text lacks further analysis of the definition as well as a reference to the WHO definition.

In a later discussion on organic agriculture, the complete WHO definition is used as the definition of animal health (Hovi et al. 2004; Marley et al. 2010). According to Hovi et al., this kind of definition is more fruitful to cover all aspects of health, and it also facilitates the practical planning on the farm in written health plans. The WHO definition is also used in discussions concerning laboratory animals (King and Rowan 2005) and could be used on the population level of health (see that section below).

Health as Mental and Physical Control

In an interview study, an expert Swedish veterinarian tried to define animal health as mental and physical control, namely, health is regarded as:

> the animal's ability to have control of its situation, with regard to both coping systems and the forestalling, wherever possible, of the need for coping. (Lerner 2008, p. 90)

A central aspect of this definition is coping, which is similar to Broom's definition of health (see the section "Health as Normal Biological Function"). The difference in relation to Broom's theory of coping is the idea of forecasting what will happen in the future or in other words the ability to have enough control of a situation that the animal might alter it. The animal has a mental ability to imagine a situation which could be dangerous or unpleasant and change the situation by avoiding or hindering its occurrence. The animal is therefore able to adjust to different situations or changes in the environment.

In Georges Canguilhem's theory of human health, there is a distinction between being normal and being normative. Being normative is the healthy state where one is both able to live in and change one's present environment and being able to change from one environment to other environments. The normal means being confined to live only in one environment, because one cannot change to other environments without facing problems. Typical for the latter case is a person who has a disease but still fares well in a well-adapted environment for that disease (Canguilhem 1978).

According to Lerner's (2008) expert veterinarian, the animal that controls its situation and succeeds in forestalling and adjusting is normative in the sense of Canguilhem. Merely managing to cope (as in Broom's definition) is to be normal in the sense of Canguilhem.

Health as Ability to Realize Goals

Nordenfelt (2006, 2007) has introduced his concept for human health within the field of animal health. He defines animal health in two slightly different ways. The first version belongs to a thorough attempt to analyze the field of animal health, while the second version is an attempt to include all living species.

> A is completely healthy if, and only if, A is in a bodily and mental state which is such that A has the second-order ability to realize all his or her vital goals given a set of standard or otherwise reasonable conditions. (Nordenfelt 2006, p. 147)
>
> An animal A or a plant P is healthy if, and only if, A or P has the ability to realize all its vital goals given standard circumstances. (Nordenfelt 2007, pp. 30–31)

Central to both versions is a reference to all the individual's vital goals. When it comes to humans and animals with certain mental abilities, a vital goal is something

external or internal that could contribute to the individual's long-term happiness. These goals could be basic needs, such as getting water and food, as well as more sophisticated ones, such as learning. In other animals and plants, where these mental abilities are absent, vital goals are states which contribute to vitality. For plants vitality is being able to develop or grow.

Unfortunately, something central might be lost when changing the idea of vital goals from a mental aspect to an aspect of vitality. The definition has been influential in the philosophy of medicine because it focuses on other things besides mere biological function, such as a long-term happiness which is normative and individual. Limiting vital goals to vitality and defining it in terms of development and growth might lose the normative aspect crucial to differentiating it from a health as normal biological function definition.

Plant Health Definitions

Plant health has received even less treatment than animal health. Besides Nordenfelt's attempt, the recent discussions of plant health which have emerged are influenced either by the discussion of human health (Döring et al. 2012) or by Gunnarson's categorization of animal health (Kůdela 2011). Both these analyses also compare concepts developed within plant science, but neither of them proposes a definition of health.

Central to plants is that in their case the concept of mental health is irrelevant. This is due to the lack of a nervous system. One example of a definition of plant health referred to in both papers mentioned above is George N Agrios's:

> A plant is healthy, or normal, when it can carry out its physiological functions to the best of its genetic potential. (Agrios 2005, p. 5)

One particular distinctive aspect of plant health compared to animal health is that for plants the focus is on disease prevention rather than cure. If a plant is diseased, one eradicates it in order to prevent other plants from becoming diseased.

Population Level of Health

As mentioned earlier, one can consider animal and plant health on the population level. For domestic animals, "herd health" is the common term (Lerner and Berg 2015). Herd health is mainly used for monitoring purposes and is based on statistical measurements of disease frequencies in populations. Definitions of health that can be applied for herd health are either a homeostasis definition of health (Vaarst and Alrøe 2012) or the WHO definition for human health (Blaha 2005) which belongs to the category of health as well-being.

For wildlife, the term "wildlife health" is used. Wildlife health has been more and more interpreted in terms of resilience. Resilience can be seen as sustainability of a population. Sustainability depends on the population's ability to cope with changes in the environment. The population's ability to cope is the "result of interacting biologic, social, and environmental determinants that promote and maintain health" (Stephen 2014, p. 429).

Resilience has also been proposed in a wider context to be a universal criterion for health (Döring et al. 2015). Resilience could be applicable as a criterion for health on individual, population, and ecosystem level, for plants, animals, and humans. According to Döring et al., resilience differs somewhat from the concept of homeostasis. The concept of resilience emphasizes the successful response to a disturbance and the recovery process, where homeostasis emphasizes a more static equilibrium.

Plant health is both defined on an individual and a population level (Döring et al. 2012), and in some cases, a proper distinction between the individual and the population level is troublesome. For those plant species that reproduce nonsexually by fragmenting, the distinction between population and individual is blurred. Individuals that are formed by nonsexually fragmenting share similar DNA but are separated physically from each other.

Universal Concepts of Health

In the literature one finds two approaches to defining a universal concept of health. The first is to find a definition that applies universally to animals and humans. The second is to find a definition that applies to plants, animals, and humans. The former seems easier due to the similarity between many animals and humans, but the animal kingdom is diverse and there are huge differences in the mental constitution when one compares, for example, insects and mammals. To be able to find a common ground for a definition of health, one has to analyze similarities and differences between different groups of species.

Nordenfelt (2006), in his comparison between animals and humans, finds three similarities between definitions of animal and human health:

1. "In both areas [there are] theories of biological, natural or normal, function
2. In both areas [there are] ideas of biological and psychological balance or homeostasis
3. In both areas [there are] ideas focusing on well-being" (Nordenfelt 2006, p. 139)

The first similarity, the category of health as normal biological function, includes definitions such as Broom's idea of coping. Broom also claims that his definition is applicable to all animals. Other options from the field of human health could possibly be adapted to animal health. The nineteenth-century physician and public health pioneer Rudolf Virchow defined health in terms of vital cells. Health persists

as long as the number of non-vital cells is fewer than the vital (Virchow 1881). Cells are similar all through the animal kingdom. Christopher Boorse's biostatistical theory of health is applicable to humans, animals, and also plants because it focuses on normal functioning, which could be further defined for different kingdoms (Boorse 1997).

The second similarity, the category of health as homeostasis, is applicable to both animals and humans.

The third similarity is mainly an adaption of the WHO definition for humans to the field of animal health and falls within the category health as well-being.

There were also three differences in Nordenfelt's analysis:

1. "Health is regarded as a controversial concept (almost) only in the human discussion
2. In the human health discussion there are a multitude of psychosocial concepts of health, emphasizing the healthy person's ability to realize goals in society
3. Theories relating health to production occur only in the animal discussion" (Nordenfelt 2006, p. 139)

The first difference is that naturalistic positions seem to be more common than normativist positions in the nonhuman area. Bernard Rollin claims that definitions made by veterinarians are more reductionistic and mechanistic than those made by physicians (Rollin 1983), implying a naturalistic position. Later, Donald Broom argued that an assessment of an animal's welfare including health "should be carried out in an objective way" (Broom 2001, p. 4). Others, like Nordenfelt, state that the definition of health in animals still must be approached from a normative standpoint (Nordenfelt 2006).

The second difference seems to be a key issue for the separation. Mental health in animals is still somewhat questioned as being a part of the concept of health (Broom and Kirkden 2004). The reasons are mainly due to the behavioristic heritage in the research field of animal behavior. At present day more and more knowledge is gained about the mental abilities of animals, and mental health is truly possible to be ascribed to at least parts of the animal kingdom. Nordenfelt claims that his idea of health is applicable both to humans and animals. Still, definitions based on certain kinds of mental abilities such as self-consciousness might limit the universality of the health concept.

The third difference found might not be sufficiently convincing. In the philosophy of medicine, Martha Nussbaum (2011) states in her list of capabilities that health for humans also includes reproductive health. Still, one seldom talks about health as production in humans, and as shown above, health defined purely in production or reproduction is troublesome. Rather, production and reproduction might be an indicator of normal physiology.

A universal definition of health for plants, animals, and humans is also possible. Somatic definitions such as health as normal biological functioning or health as homeostasis or resilience could be applicable to all three groups. Also, Nordenfelt

(2007) has tried to extend his definition of health as the ability to fulfill vital goals to cover plants, animals, and humans.

Definition of Key Terms

Behaviorism	The idea that behavior can be explained in stimulus–response patterns, where a stimulus always results in the same response. Behavior could then be explained without referring to cognitive faculties.
Ethology	The scientific field for the study of animal behavior.
Naturalistic definition of health	A definition without normative criteria.
Normative definition of health	A definition with normative criteria.

Summary Points

- There exists a scientific discussion on how to define animal as well as plant health.
- Compared to the philosophical discussion of human health, the philosophical discussion of animal and plant health is underdeveloped.
- The philosophy of human health has influenced both the field of animal health and plant health.
- For animal health at least six categories of health definitions have been presented in the literature.
- There are definitions that might be universal and applicable to plants, animals, and humans.

References

Agrios GN (2005) Plant pathology, 5th edn. Elsevier Academic Press, San Diego

Bermond B (1997) The myth of animal suffering. In: Dol M, Kasanmoentalib S, Lijmbach S et al (eds) Animal consciousness and animal ethics: perspectives from the Netherlands. Van Gorcum, Assen, The Netherlands, pp 125–143

Blaha T (2005) Tiergesundheit und tiergesundheitsfürsorge im wandel der zeit. Dtsch tierarztl Wschr 112:284–285

Boden E (ed) (1998) Black's veterinary dictionary, 19th edn. A & C Black, London

Boorse C (1997) A rebuttal on health. In: Humber J, Almeder R (eds) What is disease? Humana Press, Totowa, pp 3–134

Broom DM (2001) Coping, stress, and welfare. In: Broom DM (ed) Coping with challenge: welfare in animals including humans. Dahlem University Press, Berlin, pp 1–9

Broom DM (2011) A history of animal welfare science. Acta Biotheor 59(2):121–137
Broom DM, Kirkden RD (2004) Welfare, stress, behavior, and pathophysiology. In: Dunlop RH, Malbert C-H (eds) Veterinary pathophysiology. Blackwell, Ames, Iowa, pp 337–369
Canguilhem G (1978) On the normal and the pathological. D. Reidel Publishing Company, Dordrecht/Boston/London
Dawkins MS (1980) Animal suffering: the science of animal welfare. Chapman and Hall, London
Day C (1995) The homoeopathic treatment of beef and dairy cattle. Beaconsfield Publishers, Beaconsfield
Döring TF, Pautasso M, Finckh MR et al (2012) Concepts of plant health: reviewing and challenging the foundations of plant protection. Plant Pathol 61:1–15
Döring TF, Vieweger A, Pautasso M et al (2015) Resilience as a universal criterion of health. J Sci Food Agric 95:455–465
Ekesbo I (2011) Farm animal behaviour: characteristics for assessment of health and welfare. CAB International, Wallingford
Gunnarsson S (2006) The conceptualisation of health and disease in veterinary medicine. Acta Vet Scand 48:20. doi:10.1186/1751-0147-48-20
Hovi M, Gray D, Vaarst M et al (2004) Promoting health and welfare through planning. In: Vaarst M, Roderick S, Lund V et al (eds) Animal health and welfare in organic agriculture. CABI Publishing, Wallingford, pp 253–277
Irwin TH (2003) Aristotle. In: Craig E (ed) Routledge encyclopedia of philosophy. Routledge, London. http://www.rep.routledge.com/article/A022SECT17. Assessed 19 Dec 2012
Kelly J (2000) Nutrition of the dairy cow. In: Andrews AH (ed) The health of dairy cattle. Blackwell Science, Oxford, pp 49–88
King L, Rowan AN (2005) The mental health of laboratory animals. In: McMillan FD (ed) Mental health and well-being in animals. Blackwell Publishing, Ames, Iowa, pp 259–276
Kůdela V (2011) On the need for revision of some names of plant health malfunctions and their categorisation. Plant Protect Sci 47(4):133–148
Lerner H (2008) The concepts of health, well-being and welfare as applied to animals: a philosophical analysis of the concepts with regard to the differences between animals. Ph.D. dissertation, Linköping University, Linköping
Lerner H, Berg C (2015) The concept of health in One Health and some practical implications for research and education: what is One Health? Infect Ecol Epidemiol 5:25300, 10.3402/iee.v5.25300
Marley CL, Weller RF, Neale M et al (2010) Aligning health and welfare principles and practice in organic dairy systems: a review. Animal 4(2):259–271
McGlone JJ (1993) What is animal welfare? J Agric Environ Ethics 6(Special suppl 2):26–36
McMillan FD (ed) (2005) Mental health and well-being in animals. Blackwell Publishing, Ames, Iowa
Nordenfelt L (2006) Animal and human health and welfare: a comparative philosophical analysis. CABI, Wallingford
Nordenfelt L (2007) Holistic theories of health as applicable to non-human living beings. In: Kincaid H, McKitrick J (eds) Establishing medical reality: essays in the metaphysics and epistemology of biomedical science. Springer, Dordrecht, pp 23–34
Nordenfelt L (2013) Identification and classification of diseases: fundamental problems in medical ontology and epistemology. Studia Philos Estonica 6(2):6–21
Nussbaum MC (2011) Creating capabilities: the human development approach. The Belknap Press of Harvard University Press, Cambridge, MA/London
OIE (2015) OIE-listed diseases, infections and infestations in force in 2015. http://www.oie.int/animal-health-in-the-world/oie-listed-diseases-2015/. Accessed 27 Dec 2015
Rollin BE (1983) The concept of illness in veterinary medicine. JAVMA 182(2):122–125
Rollin BE (1998) The unheeded cry: animal consciousness, animal pain, and science, Expanded edn. Lightning Source, Milton Keynes
Stephen C (2014) Toward a modernized definition of wildlife health. J Wildl Dis 50(3):427–430

Tengland P-A (2006) The goals of health work: quality of life, health and welfare. Med Health Care Philos 9(2):155–167
Vaarst M, Alrøe HF (2012) Concepts of animal health and welfare in organic livestock systems. J Agr Environ Ethics 25:333–347
Virchow R (1881) An address on the value of pathological experiments. Br Med J 2(1075):198–203
WHO (1948) Official records of the world health organization, no. 2. WHO, Geneva, p 100
Zimmerman KL, Wilcke JR, Robertson JL et al (2005) SNOMED representation of explanatory knowledge in veterinary clinical pathology. Vet Clin Pathol 34(1):7–16

Genetic Information in Medicine: Its Generation, Significance, and Use

21

Angus Clarke

Contents

Abstract

"Genetic information" may refer to information about a person's family history, raw DNA sequence data, or an interpretation derived from the raw data. This chapter addresses what counts as genetic information, with a focus especially on genetic information about humans and the limitations on what can be known.

Family history provides information relating to the risk of specific disease without requiring any DNA sequence data. Another type of indirect genetic information concerns the heritability of quantitative traits and complex disorders; this can demonstrate that genetic factors are involved without specifying precisely what they are.

A. Clarke (✉)
Institute of Medical Genetics Building, Cardiff University School of Medicine, Cardiff, UK
e-mail: clarkeaj@cardiff.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_13

Genetic linkage studies provide information about the co-localization of a disease-related gene and a nearby marker on a chromosome. Tracking the marker can be used to infer whether someone has inherited the corresponding disorder. In contrast, association studies are performed on large populations and identify many dispersed genetic factors that, jointly and through interactions with the environment, influence the chance of developing the common complex degenerative disorders of Western society. Confidence in the accuracy of individual predictions based on linkage studies is often very high, while genetic association studies provide information that applies robustly to a population but not so readily to making predictions about specific individuals.

The distinction between "raw data" and "interpretation" is unclear and fuzzy. With the high-throughput methods now in use, it is not possible to generate sequence information about a person's genome – or even a single gene – without an interpretive step based upon sequence information generated earlier from other individuals, which has been accumulated as a body of knowledge. Interpretation by comparison with previous findings will leave some variants as being of uncertain significance (VUSs). This naturally raises the question of future reanalyses and reinterpretations of such VUSs or genome sequence data in general. In addition, unsought information may be found concerning risks of other diseases or of misattributed relationships within a family.

Information concerning many rare disorders has accumulated through a back-and-forth flow of questions and answers between patients/families and clinicians/scientists, to the benefit of both parties.

Ownership of genetic information about individuals' DNA sequence is probably an unhelpful concept, but control of the uses to be made of such information will be important.

Many ethical questions are raised by developments in DNA sequencing, as it has led to an explosion in the generation of sequence information. These include the information that should be given to a patient about their genome sequence, what information the patient or family should then disclose to relatives, and whether there is an obligation to find out all available information about one's own genome and then perhaps use it to shape the genome of one's children.

Introduction

"Genetic information" conveys very different meanings in different circumstances. Before deoxyribonucleic acid (DNA) had been recognized as the molecule of heredity, the detailed dissection of gene action was achieved by painstaking experiments, especially breeding experiments, that described the consequences of many genetic phenomena. This formal, classical, or pre-molecular genetics defined the phenomena of heredity, development, and gene action many decades ago in *Drosophila* and other model organisms that molecular genetics is still working to describe and explain. "Genetic information" may refer to the raw DNA sequence

of a species: a flowering plant, a higher primate, or a bacterium. Or it may refer to claims about the common evolutionary origin of two species from a predecessor or the common origin in the evolutionary past of the sequences of two now distinct genes within a species. In either of those cases, the "raw" sequence data has undergone a step of analysis – alignment and then comparison – to permit an interpretation of an evolutionary relationship. These interpretations relate to the genetic structure and evolution of one or more species, but more wide-ranging interpretations are also commonly sought on the basis of DNA sequence information. The interpretation may relate to physical differences between or among individuals and groups or to the occurrence of disease. In these interpretations, relating DNA sequence to physical or behavioral traits or to disease, there is much more at stake than evolutionary theory because there are implications of the findings and interpretations for the health care of individuals, for the justice of social organization, and for the shape of public policy.

An additional type of genetic information arises from the search for patterns in genomic information, such as the effect of sequence context on genome function (including gene expression) or the tendency to mutation. Such second-order information may be termed meta-information, but this will not be considered further here (Evans and Foster 2011).

The raw data that is drawn upon to warrant functional interpretations of the genetic constitution of humans is, in effect, a long string of the chemical components of the 92 DNA molecules that constitute the usual diploid genome of the human (with two copies of each type of chromosome). There are 23 pairs of chromosomes, 46 in all, each of which consists of two long DNA molecules oriented in opposite directions and associated very closely together, hence 92 molecules in all. The two molecules in each chromosome are held together by the hydrogen bonds that allow pairing between the nucleoside bases, with a choice of four "options" at each of the $\sim 10^9$ sites in the haploid genome. The choice of base at each site on one of the two strands in each chromosome (the more newly synthesized strand) is completely determined by the choice of base on its pairing partner on the other, older strand. The simplest and most effective way of expressing the genome sequence of an individual is to give the sequence of bases along one of the two strands of DNA in each chromosome. Even with these constraints, and ignoring the possibility of variation in length (by the insertion or deletion of bases), there is an astronomical number of potential human genome sequences, in that there are 10 to the power of 9 base pairs in one copy of the genome with a choice of four bases at each of these sites, so 4 to the power of 10 to the power of 9 potential "versions" of a haploid genome (as in a gamete) or the square of that in the diploid genome of an individual. This string of bases – specifying which of the four possibilities (A, C, G, or T) is present at each of the 2×10^9 sites – is the genome sequence of an individual human.

Raw DNA sequence data can be generated in several ways but two approaches need to be sketched here, the "conventional" approach known as Sanger sequencing and the newer, more powerful, approach termed "high-throughput" or "next-generation" sequencing.

DNA Sequence Data

Two fundamentally different approaches have been taken to the generation of DNA sequence information. One approach is targeted to a precise area of the genome and entails selecting the stretch of chromosomal DNA whose sequence is to be determined. The other entails the generation of sequence of whatever DNA sample is being examined; the location of this source of DNA (e.g., the gene or chromosome on which it is located in the DNA source) has to be determined after the sequence has been generated.

The conventional, long-established Sanger method of DNA sequencing selects segments of DNA for amplification, and the sequence of these amplified fragments is then determined. This approach is still widely used in diagnostic laboratories, either as the primary method or in confirmation of less trusted findings from newer methods. The newer, high-throughput approaches to DNA sequencing rely upon the sequencing of DNA fragments first and then, subsequently, their assignment to a particular gene or chromosome, or even species, on the basis of the sequence information itself. This "data first" approach, leaving to later the assignment of a sequence to a chromosomal location, was developed as "shotgun sequencing" by the commercial challenger to the publicly funded Human Genome Project (HGP) and has since become the dominant approach in large-volume sequence generation. The assignment of a specific sequence to a site within a gene or on a particular chromosome depends upon the already available human genome sequence; genome sequencing (or resequencing) today is much more straightforward because only the sequences that differ from the reference human genome sequence, established as the usual version at each base pair, will require interpretation. The reference genome sequence is not found in any one individual but is an idealized "average" at each base.

Genetic information provides information about an individual's genetic constitution and how this influences their current or future health. It may also give information about the likely response of an individual to different drug treatments as well as relationships or personal identity, ancestry (population of origin), and a number of non-disease traits. The analysis of biopsies from a tumor, or samples of free DNA in the patient's plasma, can give information about the mutations present in a malignancy and how best to treat it.

One suggestion is that infants could have their genomes sequenced at birth and then this information could be kept as a resource throughout their lives. Difficulties with that scenario include likely developments in genome sequencing, information technology, and bioinformatics. Thus, the quality (coverage and depth) of sequencing is improving and may come routinely to include not merely the DNA sequence but also information about the functional status of the genome in a range of tissues, as indicated by methylation and other markers of chromatin conformation and gene expression. Today's genome sequence may soon be regarded as inadequate as a guide to the influence of genetic factors on the body's growth, development, and disease: additional information will come to be regarded as necessary. Furthermore, both the information technology (IT) software and hardware systems used to

analyze, interpret, and store the genome data are evolving, while the updating and transfer of data onto new systems becomes costly. It may make better sense to analyze a genome when the information is required and then to store the report and conclusions but discard the data (Chadwick et al. 2013).

"Genetic information" can be understood as either raw data or its coherent interpretation. However, this distinction between "raw (sequence) data" and "interpretation" is unclear and fuzzy: it is unsustainable. As we will see below, it is not possible to use the new, high-throughput methods to generate sequence information about a person's genome – or even a single gene – without an interpretive step that employs reference to sequence information generated earlier, which has been accumulated as a body of prior knowledge.

Family History

Still one of the best predictors of disease risk is a person's family history of disease. An individual's knowledge of their family history, however, is often fragmentary and – once beyond the immediate family – a matter of anecdote or hearsay. Confirmation of disease details may be available to the health professional through a disease registry or through access to the medical records of other family members, including reports of an examination conducted *post-mortem* or laboratory confirmation carried out on specimens, such as a tumor biopsy or an organ stored from examination *post-mortem* or a newborn screening bloodspot. However, the confirmation of a diagnosis is often difficult or impossible, especially if a death occurred more than a decade ago. Information circulated within a family is often insufficiently precise or even erroneous: thus, someone said to have had Parkinson's disease may have had Huntington's disease, and a report of "liver cancer," "stomach cancer," or "brain cancer" may well not indicate the site of the primary tumor.

Compounding the inadequate oral history of diagnoses may be disagreement or ignorance within a family about the details of relationships. This may be less of a problem in Scandinavia, where excellent genealogical records are often available in parish churches, and in Iceland the relationships between many individuals are known to the researchers in the genomics venture company deCode (through their genome sequencing of many Icelanders) even if they are unknown to the individuals themselves. Most societies, however, are not "blessed" by such good records of the past several centuries or by such extensive genome sequencing of the population alive today. The fact that multiple versions of a family tree may exist within the "family records" in a clinical department, differing slightly depending upon who contributed the information, may of course present challenges to the clinicians, who will usually wish to insulate these differing versions so as to avoid the inappropriate disclosure of information about one family member to another, such as abortions, misattributed paternity, and adoptions, either into or out of the family. The clinical team may in fact hold a more accurate, composite version of the family tree than is available to any one family member, especially if it incorporates the results of

genetic testing. This composite family tree may also need to be treated as confidential material.

Another factor that can minimize the value of family history is a tendency for a family to be of small size – with few individuals, perhaps only one or two siblings within the nuclear family per generation for several generations. Members of such a family are much less likely to be aware of an inherited risk of disease, if there is such a risk, than in families with large sibships. Citizens of China, in which the one-child policy of 40 years has only just become a two-child policy, are much less likely than others to have aunts, uncles, or cousins, so that family history is less likely in China than elsewhere to give warning of potential disease risks.

Family history therefore counts as a form of genetic information but its basis is of varied quality; its applicability depends upon the assessment of its quality.

Heritability of Complex Traits and Diseases

A disorder inherited through a family over generations will usually be clearly genetic in origin, although one needs some caution in drawing conclusions. Recall that surnames (in the Western system of naming) and wealth (in the traditional Western system of the inheritance of land and titles) are also transmitted from one generation to the next through a family. Language, culture, cuisine, and even nutritional deficiencies are also often "familial."

In contrast to the strong, single (Mendelian) genes of major effect are the multiple genetic factors of small effect, which interact with the environment and influence the occurrence of quantitative traits (such as height) and the common, complex disorders (such as type II diabetes mellitus, stroke, and coronary artery disease). Examining the degree of similarity between close relatives for these quantitative traits and complex disorders does not identify the specific genes or genetic variants involved but demonstrates the degree to which genetic factors, as opposed to environmental factors, influence the phenotype. The proportion of the variation in a trait that can be attributed to variation in genetic factors, observed in that population and under those circumstances, is known as the heritability for that characteristic.

The value of heritability, ranging from 0 % to 100 %, varies for the same trait between populations and can also change over time as circumstances change: the heritability of a trait is not a fixed and permanent value. Furthermore, a high heritability does not mean that the feature of interest will not change in response to a change in the environment. This is often misunderstood, with potentially important political consequences.

Recent studies in the complex disorders have attempted to identify the genetic variants that account for the high heritability (often around 50 %) of many complex disorders, but they have mostly been unsuccessful in identifying the factors that account for more than a modest fraction of the heritability. Much of this "missing heritability" is most likely accounted for by gene-gene and gene-environment interactions, which are so difficult to study in humans for ethical and practical

reasons. However, these methods allow us to assess the extent to which genetic variation impacts on the phenotype.

Linkage and Linkage Disequilibrium

The assignment of (Mendelian) disease genes to a specific chromosome began with the recognition of the X chromosome and sex-linked inheritance: once a disorder was recognized as following sex-linked inheritance, it could at once be assigned to the X chromosome. Very few other assignments of human traits or diseases to chromosomes were achieved until the 1980s. From then on, genetic linkage studies were conducted in order to localize and then isolate disease-related genes. They utilized sites of genetic variation (usually both common and benign) that were spread throughout the genome; these were used to track disease genes by looking for the consistent co-inheritance of the marker (the site of known genetic variation) and the specific disease of interest. This then led, over two decades, to the generation of a map of the chromosomal location of many disease genes and the markers with which they were co-inherited, thereby permitting the assignment of specific traits or diseases to chromosomes and then the determination of the linear sequence of these sites along a chromosome. This amounted to a high-level, low-resolution gene map across all chromosomes and was an important form of modern genetic information. The polymorphic variants employed in these studies were not usually of intrinsic significance to the disease in question except in helping to locate the gene. This knowledge of abstract genetic information – the loci A,B,C, and D are in fact located in the sequence B,D,A,C along a specific chromosome – was of major importance in the early stages of the project to sequence the human genome, the Human Genome Project (HGP), as it formed the backbone of landmarks. Filling in the gaps between the frameworks of loci was the task of the HGP, i.e., the high-resolution detail of the DNA sequence.

The clinical application of such gene linkage relationships has been largely replaced by the ability to sequence genes and genomes directly without the need to infer the presence or absence of disease-causing mutations through the study of polymorphic markers. The circumstances where linkage methods are still applied include preimplantation genetic diagnosis, where linkage methods are more reliable than the detection of specific sequence variants, and attempts to determine the pathogenicity of molecular variants of uncertain significance. In such applications, the chance of the result being misleading is usually known precisely; it is the chance of a recombination event occurring between the trait and the marker(s) being used to track it. When close flanking markers are employed – i.e., when sites of benign genetic variation are tracked immediately adjacent to, and either side of, the gene of interest – the chance of a double recombination event occurring will usually be very small; this is the chance of two separate recombination events occurring, one either side of the gene, between the gene and the marker on that side of the gene. If a double recombination event has occurred, then studying the markers would give a misleading result. Thus, if the markers both suggested that the risk to the individual was low

but in fact a double recombination event had occurred, the copy of the gene that had been transmitted to the individual would nevertheless be high risk.

Related to genetic linkage analysis is the concept of linkage disequilibrium. This was of historic significance in the recognition of Mendelian disease genes, as in the isolation of the "gene for" cystic fibrosis, officially the *CFTR* gene, which is implicated in that disease process. When a polymorphic marker is found in close linkage with a disease gene, there will sometimes be a nonrandom association between a specific allele (variant) at the polymorphic site and the presence of disease. This suggests that there has been a founder effect with one or a few original mutation events leading to the disease, which occurred as specific historic events in the (relatively) recent evolutionary past. With a Mendelian gene, finding linkage disequilibrium indicates that the polymorphism involved is indeed very close to the variation responsible for disease.

Mendelian Disease Genes and Complex Causation

Such analyses are no longer required for the mapping of Mendelian disorders, but a similar type of analysis is employed in the recognition of associations between complex disease entities and the polymorphic variation dispersed across the genome that contributes to the risk of such disorders. This requires an aside to explain the distinction between a Mendelian disease and a complex disorder. A Mendelian disease gene is one that tracks through a family as a powerful factor that, in effect, causes the disease. The gene may act in dominant or recessive fashion, when either a single altered copy of the gene will be sufficient to "cause" the disease (in dominant inheritance) or when disease arises only when both copies of the gene are altered (in recessive inheritance), so that many people may "carry" a gene alteration without any ill effect, because only a single intact copy of the gene is required to avoid disease under normal circumstances. However, for these conditions, it is the presence or absence of the one or two gene alterations (mutations) that effectively determines the presence of disease. This description is somewhat overstated, in that other factors may also be necessary for disease to arise. This may not always occur, although the presence of the altered copy or copies of the gene may be the predominant factor determining disease or risk of disease. The additional necessary factors may be the occurrence of a second mutation on the other copy of the gene (in some strong cancer predispositions) or an environmental factor (e.g., exposure to "normal," unrestricted, protein intake in addition to having two altered copies of the disease gene, as in the recessive disorder phenylketonuria). However, the genetic variants are so strongly associated with the disease that they are said more or less to "cause" it, in a shorthand version of the truth. Notwithstanding David Hume's distinction between association and causation, this manner of speaking is widespread and serves the biomedical community well, with the subtle reservations being acknowledged when precision demands it. Penetrance, the chance that the phenotype will manifest in the presence of the relevant genetic factor, is often less than 100 %, especially in late-onset disorders and most particularly in cancers, where other

mutations may need to occur for a malignancy to result. Chance can be an intrinsic component of the development of the phenotype.

In contrast to a high-penetrance, single-gene (Mendelian) disorder, a complex disease does not have one single, simple, genetic cause but, rather, is caused by many interacting factors including genetic constitution, environmental factors, life history events, and sometimes pure chance. Examples of complex disorders include the common cancers, coronary artery disease, Alzheimer's disease, hypertension, stroke, and diabetes mellitus types 1 and 2. While there are some uncommon genetic variants that act in a strongly causal fashion – effectively single gene causes of these conditions, concealed within the large mass of each disease type – these Mendelian causes of complex disease account for only a small proportion of these conditions overall. The vast mass of genetic factors influencing the risk of these complex disorders will each contribute only weakly to the occurrence of disease.

The tests of association between the complex trait or disease and polymorphic variants are usually based on genome-wide association studies (GWAS). These assess a very large number of widely dispersed markers, most often single nucleotide polymorphisms (SNPs), but do not entail examining for the rare variants that arise in Mendelian genes as strong influences on disease occurrence.

Such associations between multiple genetic factors and disease may be highly significant statistically, but the association may still be modest or weak, usually giving a relative risk of perhaps 1: 1.01 or 1:1.02. It is the cumulative effect of many such weak influences that sums up to become the genetic contribution to disease causation, sometimes referred to as the *heritability* of the disorder (the fraction of the overall variation between individuals in predisposition to the disease that can be said to be genetic). Estimation of disease risk using the typing of the alleles at a large number of polymorphic sites is sometimes made available commercially as a direct-to-consumer (DTC) test, although such tests have usually not been validated independently and have little if any clinical utility, so that they are not offered by state-funded or insurance-funded health-care systems. The action to be recommended on the basis of a modestly increased risk of a disease is likely to be compliance with standard advice for a healthy lifestyle, including the monitoring of blood pressure and perhaps serum cholesterol, as would be recommended to the general public. The worth of these tests is contested because of the limits on what they can establish about any specific individual's risks of disease.

An additional reason for caution in applying the results of risk estimates established in this way is that the rules for combination of results at different loci are assumed for the purposes of these calculations to be simple (i.e., multiplicative, as if independent), whereas work on simpler organisms suggests that the variants at many of the sites may interact in a bewilderingly complex fashion yet to be dissected in humans (Wray et al. 2013; Ritchie et al. 2015). A third reason for caution is the risk that someone with a strong family history of disease – such as breast cancer or colorectal cancer – may be misled by a reassuring DTC GWAS-based risk result, being given false reassurance by a normal or low risk of disease estimated in this way. This might lead them to decide not to seek a formal assessment of their family history, so that the opportunity for genetic testing of relevant disease genes of major

effect in an affected relative would be lost. If such a strong genetic factor was identified, there may be opportunities to prevent the cancer or undertake screening for early detection and improved treatments.

Genetic linkage studies provide indirect but often robust information about the genetic constitution of a species, a population, or an individual through a process of inference. The level of confidence in predictions based on such inference is known and is often very high. In contrast, while genetic association studies can also provide robust information about a population, the use of these studies as the basis for predictions about the future occurrence of disease in individuals is unwarranted.

Recognizing Mutation: Interpreting Variants

When sequence information is generated, it is compared with the "standard" reference genome sequences with the use of sophisticated IT systems. The development and refinement of systems for such analyses have created a new discipline known as bioinformatics. Where the patient's DNA sequence matches the reference sequence, or at least one of the many common, benign variants, there is no problem: that stretch of DNA can be assigned to a specific, already-known site in the human genome. This may be in a protein-coding area of a gene, a transcribed but untranslated section of the gene that nevertheless forms part of the messenger RNA, an intronic area within a gene, a regulatory region within or around a gene, or an intergenic (i.e., extragenic) region with (usually) no recognized function.

Where there are discrepancies between the patient's sequence and the reference version (i.e., where a variant is identified), these will be compared with other variants previously reported. If a variant is known to occur regularly, if infrequently, within a population and has no association with disease, this will be regarded as a benign variant, a polymorphism. If it has previously been reported in association with disease, and if a plausible mechanism can be seen for how the variant might cause disease, then it will be regarded as disease causing. In between these extremes of clarity lies a whole range of uncertainties with differing levels of confidence in the interpretation of the sequence information. If a variant alters the amino acid sequence of the corresponding protein, especially in an important functional domain of the protein or at a highly conserved amino acid (in evolutionary terms), then it will be regarded as likely pathogenic, although additional checks may be made to assess the plausibility of this attribution such as examining the parents clinically and taking parental samples, if they are available, to determine whether the variant has been inherited or has arisen de novo.

Another step that could in principle contribute to a decision about the clinical impact (often, the pathogenicity) of the sample is functional analysis of the variant. It will sometimes be helpful to assess the influence of the variant on gene expression, on RNA splicing, or on the function of the altered protein. However, this type of assessment would often amount to a research project, and resource constraints do not usually permit a diagnostic laboratory to pursue such avenues.

Assignment of a sequence variant to the class of disease-causing mutations, benign polymorphism, or an intermediate, less certain category entails a complex process of data interpretation that draws on the accumulated knowledge of DNA sequences from many other individuals, including sometimes other species. Additional investigations may be required to assess an interpretation and to determine whether (or when) it requires revision. Such assignments are not stable but may need revision as new knowledge accumulates. Some variants will be left in a provisional category as variants of uncertain significance, but such VUSs cannot be regarded as stable: as new knowledge accumulates, they may subsequently need to be reviewed and assigned to a different category (e.g., definitely benign or definitely disease causing). Some confidently made initial assignments may also have to be revoked and modified.

High-Throughput or Next-Generation Sequencing (NGS)

Most approaches to high-throughput sequencing rely upon a method of determining the sequence of short- to medium-length fragments of DNA (up to several hundred base pairs). This operation is performed rapidly and, ideally, will determine in an unbiased fashion the base sequence of any sample of DNA that has been prepared as fragments of the appropriate size for the method being used. Several applications of NGS rely upon the unbiased nature of the sequencing method, as the relative representation (the relative copy number) of different versions of the sequence may be critical.

The quality of NGS is assessed by reporting the depth and spread of coverage, with the probability of detecting both alleles of a heterozygous base or polymorphism increasing the more often that stretch of DNA has been sequenced. With higher depth of coverage, the chance of detecting low-level mosaicism also increases. The depth of cover required for a diagnostic laboratory to be confident that it has excluded a mutation when it has failed to detect one is greater than would usually be expected in a research laboratory or in investigations of evolutionary history.

Some applications of NGS rely upon an enrichment step before sequencing, such as an enrichment for many or all transcribed genes, which focuses the sequencing method on the exome (the 1–2 % of the genome that is translated into protein). This generates far less information than performing whole genome sequencing (WGS), and the interpretation is therefore less onerous. However, the enrichment step may introduce distortion into the process so that coverage of the exome may be uneven, and some transcribed regions may be underrepresented with, as a result, an increased chance of failure to detect a mutation.

One example of the need for very deep sequencing, requiring the manyfold sequencing of the relevant areas of the genome, is in the genotyping of fetal DNA present in the blood plasma of a pregnant woman. Up to 10 % of the free DNA in maternal plasma can be fetal in origin, and so it will require many more copies of a gene to be sequenced in the maternal plasma to warrant confidence in determining

the fetal genotype than if the sample was derived directly from the fetus. Such analyses amount to a simple matter of probability, as long as the DNA sequencing process is indeed unbiased, with the power of the technique depending on the number of copies sequenced of the site of interest. It is possible to determine the entire fetal genome sequence from the free DNA in maternal plasma, but it can also be used in a more focused fashion to see if the fetus may be affected by a condition for which one or both parents are carriers. If the father carries a serious disorder caused by a mutation not present in the mother, then the failure to detect the father's mutation in the maternal plasma will be much more reassuring after any particular depth of sequencing of that base pair (or region) than if the mother is heterozygous for the disease allele (i.e., if she carries it). In the first case, one only expects to detect the mutant allele in the maternal plasma if the fetus carries it, and it would be expected to account for either 0 % or 5 % of the alleles found: one has to distinguish 0 % from 5 % of alleles in the maternal plasma. In the second case, sufficient alleles must be sequenced to distinguish reliably between 45 % and 50 % of the alleles being mutant. It is much simpler to be confident of having either detected or failed to detect a mutation than of having distinguished between two very similar ratios between the two alleles: that is why the search for fetal sequences derived from the father in the maternal plasma is so reliable in fetal sexing and for fetal rhesus genotyping.

Another context where read depth of specific sequences appears critical, and raises comparable difficulties in clinical practice and doctor-patient communication, is in the use of free plasma DNA to monitor the bulk and evolution of tumors, especially when looking for remission or early indications of tumor recurrence. Both these areas are in the early stages of their clinical application, but the landscape of potential difficulties is taking shape.

High-throughput or next-generation sequencing (NGS) has to meet prespecified quality standards for an interpretation to be clear, and, even then, the interpretation may be problematic in several respects. It may fail to meet the necessary standards and then it must be repeated if trustworthy information is to be provided. Interpretation of NGS analyses is intrinsically probabilistic in nature and some – either patients or clinicians – may find that to be conceptually or emotionally challenging when they seek binary (true/false) answers to their questions. The use of NGS analyses of DNA in maternal plasma for prenatal genetic diagnosis raises especially challenging issues.

Data-Rich Biology and Unsought Information

An important aspect of any genomic investigation, whether the use of a DNA array to perform array comparative genomic hybridization (aCGH) or of NGS to sequence an exome or genome, is the detection of unsought but possibly important information. Such findings are unrelated to the initial reason for genetic testing and are

known as incidental findings (IFs) or secondary findings. While IFs used to occur very occasionally with conventional chromosome analysis or with the earlier, very highly targeted molecular genetic approaches, they occur much more frequently now that genetic investigations are simultaneously both broad (genome-wide) and very detailed (determining the full DNA sequence). These findings, once VUSs have been excluded, can be broadly divided into four categories: (i) important findings potentially relevant to the present or future health and well-being of the patient or close family members, for which medical interventions could be usefully enabled by an early recognition of the genetic results; (ii) similarly important findings but where no medical interventions are likely to be available to make a difference to outcomes; (iii) information about genetic carrier status for a recessive, sex-linked, or chromosomal disorder that is unlikely to be relevant to the health of the patient but may have implications for the health of their future children or of the children of other members of the family; and (iv) information about relationships within the family (such as paternity and incest) and information about ancestry (population of origin).

There has been much debate as to whether clinicians and/or researchers have an obligation to disclose unsought but potentially important information that comes to light during a genomic investigation. Much of this hinges on the agreement between the patient and the clinician or research team at the start of the process: was it explained to the patient (or parents) that such results might emerge and did the patient (or parents) consent to receive such information? When should information be disclosed without the patient having agreed in principle to receiving IFs? When, if ever, should IFs be disclosed despite the patient having stated explicitly that they would *not* wish to be given such information? We will not address these questions here, but it is important to recognize that any genomic investigation raises these challenging issues. Further, this question is not one that a molecular scientist can completely evade by leaving it to the clinician's discretion because the formulation of the laboratory report, as it appears in the patient's medical notes, has extensive implications for clinical practice and for the disclosure of information to the patient or family members, whether deliberate or inadvertent. Some close relationships identified in the laboratory may also indicate that a crime has occurred, perhaps child sexual abuse within the family, and there may be a legal duty to disclose such results to the legal authorities on those grounds.

It is also possible for an individual's identity to be inferred from their genome data, along with highly confidential information about their family relationships and their susceptibility to disease. Inadequate data security is therefore a concern, as well as the public release of the sequence of individuals' genome data held in research biobanks (Malin et al. 2011). An individual's genome sequence needs to be treated with respect as private and personal.

Information generated in the process of attempting to answer a specific clinical question may provide an answer to questions that had not even been asked. The status of such unsought information (known as incidental findings) is problematic, although it may sometimes be of direct medical benefit to the patient and at least some other members of the family (Clarke 2014).

Information Flow in Genetics

It may be helpful at this point to relate the origin or source of genetic information to its potential applicability and also perhaps to the concept of causation. There is a constant interplay between the gathering of phenotypic information and biological samples for analysis from families and the giving of information back to the contributing families and to others.

In the context of clinical genetics, the back-and-forth flow of information between patients and families, on the one hand, and professionals, on the other, has been most fruitful. The recognition that two or more children have similar facial features in association with comparable developmental difficulties has enabled numerous diagnostic entities to be established. Listening to the family experiences and observing the features and behaviors of the affected individuals allows professionals to build up a detailed description of the disorder. Seeing a few additional cases with similar features then leads to the recognition of a pattern, so that the professionals can pass on their accumulated knowledge of the condition to the next family seen. Sharing the information – the accumulated experience of one professional in the form of publications or presentations – with colleagues transforms it into information available to all. Initially, the diagnosis is based on pattern recognition, and this takes on something of the quality of an explanation for the problems observed. However, what the professionals have achieved in such a case is not strictly a causal explanation of the genetic mechanisms underlying the condition but rather a high level of confidence that the phenotypic pattern recognized by the clinician is indeed genetic in origin, so that they can give strong assurance that there is an explanation to be found.

As knowledge accumulates further, the chromosomal or molecular genetic basis of the condition may be identified. Thus, as the techniques for studying chromosomes were refined, the deletion found on chromosome 15q in two thirds of children with Prader-Willi syndrome was recognized and this brought the community of "professionals-plus-families" much closer to an explanation for the occurrence of the condition. There might still be no detailed explanation for each link in the pathogenetic chain stretching from the deletion to the phenotype, but the two ends of the causal chain (or causal "web") are well described, and some of the intervening processes have been clarified (Ernst 2008). Once additional information becomes available – such as the finding that (apparently) the same deletion is often observed in children with Angelman syndrome, a very different pattern of physical features and developmental difficulties – then the stage is set for a further iteration of the cycle of information flow between patient/family and clinician/laboratory.

In disorders where gene mapping was achieved as a result of genetic linkage studies, the families often cooperated and collaborated with the professionals and researchers, becoming partners in the enterprise. Both affected and unaffected members of many families not only contributed samples of blood and tissue but also detailed descriptions of life with their disorder, and they may have raised funds to promote the research.

Clinicians have learned from the families of affected individuals and have collated this information collectively, as a profession, and then fed it back to the same and other families. The recognition of causal cytogenetic or molecular genetic explanations for a disorder fits into this process, with clinical phenotypic information enabling the laboratory work to proceed using samples from a homogeneous group of patients, thereby maximizing the chance of finding a common underlying pathogenetic process.

Producing Results for Patients: Interpretation Through Comparison

When genetic investigations are performed so as to give results (a category of information) to patients or those at risk of disease, several of the steps described above have to be coordinated. The relevant elements in the patient's genome will be analyzed, often by DNA sequencing, and this raw data will be interpreted in the light of the standard, reference sequence of the human genome and previous sequencing of patients with the disease under consideration.

If a variant is found that has previously been reported as strongly associated with the disorder, then the interpretation may be straightforward, although the penetrance of the variant (the chance that it will produce a phenotype in any particular individual) may not always be clear-cut. If the variant has not been reported before, then the laboratory will assess how likely the variant is to cause a disease phenotype. This process can become complex and expensive, so that diagnostic laboratories may not always be able to undertake the full set of studies required to produce the best interpretation possible.

Steps that may (in principle) be taken to achieve an interpretation include:

- Checking literature, websites, and databases for previous reports of the same variant (or closely related variants)
- Examining relatives to see if the disorder in question is familial or sporadic
- Testing relatives (especially parents) for the variant to see whether it has occurred de novo or has co-segregated with the disorder in the family
- If the variant leads to a change in the amino acid present at the corresponding point in the protein, then assessing the resultant degree of change in the size and character of the amino acid side chain to assess pathogenicity
- Assessing the degree of evolutionary conservation of the altered amino acid or protein domain, perhaps comparing the *H. sapiens* reference sequence with other mammals, vertebrates, insects, lower animals, and yeast
- Using software to perform in silico prediction (i.e., by computational modeling) of any effects of the variant on splicing of the transcribed RNA
- In vitro functional assessment of the effects of the variant on transcription of the gene

- If the structural characteristics of the relevant protein and its function are known, then using structural modeling to assess the functional consequences of the variant
- Creating a model organism with the same variant in its homologue of the gene to assess any resultant phenotype it may display

The information generated by these efforts can then be used in the interpretation of the patient's raw sequence data or aCGH result, to give a more robust interpretation.

In practice, the functional, the protein structural modeling, and the model organism steps are largely research procedures and not available to diagnostic laboratories, although that may change within just a few years.

As discussed above, the results of SNP-based genome-wide association studies may give highly significant results, but they are generally of little or no clinical utility as the relative risks of disease associated with each variant are usually little different from unity and the combination – the merging – of multiple odds ratios into a single risk figure depends upon models of uncertain (unproven) validity. There are few circumstances, therefore, in which it is reasonable to apply such population data to gain information about specific individuals in relation to risk of disease, although interesting information about ancestry may be inferred.

Ownership, Control, and Access

Once the raw genetic data have been interpreted in a coherent and defensible manner, where does it go? Who has a right to the information and who can control access to it?

The concept of ownership of genetic information is often drawn upon in discussions on this topic but may be criticized as intrinsically incoherent and unsustainable. The critique of genetic information as "stuff," to which ownership could be applied, has been mounted by, among others, Manson and O'Neill (2007, pp. 98–99). Their argument challenges the "conduit model" of genetic information. Information is not a "thing" like a football, as it can be both passed and kept at the same time. It has to be conceptualized accordingly, with more attention to access and application than ownership per se.

In the context of basic research, the techniques developed for use in the laboratory may be subject to patent law, but the facts of nature that are discovered through research are not patentable, as patenting requires that the process be novel, nonobvious, useful, and – crucially – an invention (Mackenzie 1999). Attempts to patent normal gene sequence information met with some early success, but they have, eventually, been overturned in both Europe and the USA (Kesselheim et al. 2013). There is also something especially objectionable to the idea of patenting normal human gene sequences, akin to notions of a trade in body parts (Macklin 1999).

Research into differences between populations is important and has useful applications but can also generate social and ethical difficulties. Finding the frequencies of recessive disease alleles in different population groups may be useful in providing accurate information about reproductive risks in genetic counseling. Tracking the migration and evolution of early humans across the globe may also be of great interest, although care must be taken by researchers not to cause offense to groups who provide samples for such research without appreciating that the fruits of the research may conflict with their people's collective myths of origin.

In the context of clinical research, genetic information is increasingly being used to develop new, rational therapies designed on the basis of the knowledge of disease mechanisms gained through genetic studies. This might take the form of targeting drugs to act on a metabolic or signaling pathway implicated as a result of genetic investigations. Repeated analysis of tumor DNA may be helpful in tracking acquired resistance to treatments and thereby selecting the most effective cancer chemotherapy. Another approach is the direct supplementation of gene action through gene therapy (inserting an intact copy of the gene where the disease results from insufficiency of gene product) or through protein supplementation, as in various types of enzyme replacement therapy. The techniques used to synthesize or deliver the therapeutic agent are open to patenting in such research.

Moving to the context of specific families, there has been much interest in the disclosure of genetic information within families, especially when information about one individual may be relevant to the health and well-being of another. When family communication works well, there is no problem in that those who need to know that they are at risk of developing a late-onset disorder, such as an adult-onset form of cancer, are given this information in good time, so that they can arrange the appropriate medical surveillance. Equally, those at risk of bearing children affected by a serious inherited disorder are given this information ahead of any pregnancy, so that they can learn the facts and make a calm decision as to how to approach reproduction. When there are impediments to the flow of such clinically relevant information within a family, however, the question arises as to whose duty it is to transmit the information and what professionals should do if family members fail to perform this task. In essence, does information about the genetic constitution of an individual belong to that person alone or also to other members of the family? (Forrest et al. 2003; Parker and Lucassen 2004).

A particularly interesting aspect of the discussions around personal genetic information is the extent to which one ought to know all there is to be known or at least all that may be useful. Is it ever responsible, or even permissible, to choose not to know information about one's genetic constitution that could have important effects? (Chadwick 1997). The potential for paradox has to be circumvented – it may be difficult to assert that one does not wish to know a fact, without some knowledge that there is a fact to be known (Tymstra 2009). However, willful ignorance of the relevant facts, it could be argued, undermines any claim to be making decisions as an autonomous individual. Autonomy can be seen as different from whim or caprice and has to be grounded in reality, taking into account the facts as they are (Husted 1997). There are those who think that the obligation to know all

there is to be known extends to a duty to know the genetic constitution of one's potential future children – and only to bear (i.e., to carry to term) the best of these potential children (Savulescu and Kahane 2009). These are further aspects of genetic information that cannot be pursued here.

There are additional concerns about access to genetic information by third parties, such as insurers, employers, and governments. The perceived importance of ensuring privacy for information about individuals' genetic information is likely to be the strongest where funding for health and social care depends upon health insurance policies, taken out by individuals with private companies, because of the concern that such companies will charge those at increased genetic risk of disease an increased, perhaps prohibitively increased, level of premiums. This could effectively deny health care to some who need it most. State-funded schemes are less likely to discriminate in this way and more likely to spread the higher cost of care for some individuals across the population as a whole through general taxation. Confidence in the long-term future of a state-supported health-care scheme may therefore be critical in shaping attitudes to the privacy of genetic information. It is important to note here that genetic sequence data is intrinsically identifying, so that the public release of genome sequence data through biobank initiatives entails a risk that personal information could become known about the individual contributors to a research biobank (Malin et al. 2011). One approach to this problem is to address these misapplications of genome data rather than erecting ever more barriers to the sharing or pooling of genome data, as the barriers are likely to prove futile and fail to protect personal privacy.

Definitions of Key Terms

Association	The nonrandom association between a specific allele (version) of a polymorphic (variable) site in the genome (usually a single nucleotide polymorphism or SNP) and a particular phenotype. Typically, a strong statistical demonstration can be achieved of multiple but weak associations between a phenotype and a range of widely dispersed SNPs.
Complex disorder	The common degenerative disorders of Western society, including coronary artery disease, stroke, diabetes mellitus types 1 and 2, Alzheimer's disease, inflammatory bowel disease, and the common cancers of breast and bowel, are influenced by many genetic factors in interaction with each other and with the environment.
Exome	The portion of the genome that is copied into messenger RNA, as part of the process of protein synthesis (about 1–2 % of the genome).

Genetic linkage	The tracking together of two sites of DNA variation located on the same chromosome through the generations of a family. The closer the two sites are together on the same chromosome, the less likely they are to be separated in the production of gametes by recombination events at meiosis. In gene mapping research, a disease gene would be localized by finding a site of polymorphic variation that tracked through the generations of a family along with the disease in question, indicating that the marker and the disease gene were located close together on that chromosome.
Genome-wide association study (GWAS)	The simultaneous search for associations between one phenotype and the various alleles at a wide range of polymorphic sites (SNPs) from across the genome.
Heritability	The proportion of the variance (loosely, the variability) in a measure of phenotype found in a specific population that can be attributed to genetic factors.
Incidental finding (IF)	The generation of information about a patient's present or future health that has emerged from a genetic investigation initiated for some other reason. This is especially likely to occur with genome-wide tests such as array CGH, exome analysis, or whole-genome sequencing.
Single nucleotide polymorphism (SNP)	A polymorphic (variable) site in the genome where the less common allele has a population (allele) frequency of at least 1 %.
Variant of uncertain significance (VUS)	The presence at a site in the genome of a variant base in the DNA sequence, where the functional impact of the variant (such as its pathogenicity) is unclear.

Summary Points

- Genetic information has multiple meanings, including family history of disease, raw DNA sequence, or a carefully interpreted laboratory report that defines the cause of a patient's inherited disorder.
- Most interpretations of DNA sequence data rely upon comparison of the new sequence with reference sequence data and with previously reported variants. Interpretation is largely inference on the basis of comparison.

- The methods of establishing DNA sequence information have developed so that large numbers of single molecules can now be sequenced at low cost ("next-generation sequencing"). This has transformed the practice of biology and is having increasing impact on medicine and health care.
- With the generation of large volumes of DNA sequence data, it is the interpretation rather than the sequencing itself that proves to be challenging, especially the variants of uncertain significance (VUS) that make clear the provisional nature of interpretation and raise the question of future reanalyses and reinterpretations of genome sequence data.
- Generating important information that has not been sought is also a professional challenge for medicine and genetic counseling (the incidental findings).
- The privacy of genetic information is important but abuse may be prevented more effectively by control of the uses to which genetic information is put than by attempts to "guard" it more closely.

References

Chadwick R (1997) The philosophy of the right to know and the right not to know. Chapter 1. In: Chadwick R, Levitt M, Shickle D (eds) The right to know and the right not to know. Ashgate, Aldershot/Brookfield, pp 13–22

Chadwick R, Capps B, Chalmers D, Clarke A, Clayton EW, Liu E, Lysaght T, Mulvihill JJ, Marianne Winslett M (2013) Imagined futures: capturing the benefits of genome sequencing for society. Human Genome Organisation

Clarke AJ (2014) Managing the ethical challenges of next generation sequencing in genomic medicine. Br Med Bull 111(1):17–30

Ernst Z (2008) Philosophical issues arising from genomics. Chapter 13. In: Ruse M (ed) The oxford handbook of philosophy of biology. Oxford University Press, Oxford, pp 310–326

Evans JA, Foster JG (2011) Metaknowledge. Science 331:721–725

Forrest K, Simpson SA, Wilson BJ, van Teijlingen ER, McKee L, Haites N, Matthews E (2003) To tell or not to tell: barriers and facilitators in family communication about genetic risk. Clin Genet 64:317–326

Husted J (1997) Autonomy and a right not to know. Chapter 6. In: Chadwick R, Levitt M, Shickle D (eds) The right to know and the right not to know. Ashgate Publishing, Aldershot/Brookfield, pp 55–68

Kesselheim AS, Cook-Deegan RM, Winickoff DE, Mello MM (2013) Gene patenting – the Supreme Court finally speaks. N Engl J Med 369:869–875

Mackenzie R (1999) Paradigms of author/creator property rights in intellectual property law. Chapter 13. In: Thompson AK, Chadwick RF (eds) Genetic information: acquisition, access and control. Kluwer Academic/Plenum, New York, pp 139–148

Macklin R (1999) The ethics of gene patenting. Chapter 12. In: Thompson AK, Chadwick RF (eds) Genetic information: acquisition, access and control. Kluwer Academic/Plenum, New York, pp 129–137

Malin B, Loukides G, Benitez K, Clayton EW (2011) Identifiability in biobanks: models, measures, and mitigation. Hum Genet 130(3):383–392

Manson N, O'Neill O (2007) Rethinking informed consent in bioethics. Cambridge University Press, Cambridge

Parker M, Lucassen AM (2004) Genetic information: a joint account? BMJ 329:165–167

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D (2015) Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet 16:85–97

Savulescu J, Kahane G (2009) The moral obligation to create children with the best chance of the best life. Bioethics 23:274–290

Tymstra T (2009) 'If ony I had (not) known that': about risk-individualization, chance specification and the loss of certainty of not-knowing. Chapter 7. In: Rehmann-Sutter C, Müller HJ (eds) Disclosure dilemmas. ethics of genetic prognosis after the 'right to know/not to know' debate. Ashgate, Farnham/Burlington, pp 85–95

Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM (2013) Pitfalls of predicting complex traits from SNPs. Nat Rev Genet 14:507–515

Conceptualization of Genetic Disease

22

Catherine Dekeuwer

Contents

Abstract

The expression "genetic disease" suggests the idea that one or more genes cause a disease. Since the identification of the so-called disease-causing genes, many genetic variants associated with common diseases such as diabetes or cancer have been identified. But if the involvement of genetic factors in the development of a condition justifies labeling it "genetic," the definition of genetic disease is likely

C. Dekeuwer (✉)
Jean Moulin University Lyon III, Lyon, France
e-mail: catherine.dekeuwer-carrier@univ-lyon3.fr

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_81

to become trivial. Indeed, all human diseases would be genetic, since virtually every medical condition, disease susceptibility, or response to treatment is caused, regulated, or influenced by genes. Moreover, genes alone are not responsible for the development of diseases: their etiology is complex with many factors interacting with each other. So when is giving primary importance to genetic factors justified? The history of medicine may help us see which biological and technological advances have influenced conceptualization of genetic disease, while philosophical analyses of the causal selection problem may help us better understand how one might justify claims that genes are the main causes of certain diseases. The last part of the article examines interactionist attempts to deal with genetic factors involved in human diseases and particularly network medicine.

Introduction

The expression "genetic disease" suggests the idea that one or more genes cause a disease. For example, phenylketonuria, cystic fibrosis, Huntington's disease, or certain forms of breast and ovarian cancers are generally considered to fit this description, which has arisen from the localization and identification of the so-called disease-causing genes at the end of the twentieth century.

The history of genetic medicine, however, which began with an enthusiastic program for discovering disease-causing genes, now seems to be hindered by the observation that all diseases might be caused by some genes. Many genetic variants associated with common diseases have been identified such as the sporadic form of Alzheimer's disease, obesity, diabetes, sporadic cancers, antisocial personality disorders, and even tuberculosis. Consulting the catalogue of human genes and genetic disorders (Online Mendelian Inheritance in Man) raises even more questions about the definition of genetic disease, as diseases linked to mitochondrial DNA mutations, epigenetic modifications, and even congenital disorders are referenced. This has led some scholars to ask whether cancer or tuberculosis could be labeled "genetic diseases" and raises the problem of the triviality of the concept. Extending the concept of genetic disease empties the term of its specificity and so no longer conveys any useful information. Moreover, genes alone are not responsible for the development of diseases: their etiology is complex with many factors interacting with each other. This raises a key question: when is giving primary importance to genetic factors justified?

In order to set some signposts in what looks like a fuzzy landscape, the first section starts with the history of medicine, so as to recount the biological and technological advances that have influenced the conceptualization of genetic disease, as well as drawing attention to some potential shifts in that conceptualization. The second section will then consider philosophical analyses of the causal selection problem with a view to discussing the ways in which one may justify the claim that the genes are the main causes of certain diseases. The last section will turn to some contemporary interactionist developments in biology and medicine that shed light on the somewhat puzzling use of the concept of genetic disease.

From Predictive to Personalized Medicine

In the 1970s, French Nobel Prize J. Dausset developed the concept of predictive medicine (Dausset 1972, 1996). The discovery of correlations between histocompatibility antigens and susceptibility to illness, the hypothesis that a set of genetic factors determines for each individual her susceptibility to common diseases like cancers, and the first advances in the attempt to map the human genome supported the idea that twenty-first-century medicine will be predictive. Intended to apply to healthy individuals, predictive medicine was conceived as the practice of determining whether a person's susceptibility to specific diseases was increased and of proposing appropriate measures of prevention.

The similarity of Dausset's project with the means and goals of contemporary genomic medicine is striking. The hope that medicine will prevent common diseases like cancers, diabetes, asthma, mental illnesses, or heart disease supported major biological projects such as the Human Genome Project (HGP) and genome-wide association studies (GWASs). The National Human Genome Research Institute's presentation of the GWAS project states, for example: "the impact on medical care from genome-wide association studies could potentially be substantial (. . .) In the future, after improvements are made in the cost and efficiency of genome-wide scans and other innovative technologies, health professionals will be able to use such tools to provide patients with individualized information about their risks of developing certain diseases. The information will enable health professionals to tailor prevention programs to each person's unique genetic makeup. In addition, if a patient does become ill, the information can be used to select the treatments most likely to be effective and least likely to cause adverse reactions in that particular patient" (https://www.genome.gov/20019523, accessed 20 July 2015). This project, often called "personalized medicine," is today at the forefront of much medical research.

The history of the changes that occurred since the 1970s can be described as a two-phase process. Some distinctions between genetic and genomic medicines, genetic causes and genetic factors, monogenic and common diseases can be found in scientific literature. For example, some researchers have a preference for the expression "genomic medicine" over "genetic medicine," for it implies a focus on medicine that uses new technological methods of genomic data acquisition and analysis to study the complex interaction between genetic and environmental factors in common diseases and, as such, it is clearly distinct from the attempt to study Mendelian monogenic diseases. However, most of the time, there are no such clear distinctions. A brief survey of the history of medicine sheds light on how anarchical the use of the concept of genetic disease is today.

Monogenic Diseases Are Not Simple

In the 1980s, the identification of so-called disease-causing genes led the predictive medicine project to be temporarily put aside. Researchers carried out linkage analyses in order to map the genes for monogenic diseases such as thalassemia,

Huntington's disease, or cystic fibrosis. Even if known environmental factors (e.g., altitude in the case of thalassemia) or unknown phenomena play a part in the age of onset of the symptoms or their severity, the penetrance of these monogenic diseases is almost 1. This means that almost all the individuals carrying a particular variant of a gene (the so-called genetic mutation) also express the pathological phenotype. At the same time, molecular medicine was also rapidly expanding. Following the "central dogma" of molecular biology, it sought to explain why a given genetic change should result in a particular clinical phenotype, how a change in a particular DNA sequence could modify the quantity or function of the gene's product, and why this change is pathogenic for a cell, a tissue, or a stage of development.

As a small proportion of breast and ovarian cancers are inherited in an autosomal dominant manner, BRCA1 and BRCA2 genes were localized in 1990 and 1994 using linkage analysis. The presence of BRCA1- or BRCA2-specific variants significantly increases the risk of developing such cancers, but the penetrance of the pathological phenotype is incomplete: a proportion of individuals carrying the identified variants will never develop the symptoms. For instance, in 2007, the cumulative risk of developing breast cancer for a person carrying a BRCA1 mutation at age 70 years was estimated at 57 %, while the risk is about 10 % in the general population (Chen et al. 2007). The concept of genetic predisposition was born, referring to a highly increased risk of developing a disease associated with specific genetic variants. The molecular effects of the BRCA1 and BRCA2 variants began to be studied.

The localization and identification of some disease-causing genes and genetic predispositions led to the rapid development of specific medical practices, such as genetic counseling, aimed at helping individuals to make autonomous decisions regarding genetic testing. As more and more genetic presymptomatic and prenatal tests were available, raising important ethical questions, several countries chose to supervise genetic medicine by legislative means. Genetic medicine is today a normalized practice based on the knowledge of very high correlations between the presence of specific genetic variants and the occurrence of certain diseases in a family. Increasingly, the specific molecular effects of these variants are also being identified.

Recalling this history underlines three characteristics of the concept of genetic disease. (1) A genetic disease is inherited according to Mendel's laws. (2) The development of symptoms is highly correlated with the presence of specific genetic variants whose knowledge supports genetic testing. (3) One can explain the contribution of the variants to the molecular transformations responsible for the symptoms. None of these three characteristics is sufficient for labeling a disease "genetic," but they work together as landmarks when reflecting on how to conceptualize genetic disease.

Things became more difficult as knowledge of the relationship between genotype and phenotype developed further. For example, phenylketonuria used to be considered as a monogenic disease caused by some specific variations in the PAH gene. But, in absence of PAH mutations, an abnormal BH4 protein may result from other genetic variations and also cause the specific symptoms of phenylketonuria.

Scientists have to conceptualize this phenomenon, called "genetic heterogeneity." They have also the hard task to understand the phenomenon of incomplete penetrance as well as the variable expressivity of diseases. This means to understand why – considering the same genetic variant – symptoms may be more or less severe and why some patients develop one set of symptoms, whereas others develop a different set of symptoms. Some studies (see Friebel et al. 2014) have tried to show the influence of environmental factors on the development of symptoms, such as pregnancies or history of breast feeding, for breast and ovarian cancers. Genetic factors and genetic regulation may also explain the phenotypic variations of the same monogenic disease. For example, the CFTR gene was established in 1989 as the "cystic fibrosis-causing gene," but variations in lung and digestive symptoms correlate poorly with variants of the CFTR gene as well as with environmental factors. Considering that the phenotype may be polygenic (i.e., modified by the presence or absence of genetic variants other than the CFTR gene), "modifier genes" have been identified, which influence the phenotype. Cumulative genetic effects, genetic regulation, and the functional role of specific genetic variants in the control of infection, immunity, or inflammation are all working hypotheses aiming to account for the phenotype–genotype correlation in cystic fibrosis disease (Dipple and McCabe 2000; Badano and Katsanis 2002).

Disease-causing genes, as well as genetic predispositions, do not work in an "all or nothing" way. Understanding the variable expressivity of monogenic diseases and the differences in the severity of symptoms or in the timing of their appearance requires the study of several genetic factors that interact with each other, as well as of the interactions between genetic and environmental factors. It also requires the understanding of their involvement in the biological pathways that determine occurrence of illness in individual persons. In other words, the apparent simplicity of the concept of genetic disease has given way to an appreciation of its complexity.

The Black Box Strategy of Genome-Wide Association Study

Association studies have led to the identification of many genetic variants statistically correlated with common diseases (see Visscher et al. 2012). As knowledge of the mode of inheritance of the disease is not necessary to complete association studies, they are particularly useful for studying common diseases. The recent GWASs are based on the "common disease–common variant" hypothesis (Reich and Lander 2001), which states that common diseases are attributable in part to genetic variants very common in the population. These variants are not sufficient to cause a condition, but they are statistically more frequent in the affected population than in the non-affected one. The strategy of GWAS is to study large populations of healthy and affected individuals and to find some differences in the frequency of markers in individuals who are or are not affected by the condition (the most commonly used markers are called "single-nucleotide polymorphisms" or SNP). Thus, nothing is known about the biological effect of the genetic component highlighted in this

way: GWAS, like other association studies formerly used, is a "black box strategy." For this reason, a genetic variant that has been identified by association study cannot be considered crudely as a cause of the disease. The concept of a genetic factor thus turns out to be rather empty, for it amounts only to a specific SNP statistically correlated to a disease that confers at best a slightly increased relative risk in comparison to the general population. Problems raised by these association studies such as size and choice of populations, reproducibility of the study, or the threshold of significance of the correlations are now well known (Visscher et al. 2012).

GWAS also raises some additional difficulties. First, it gives no clue as to the genetic architecture of common diseases, i.e., the way genotype and phenotype are related. Moreover, these common genetic variants confer only a relatively small increment in risk and explain only a small proportion of heritability. For example, even if the estimated heritability of a complex trait is about 80 %, the genetic variations associated with this trait generally explain only about 5 % of that heritability, despite studies of thousands of people. This phenomenon, called "missing heritability," has led some researchers to state a new hypothesis: "common disease–rare variant." According to that hypothesis, the effects of low-frequency variants of intermediate effect and the interaction between several genes or the effect of epigenetic factors might contribute to the explanation of missing heritability (see Manolio 2009).

However, the main limitation of association studies remains their black box strategy, which makes it necessary to add further biological investigations about the potential functions of the DNA regions identified. It is also important to note that association studies are not the only example of this limited strategy. For instance, inheritance of Huntington's disease was first associated with a genetic marker, and the first genetic tests were set up without any knowledge of the sequence and the function of the gene involved in the disease (Bates 2005). To take another example, the history of the identification of the CFTR gene shows how the researchers confined a locus by means of a genetic marker, picked out candidate genes, and sequenced them. This strategy, called "reverse genetic," led them to select the CFTR gene, whose function was congruent with biological and pathological evidence (see Dekeuwer 2015). Today, GWAS is frequently contrasted with candidate gene approach since the former "wide" strategy is not compatible with the selection of prespecified genes of interest, i.e., candidate genes. However, when conducting *medical* research, it is necessary to complement the identification of a set of SNP markers correlated with a particular condition with the study of the potential biological pathways in which the related DNA region might be involved.

Thus, the contemporary and somewhat arbitrary use of the concept of genetic disease is the consequence of some still more complex analyses of the human genome. It is also the consequence of the extension of our knowledge of the genetic determinants of diseases. The developments of classical Mendelian genetics also cast doubt on the concepts, taking us back to the complexity of the interactions involved in pathological pathways. It now seems necessary to raise the question whether genetic factors should at all be considered the most important.

Genes that "Make the Difference"

The etiology of biological traits is extremely complex and involves a large number of causally relevant factors that interact with each other. The causal selection problem is the problem of how to select the cause "that makes the difference" and conceptualize it. Some philosophers (Hesslow 1984, 1988; Norell 1984; Gifford 1990; Gannett 1999; Magnus 2004; Smith 2007; Dekeuwer 2015) have tried to give critical assessment of the criteria that may resolve the causal selection problem. However, since all these criteria show insufficiencies, and as it is difficult to envision a trait in which genes are not involved, it seems impossible to prevent the concept of genetic disease becoming absolutely trivial. This section summarizes some arguments that support the view that genes could make the difference and discusses the consequences of an unjustified extension of the concept of genetic disease.

Genes as Specific Causes

The first strategy to solve the causal selection problem tries to specify the causal relationship between gene and phenotype. Criteria of necessity, sufficiency, and directness have been discussed, but the strongest criterion is that of specificity. As Fred Gifford argues, in order to be considered genetic, "a trait must be described or individuated in such a way that it is properly matched to what the gene causes specifically" (1990, 329). In the complex process of causality, some causal factors are indeed "universals" because they are present in all cases of both the occurrence and the nonoccurrence of a phenotype. According to Gifford's example, the absence of a 1,000 °C temperature is a necessary condition for having a given condition. But it does not constitute an explanation of this particular condition, since human life in general requires that factor. On the contrary, a genetic effect is "specific" if the modification of a gene has some effect on the considered phenotype but not on other ones. Moreover, in Gifford's account, this phenotype must be individuated (i.e., correspond to one precise unit of description), neither too broad nor too narrow. Hypercholesterolemia, for instance, cannot be considered a genetic disease because one can distinguish familial and sporadic forms of this disease.

Even though Gifford's account of genetic disease has been criticized, it constitutes a good criterion to resolve the causal selection problem. To be sure, it is not an absolute criterion because for some diseases, there are many necessary conditions which, properly construed, are neither universal nor specific. Nonetheless, it constitutes a clear echo of the conceptualization of genetic disease analyzed in the first section of the present article. The concept of specific effect refers indeed to molecular genetics, and Gifford explains that one should accord less importance to the question of knowing if a trait is genetic than to the steps genetic variations modify in biological pathways.

The Epidemiological Strategy

The second relevant approach to genetic causality refers to populations rather than individuals. It states that a disease is genetic with respect to a population P if genetic factors make the difference between the persons who suffer from the disease and the rest of the P population (see Smith 2007). In this epidemiological account of genetic disease, it is entirely possible that the analysis of one population will label a disease "genetic," while a similar analysis of another population will label the disease "environmental." Philosophical discussions often refer to the example of a water supply contaminated by a pathogen. Only half of the villagers fall ill and the researchers assume that their genes confer some resistance to the pathogen. On the scale of the village, the infection is genetic because genes are practically sufficient to bring about the condition. On the planetary scale, however, the disease would probably be labeled "environmental" and the pathogen considered as the cause of the disease.

Relativity is thus the keyword for this conceptualization of genetic disease, and some scholars reject this strategy for solving the causal selection problem because the concept of genetic disease becomes relative to population. After all, the problem of causal selection is simply pushed back to the problem of population selection. Moreover, this epidemiological account of genetic disease is completely at a loss when faced with the task of explaining individual occurrences of a disease. Again, the attempt to give an absolute criterion in order to label a disease as genetic ends in failure.

A Pragmatic Solution to the Causal Selection Problem

Reflecting on the differences between genetic and genomic medicines and on the role of geneticists in the personalized medicine of the future, Charles Epstein stated that even if the genetics of common diseases is more complex than the genetics of monogenic conditions, it is "still genetics nonetheless" (2006, 437). The fact that genomic tools make it possible to acquire the knowledge required to identify the genetic components of these diseases, as well as the fact that genetic testing procedures may be carried out to predict who is at risk, constitute for him significant arguments.

More generally, the pragmatic position states that a cause makes the difference when it is the most "manipulable." The above example of the water supply shows that we consider the pathogenic agent to be the main cause of the disease because we believe that we can manipulate this cause: we can clean the water supply, for instance. However, practical as well as theoretical arguments can be opposed to the position that gene factors could be the main cause of genetic diseases.

Thus far, genetic manipulation has not proved to be a simple strategy for managing disease. Difficulties in finding safe methods to introduce a therapeutic gene into human cells as well as ethical problems regarding human genetic manipulation are strong impediments to gene therapy advances. Even if gene transfer

protocols have been approved for human use in inherited diseases, cancers, and common diseases, such as blindness, hemophilia, thalassemia, melanoma, or Parkinson's disease (see the "Gene Therapy Clinical Trial Worldwide" database and Misra 2013, gene therapy is still an exception, and the property of "manipulability" cannot be considered as a solid criterion for the causal selection problem.

Should this pragmatic approach thus be reduced to its preventive dimension? In that case, a disease would be labeled "genetic" when a genetic test could be put on the market which predicts a relative risk of developing the condition. This position falls victim to a more general objection: given the role played by genes in all traits, any phenotype will be genetically manipulable in principle. However, according to our knowledge of the etiology of phenotypes, any trait that is genetically manipulable in principle will also be environmentally manipulable in principle. So the causal selection problem reappears, this time in a pragmatic guise: how does it benefit us to describe a disease as genetic?

The Problem of Geneticization

Some philosophers are still working hard to resolve the causal selection problem, but one can also ask what the consequences of giving up on that problem would be.

As medical research led to an extension of the concept of genetic disease, a neutral concept of geneticization was proposed. This neutral concept holds that the process of geneticization consists in a change in the explanation of a disease such that it is considered to take place at the molecular level. Some sociologists have proposed local and precise descriptions of that change for particular diseases such as cystic fibrosis or diabetes (Hedgecoe 2002, 2003; Keer 2004, 2005).

The difficulties in finding a satisfactory solution to the problem of causal selection may however lead to a much more critical position. In the 1990s, some sociologists and philosophers (Lippman 1991; Nelkin and Lindee 1995; Hoedemaekers and Ten Have 1998) described the rapid expansion of genetics in medicine and society, which they called "geneticization" in a critical sense. They stated that there were ideological reasons for favoring genetic explanations of differences in abilities, behavior, and disease. For them, labeling a disease as genetic refers more to a set of social and political commitments about the best way to allocate resources than to biological or medical considerations. Even today, the concept of geneticization is still associated with deterministic and reductionist descriptions and interpretations of human life and behavior. It also underlines how using genomic vocabulary and metaphors may have perverse social consequences. Indeed, the confusion of "genetic" and "inevitable" may lead to give up on costly public policies for caring and protecting people. Furthermore, it leads to the belief that the only solution to a genetic problem is the selection of individuals based on genetic criteria, a solution some might view as objectionably eugenic.

The problem of causal selection may have no solution as long as philosophers continue to seek a unique and absolute criterion that could justify holding a genetic factor to be the main cause of a complex biological phenomenon such as disease.

However, the three abovementioned strategies constitute good points of reference when it comes to the question of whether it is justified to label a particular condition as genetic (see also Dekeuwer 2015). This section nonetheless has also highlighted what is perhaps a more difficult problem than the causal selection problem, namely, that a viable criterion of genetic disease may not be sufficient to hold back a dangerous process of geneticization.

Interactionism and New Perspectives in Medicine

This last part of the chapter endorses the causal parity principle (also called "causal democracy" principle) and considers the interactionist attempts to deal with genetic factors involved in human diseases. According to this principle, there is no reason to give any special privilege to genetic causality. Two versions of interactionism that maintain this principle open new perspectives when it comes to conceptualizing genetic diseases.

From Genetic Predisposition to Interactive Predisposition

Research into the determinism of biological traits typically regards the phenotype as the product of genes on the one hand and the environment (i.e., everything else) on the other. Researchers inquire how the phenotype varies as the DNA sequences are held constant and other factors change. The genotype's reaction norm – a very common tool in biology – is the graphic representation of this strategy.

Philip Kitcher (2003) showed that this tool is scientifically valid and respects the principle of causal parity. Indeed, isolating certain causal factors by holding them constant in order to see how the effect varies when other factors are modified is justified. Following the principle of causal parity, it is then possible to study phenotype variation as a function of a genotype held constant when environmental factors change. This is how most biological and medical research operates. But it would also be possible to carry out a study in which an environmental factor would be held constant and the variation of phenotypes would be observed as a function of this factor. In this case, we could see that for a single genotype, the phenotype varies as a function of the environmental factor. Kitcher underlines that if both designs are congruent with the principle of causal democracy and pertinent for studying the determinism of a complex trait, scientists often use the reaction norm in a genocentric way, showing how keeping a genotype constant in various environments leads to an invariable phenotype.

This genocentrism may lead to misleading views about genetic determinism and has urged some scholars to critically investigate biological and medical studies of complex traits, as did, for instance, James Tabery (2009). In a 2002 study, Avshalom Caspi et al. studied an interaction between a gene controlling neuroenzymatic activity (MAOA), exposure to maltreatment in childhood, and antisocial personality

disorder (ASPD). What made this article so interesting is the fact that the authors found a particular genetic factor and a particular environmental factor which, when joined, were highly correlated to a mental disorder. However, as these results have been reported in the media as the discovery of a "Murder Gene" and the interaction conceptualized as a genetic predisposition to violence, Tabery established a distinction between genetic predisposition and interactive predisposition. He also underlined the ethical consequences of the misconstruction of the concept of genetic predisposition to violence.

At first glance, the reaction norms published by Caspi et al. show that individuals with high-MAOA activity gradually increase their risk of developing ASPD as incidents of childhood maltreatment increase. But to the contrary, individuals with low-MAOA activity drastically increase their risk of developing ASPD as incidents of childhood maltreatment increase. This was interpreted as showing a genetic predisposition to ASPD.

According to Tabery's definition, a genetic predisposition consistently increases the risk of developing the phenotype in each tested environment. In the case of a genetic predisposition to ASPD, individuals from the low-MAOA activity group would maintain their relatively elevated risk for ASPD in *every* tested environment (none, probable, severe maltreatment in childhood). Moreover, the genetic difference between the low-MAOA activity and the high-MAOA activity groups would *consistently* put individuals from the low-MAOA activity group at an increased risk of developing ASPD compared to the individuals from the high-MAOA activity group. However, the reaction norms from the two groups clearly cross over. Their representation shows a change between the two groups across the different environments: in the absence of childhood maltreatment, the high-MAOA activity group scored higher than the low-MAOA activity group. In other words, high-MAOA activity is protective or aggravating, depending on the environment. Thus, Tabery elaborates the concept of interactive predisposition, which refers to the presence of a genetic difference between various groups that "both *increases* and *decreases* the probability of individuals from one group, *in comparison to individuals from the other group*(*s*) developing a particular phenotypic trait *depending on the environmental conditions experienced*" (2009, 35).

This work constitutes a good example of an attempt to perfect the much used conceptualization of genetic predisposition. It respects the principle of causal parity while pointing out the ethical and social risks of conceptual misconstruction. It thus shows how further studies employing an interactionist perspective could improve concepts originating from genetics.

The Concept of Developmental Interaction

Since the work of Susan Oyama (1985) and Richard Lewontin (2000), the Developmental System Theory (DST) has tried to understand the development of biological organisms from an interactionist point of view. But, far from improving classical concepts, it rejects some foundational biological notions. For DST, the assumptions

that are required to build the reaction norm, for instance, the genotype–phenotype distinction, must be reconsidered.

DST has the merit of drawing attention to some conceptual difficulties involved in contemporary studies. The partition between "gene" and "environment," for instance, is very fuzzy. This is not only because the gene is difficult to define as a concept (see Beurton et al. 2000). It is also because one cannot bring together into the same category epigenetics, internal environment (e.g., the cell functioning), external environment (altitude, exposition to a pathogen, and so on), and social environment. If not on the assumption of genocentrism, how could the partition between genes on the one hand and "all the rest" on the other be justified? Indeed, it is only by incorporating many more layers of structures and processes than the gene–environment interaction that we may account for the processes of development, including diseases.

DST also contests the idea that the partition between genetic influence and environmental influence could be quantitatively assessed. The biometric concept of interaction, introduced by Ronald Fisher, defines genetic and environmental influences statistically. Consequently, biometric studies yield only statistical and not biological evidence for genetic influence. On the contrary, interaction should be conceptualized as the result of differences in unique developmental combinations of a particular set of factors that have to be identified.

For DST, the development of an organism – and diseases are particular instances of the development of an organism – has to be understood as a co-constructionist process. Co-constructionist interactionism emphasizes that an organism constructs its environment, just as the environment constructs the organism. An organism inherits resources from its environment and modifies it. That modified environment in turn may modify gene regulations as well as the distribution of genetic variants in the population. In this kind of dynamic approach, development at each stage builds on the results of developments at earlier stages.

DST offers important criticisms of other approaches in biology, but it is yet undecided whether it will lead to a major shift in the conceptualization of genetic diseases. Although this theory has the major advantages of respecting the causal democracy principle as well as carrying out its explanations at the level of the organism, its lack of applications has been criticized. However, even if DST seems a long way from contemporary medical perspectives, network medicine appears today to be quite close to the basic project of DTS.

The Concept of Genetic Disease in the Context of Network Medicine

The expression "network medicine" was introduced by Albert-László Barabási in a 2007 publication entitled *Network medicine: from obesity to the "diseasome"* published in the *New England Journal of Medicine*. Its key hypothesis is that a disease is rarely a consequence of an abnormality in a single gene but reflects various biological processes that interact in a complex network. A disease reflects the perturbations of a series of linked networks that incorporate several intracellular as

well as extracellular components. Yet biological systems are not random and their many components are connected in complicated ways that can be characterized by a core set of organizing principles.

Using network theory, it is possible to study biological systems by presenting biological factors like molecules or diseases as nodes and their relationships (for instance, physical interaction or shared gene) as edges. Network medicine focuses mainly on topology to study the networks that are at play in the development of diseases. For instance, it uses metabolic networks to study protein–protein interactions but also social networks to map the spreading of disease across populations (Braun et al. 2008).

Network medicine defines the human "interactome" as the whole set of interactions that occur in a healthy human organism. Today, it encompasses protein–protein interaction networks, metabolic networks, regulatory networks, and RNA networks (see Barabási et al. 2011). For instance, in metabolic networks, nodes are metabolites that are linked if they participate in the same biochemical reaction. The human disease network or "diseasome" represents disease maps, whose nodes are diseases and whose links represent molecular relationships between, for instance, shared genes and shared metabolic pathways. These maps bring into light disease modules, each disease module being a group of interacting components (DNA, RNA, transcription factors, metabolites, and so on) which, if disrupted, results in a pathological phenotype.

Leroy Hood (Hood 2004; Galas and Hood 2009; Hood and Flores 2012) developed the idea that network medicine could lead to 4P medicine: a preventive, predictive, personalized, and participatory medicine (this 4P medicine is also called "personalized medicine" and refers to the project analyzed above in the first section). In reality, however, network medicine is more challenged by its various limitations than it is successful in revolutionizing medical practices. Among these limitations, one can recall that many factors affecting disease modules remain unknown, but also the limitations of statistical tools used to explore the role of networks in disease (see Barabási 2011). Several advances regarding the conceptualization of genetic disease in network medicine are nevertheless worth mentioning.

Network medicine studies are an essential complement of the GWAS, as their results depend on map networks, which in turn may lead to the identification of new genetic factors involved in pathological phenotypes. For instance, the underlying disease module of a specific disease is likely to include all disease-modifying genes involved in epigenetic, transcriptional, or regulatory phenomena. Thus, network medicine can guide further experimental work toward uncovering disease mechanisms as well as their complementary genetic factors. For instance, Yanqing Chen et al. (2008) identified a network of genes from mouse liver and adipose tissues for which three variants associated with obesity and diabetes have been identified.

Using the interactome and the diseasome is also useful to classify diseases (Loscalzo et al. 2007). For instance, mapping interactions of cellular components can lead to the identification of deeper functional or molecular relationships among apparently distinct phenotypes. Diseases could thus be classified not by their differences in phenotype but by their belonging to a specific disease module.

Above all, network medicine research involves a massive change in the definition of genetic disease. Scientists indeed used to seek the genetic factors involved in all diseases, each of them being defined by a specific phenotype. A genetic theory of diseases is now arising: "understanding diseases in the context of these network principles allows us to address some fundamental properties of the genes that are involved in disease. Indeed, only about 10 % of human genes have a known disease association. Thus, do disease genes have unique, quantifiable characteristics that distinguish them from other genes?" (Barabási 2011, 57). Evidences from genomic research can be used to list the genetic variants associated with each disease. But they can also be used to understand the general principle that links a set of specific genes to the phenomenon of disease. In that context, human disease genes are understood as the genes involved in disease in general. In other words, the concept of genetic disease is in the process of becoming related to a specific category of genes.

Unlike DST, network medicine does not give up on the most common biological distinctions (like the genotype–phenotype distinction, for instance). However, its hypotheses are grounded in an interactionist perspective, if not a co-constructionist one. They provide a useful framework, taking complexity into account and paving the way for redefining the concept of human disease genes. In that respect, network medicine may offer a satisfactory solution to the abovementioned problem of the triviality that arises from the overextension of the concept of genetic disease.

Conclusion

Genomic research leads to the identification of ever greater numbers of genetic variants involved in human diseases, and especially common diseases. It does not seem to be the case, however, that researchers working on the genetics of complex diseases are trying to show that asthma and bipolar disorder, for instance, are genetic diseases. Nevertheless, the increasing complexity of our understanding of genotype–phenotype relationships, as well as the fact that virtually every medical condition, disease susceptibility, or response to treatment is caused, regulated, or influenced by genes, contributes to the puzzle of how to conceptualize genetic diseases. Moreover, as the causal selection problem has not received any entirely acceptable solution, one cannot be satisfied with the conceptualization of genetic disease as a disease mainly caused by one or more identified genes.

To avoid this uncomfortable situation, various paths are possible. The insufficiencies of the conceptual solutions to the causal selection problem may lead one to consider that the concept of genetic disease is only pragmatic. Then, the first option consists in a philosophical or sociological criticism of the process of geneticization that characterizes contemporary medicine, and perhaps also society. The second option gives up on the causal selection problem and follows the principle of causal parity. Considering the interactions between genes and environment, it abandons simple views of gene action and tries to formalize the complex interactions that explain pathogenesis. The distinction between the concept of genetic predisposition

and the concept of genetic interaction is an example of such a strategy. In this context, the opposition between environmental and genetic diseases disappears, and, conversely, the specificity of the genes involved in human diseases comes to the fore. With network medicine, conceptualization of genetic disease has begun to move from an attempt to identify genes involved in particular conditions to the study of a specific class of genes with specific biological properties.

Whether it is called "personalized medicine" or "4P medicine," medical researchers share the dream of a preventive medicine based on genetic knowledge. Be it our future or not, the potential social consequences of the misuses of the concepts of genetic disease and genetic predisposition should draw our attention to how cautiously we should be in setting up new routine medical practices.

Definition of Key Terms

Dominant autosomal inheritance	Concerns the transmission of a gene that is not localized on sexual chromosome X or Y. Each parent gives one copy of each gene, which is called an "allele." When a trait is dominant, only one copy is required for the trait to be observed. When both alleles are necessary, the mode of inheritance is said to be recessive.
Epigenetic	Epigenetic phenomena refer to a set of heritable DNA modifications, as, for example, methylation. They change the expression of genes but do not involve any modification in the underlying DNA sequences. Prader–Willi syndrome, for instance, is related to epigenetic modifications of DNA.
Genetic linkage study	Linkage analysis relies on the co-segregation of a phenotype and specific genetic variants in families to localize a marker statistically associated with the disease.
Genome-wide association study	GWAS consists in scanning markers across genomes of many people to find genetic variations associated with a particular disease, for example, the Wellcome Trust Case Control Consortium published in 2007, a GWAS that identified genetic factors for Crohn's disease, rheumatoid arthritis, bipolar

	disorder, coronary artery disease, and type 1 and 2 diabetes.
Heritability	The portion of phenotype variance in a population attributable to genetic factors. This statistical measurement comes from the quantitative genetics introduced by Ronald Fisher.
Human Genome Project	HGP was an international project aimed at determining the human DNA sequence, identifying and mapping the human genes. It started in the 1990s and was declared complete in 2003.
Monogenic disease	A monogenic disease is considered as mainly caused by one gene, for example, thalassemia, cystic fibrosis, or Huntington's disease. It is transmitted according to Mendel's laws.
SNP	The single-nucleotide polymorphism markers are very frequent variations of one nucleotide distributed along the DNA. After the achievement of GWA studies, association studies used CNV (copy number variations) markers, but the results were rather disappointing.
The "central dogma" of molecular biology	This expression refers to a model formulated in 1956 by Francis Crick and republished in *Nature* in 1970, stating that DNA makes RNA and RNA makes proteins.
Variable expressivity	One talks about variable expressivity when the same genetic variant is correlated to different clinical phenotypes.

Summary Points

- Extending the concept of genetic disease too far empties the term of its precision and utility, running the risk that it no longer conveys any useful information, as virtually all diseases are genetic in the sense that genetic factors are at work in their development.
- Genetic medicine is today a normalized practice based on the knowledge of very high correlations between the presence of specific genetic variants and the apparition of certain diseases in a family.

- Genome-wide association study, like other association studies formerly practiced, is a black box strategy. For this reason, a genetic variant that has been identified by association studies cannot be considered a cause of a disease.
- The contemporary and somewhat arbitrary use of the concept of genetic disease is the consequence of some still more complex analyses of the human genome. It is also the consequence of the extension of our knowledge of the genetic determinants of diseases. The developments of classical Mendelian genetics also cast doubt on the concepts, taking us back to the complexity of the interactions involved in pathological pathways.
- For some sociologists and philosophers, labeling a disease as genetic refers more to a set of social and political commitments about the best way to allocate resources than to biological or medical considerations.
- The problem of causal selection may have no solution as long as philosophers will seek a unique and absolute criterion that could justify the choice of a genetic factor as the main cause of a complex biological phenomenon such as disease.
- Network medicine provides a useful framework, taking into account complexity and paving the way for redefining the concept of human disease genes as a set of specific genes involved in disease in general.

References

Badano JL, Katsani N (2002) Human genetics and disease: beyond Mendel: an evolving view of human genetic disease transmission. Nat Rev Genet 3(10):779–790

Barabási AL (2007) Network medicine: from obesity to the "diseasome". N Engl J Med 357:404–407

Barabási AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12(1):56–68

Bates GP (2005) History of genetic disease: the molecular genetics of Huntington disease – a history. Nat Rev Genet 6(10):766–773

Beurton P, Falk R, Rheinberger HJ (eds) (2000) The concept of the gene in development and evolution. Historical and epistemological perspectives. Cambridge University Press, Cambridge

Braun P, Rietman E, Vidal M (2008) Networking metabolites and diseases. Proc Natl Acad Sci 105 (29):9849–9850

Caspi A, McClay J, Moffitt TE et al (2002) Role of genotype in the cycle of violence in maltreated children. Science 297:851–854

Chen S, Parmigiani G (2007) Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol 25 (11):1329–1333

Chen Y, Zhu J, Lum PY et al (2008) Variations in DNA elucidate molecular networks that cause disease. Nature 452:429–435

Dausset J (1972) Correlation between histocompatibility antigens and susceptibility to illness. Prog Clin Immunol 1:183–210

Dausset J (1996) Predictive medicine. Eur J Obstet Gynaecol Reprod Biol 65:29–32

Dekeuwer C (2015) Defining genetic disease. In: Huneman P, Lambert G, Silberstein M (eds) Classification, disease and evidence. Springer, Dordrecht, pp 147–164

Dipple KM, McCabe ERB (2000) Modifier genes convert "simple" Mendelian disorders to complex traits. Mol Genet Metab 71:43–50

Epstein CJ (2006) Medical genetics in the genomic medicine of the 21st century. Am J Hum Genet 79(3):434–438

Friebel TM, Domchek SM, Rebbeck TR (2014) Modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: systematic review and meta-analysis. J Natl Cancer Inst 106(6):1–15
Galas DJ, Hood L (2009) Systems biology and emerging technologies will catalyze the transition from reactive medicine to predictive, personalized, preventive and participatory medicine. Interdiscip Biol Central 1(2):1–4
Gannett L (1999) What's in a cause? The pragmatic dimensions of genetic explanations. Biol Philos 14:349–374
Gifford F (1990) Genetic traits. Biol Philos 5(3):327–347
Hedgecoe AM (2002) Reinventing diabetes: classification, division and the geneticization of disease. New Genet Soc 21(1):7–27
Hedgecoe AM (2003) Expansion and uncertainty: cystic fibrosis, classification and genetics. Sociol Health Illn 25(1):50–70
Hesslow G (1984) What is a genetic disease? On the relative importance of causes. In: Nordenfelt L, Lindahl BIB (eds) Health, disease, and causal explanations in medicine. Springer, Dordrecht, pp 183–193
Hesslow G (1988) The problem of causal selection. In: Hilton DJ (ed) Contemporary science and natural explanation: commonsense conceptions of causality. Harvester Press, Brighton, pp 11–31
Hoedemaekers R, Ten Have HAMJ (1998) Geneticization: the Cyprus paradigm. J Med Philos 23 (3):274–287
Hood L (2004) Systems biology and the New technologies enable predictive and preventative medicine. Science 306(5696):640–643
Hood L, Flores M (2012) A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized, and participatory. N Biotechnol 29(6): 613–624
Kerr A (2004) Giving up on geneticization: a comment on Hedgecoe's "Expansion and uncertainty: cystic fibrosis, classification and genetics". Sociol Health Illn 26(1):102–106
Kerr A (2005) Understanding genetic disease in a socio-historical context: a case study of cystic fibrosis. Sociol Health Illn 27(7):873–896
Kitcher P (2003) Battling the undead. How (and how not) to resist genetic determinism. In: Kitcher P (ed) In Mendel's mirror. Philosophical reflections on biology. Oxford University Press, Oxford, pp p283–p300
Lewontin RC (2000) The triple helix: gene, organism and environment. Harvard University Press, Cambridge, MA
Lippman A (1991) Prenatal genetic testing and screening: constructing needs and reinforcing inequalities. Am J Law Med 17:11–50
Loscalzo J, Kohane I, Barabási AL (2007) Human disease classification in the postgenomic era: a complex system approach to human pathobiology. Mol Syst Biol 3(1):1–11
Magnus D (2004) The concept of genetic disease. In: Caplan AL, McCartney JJ, Sisti DA (eds) Health, disease, and illness: concepts in medicine. Georgetown University Press, Washington, pp 233–242
Manolio TA, Collins FS, Cox NJ et al (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753
Misra S (2013) Human gene therapy: a brief overview of the genetic revolution. JAPI 61:127–133
Nelkin D, Lindee MS (1995) The DNA mystique: the gene as a cultural icon. University of Michigan Press, Ann Arbor
Norell S (1984) Models of causation in epidemiology. In: Nordenfelt L, Lindahl BIB (eds) Health, disease, and causal explanations in medicine. Reidel, Dordrecht
Oyama S (1985) The ontogeny of information: developmental systems and evolution. Cambridge University Press, Cambridge
Reich DE, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17 (9):502–510

Smith GD (2007) Towards an adequate account of genetic disease. In: Kincaid H, McKitrick J (eds) Establishing medical reality – essays in the metaphysics and epistemology of biomedical science. Springer, Dordrecht, pp 83–110
Tabery J (2007) Biometric and developmental gene–environment interactions: looking back, moving forward ». Dev Psychopathol 19(4):961–976
Tabery J (2009) From a genetic predisposition to an interactive predisposition: rethinking the ethical implications of screening for gene-environment interactions. J Med Philos 34(1):27–48
The Welcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678
Visscher PM, Brown MA, McCarthy MI et al (2012) Five years of GWAS discovery. Am J Hum Genet 90(1):7–24
Zhou X, Menche J, Barabási AL et al (2014) Human symptoms–disease network. Nat Commun 5:1–10

Online Resources

Gene Therapy Clinical Trial Worldwide. http://www.abedia.com/wiley/. Accessed 20 Jul 2015
Online Inheritance in Man. http://omim.org/. Accessed 20 Jul 2015

Mind-Brain Dualism and Its Place in Mental Health Care

23

Eric Matthews

Contents

Abstract

This chapter begins by setting out and explaining the doctrine of "substance dualism", according to which the mind and the brain are distinct and mutually independent "substances". It then examines the merits and deficiencies of dualism, in comparison with those of alternative theories, in answering questions about the nature and treatment of mental disorder, its similarities and differences from bodily illness, and the relation between mental disorder and brain dysfunction. The alternative theories considered are the mind-brain identity version of materialism, and Merleau-Ponty's conception of human beings as "embodied subjects".

Introduction

The nature of human mental life and its relation to biological life, and especially the operations of the brain, is a central theme in general philosophy. Since health care is concerned with maintaining and restoring human well-being, both mental and

E. Matthews (✉)
The School of Divinity, History and Philosophy, University of Aberdeen, Aberdeen, UK
e-mail: e.matthews@abdn.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_15

bodily, this is also a theme for philosophical reflection about health care. Such reflection is not just theoretical: our answers to questions about the relation between the mental and the bodily have a bearing on the kinds of treatment which are appropriate for mental and bodily disorders.

Various questions arise. First, are mental disorders properly seen as "illness," in the sense in which that term is understood in modern, scientific, medicine? (The answer to this question requires thought about how terms like "illness" and "disease" *are* understood in modern medicine.) Second, if the answer to the first question is "Yes," then how is mental illness related to bodily, especially brain, disorder? Is mental illness totally distinct from brain disorder? Or are "mental" disorders just a subclass of brain diseases? Or is mental illness the *causal product* of brain disease? Or, finally, is all illness a combination of "mental" and "bodily" elements, so that the alleged distinction between mental health and bodily health is, at best, *pragmatic* – useful for some purposes, but with no sound philosophical foundation?

These are the questions which will be considered in this chapter. To prepare the ground, however, we must first very briefly survey some of the main positions adopted by philosophers on the general problem of relations between "mind" and "brain" and their implications for health care.

Descartes, Dualism, and Medicine

Much of the agenda for modern philosophy, and indeed for the modern world, including the development of scientific medicine, was set in the seventeenth century by René Descartes (1595–1650). Descartes was dissatisfied with what had been presented in his own education as "knowledge" about the world. Knowledge worthy of the name – what we would call "science" – should, it seemed to him, have a basis in what could meet the scrutiny of that *reason* with which all human beings were endowed. What passed for knowledge in his own day, however, was better described as a set of *opinions*, supported by tradition, authority, or uncritical common sense, rather than reason, that is, by evidence that anyone could *see* to be reliable when examined in an unprejudiced way. He therefore resolved to find such a reliable foundation, using a "method of doubt." The program was to subject to the most searching doubt all those beliefs which he had previously taken as established: if any reason to doubt them, however slight, could be found, then he should for the time being treat them as if they were false. If, however, he could find a belief which could not be doubted, then the grounds on which he accepted it could be taken as a reliable foundation on which a whole structure of dependable knowledge could be based.

Using the method of doubt led, Descartes thought, to only one belief which could serve as a foundation in this sense. This was the belief in his own existence as a thinking being, which could not be doubted because doubt would be self-defeating. Even to doubt the existence of other things, the things one thought *about*, one must indubitably exist oneself and have at least the capacity for thought (since doubting involves thinking). Descartes concluded from this that he could exist as a thinking being whether or not anything else existed. (Most critics would reject this inference,

arguing that from the fact that one may *doubt* the existence of one thing but may not doubt the existence of another, it does not follow that the *existence* of the latter is not dependent on that of the former). In the terminology that he used, that meant that the thinking self was a distinct "mental *substance*," where "substance" meant a being which could exist independently of anything else. Conversely, he concluded that all the objects that he could think about, including his own body, constituted an entirely separate substance, "material substance," since their existence does not require the presence of a thinking substance. A human being, therefore, was a composite of two quite independent substances, mental and material, or "mind" and "body." This is the doctrine known as "substance dualism," the conception of the world, including ourselves, as made up of two types of stuff – mental and material.

In order to be logically independent in their existence, two substances must be different in their "essences": that is, they must each have defining characteristics which are peculiar to them and not found in the other substance. Descartes argues that the essence of mental substance is *thought*, *reason*, or *consciousness*; that of material substance is *extension*, or the property of occupying space – having a spatial location, being able to move from one such location to another, having spatial dimensions, etc. Because these are distinct essences, nothing mental can have, say, a spatial location (e.g., a thought cannot be 2 cm from another thought); and nothing material can have any property which depends on thought (a boulder, for instance, cannot have a *purpose* in rolling downhill). Thus, the explanation of mental life must be different in kind from that of processes in the material world. We can explain why someone felt angry at someone else by giving his reasons for feeling that way (he found the other's words insulting, say): but the brain processes which occurred when he felt angry – the electrochemical movements in his neurons, for instance, cannot, for the dualist, be any part of the explanation of his anger. Equally, however, the feeling of being insulted cannot be part of the explanation of why just these brain processes occurred then: for that, according to dualism, we need a more "mechanistic" explanation in terms of purely physical and chemical processes.

One main reason which Descartes had for seeking reliable foundations for science was his conviction of the need to establish medicine on a sounder, more scientific, basis. The maintenance of health, he says in his best known work, the *Discourse on the Method*, is "the chief good and the foundation of all the other goods in this life" (Descartes 1985, p. 143). If we had a medicine which was based on a reliable understanding of the causes of the diseases and infirmities of body and mind, therefore, we could use it to free ourselves of these infirmities, and so make human beings "wiser and more skilful than they have been up till now" (Descartes 1985, p. 143).

As far as the infirmities of the body were concerned, reliable understanding, according to the argument given above, required seeing them as like the dysfunctional performance of a machine – a clock which could not tell the time correctly, for instance. The human (or animal) body was, after all, according to dualism, a part of matter or material substance, whose movements could be explained "mechanistically," as the result of the purposeless movements of particles of matter from one position in space to another – just like the movements of clockwork. Bodily

processes, both normal and abnormal, are just physicochemical processes governed by the laws of nature, that is, according to Descartes, the laws of mechanics (see Descartes 1985, p. 139). A good example of this, which impressed Descartes greatly, was the discovery by his English contemporary, William Harvey, of the circulation of the blood. This phenomenon could be explained mechanistically, Descartes reasoned, if one regarded the heart as a pump which pushed blood round the body. Then, certain kinds of heart disease could be seen as analogous to the failure of such a pump to be able to fulfill this function. This failure will itself be mechanically caused: so, if we can discover these causes, we can hope to devise methods of curing or preventing such maladies (see Descartes 1985, p. 316).

Descartes talks also of mental illness, but his account of the nature of mind makes it hard to explain what it consists in, how it can be explained, or how it can be treated. Mental substance, according to dualism, does not operate mechanistically: our bodies may be machines, but our minds cannot be. Their operations cannot be explained by the laws of physics, but only by reason. If we are to speak of mental *illness*, therefore, it must consist in failure to operate rationally. Does it make sense, however, to speak of a person as having reasons to think, feel, desire, or behave irrationally? In his *Fourth Meditation*, Descartes tries to explain why, despite the goodness of God, who does not wish us to be deceived and has given us the power of reason to discover the truth, we can nevertheless make mistakes. This is because, he says, "the scope of the will is wider than that of the intellect" (Descartes 1984, p. 40). We are, that is, inclined to rush to judgment about things that we do not fully understand. The will, however, is part of our minds and so governed by reason: how can we have rational grounds for rushing to form irrational beliefs? We could, perhaps, fail to use our powers of reasoning properly, especially about issues which require hard thought, and so get into difficulties, which could be described as mental *disorders*. Such disorders, however, could surely not qualify as *illness* in any medical sense, that is, in any sense where professional medical help is needed to help us overcome the difficulties. They require the help not of a doctor but of an educator (and some willingness on our part to make the effort to think more rationally).

Are Mental Disorders Illnesses?

We might, of course, simply deny that this is a problem: some would wish to deny, anyway, that there is such a thing as "mental illness," as opposed to difficulties which we may get into in our lives and which we may need help in dealing with – though not *medical* help. One of the best known of these "deniers" was the American psychiatrist, Thomas Szasz (1920–2012). In the book which first made him famous, *The Myth of Mental Illness* (1961), he proposed (as his title implies) that the whole idea of mental illness was a "myth." He had various reasons for this view. One of the most relevant, from the present point of view, was the claim that, according to ordinary usage, a condition could only be described as an "illness" if it is a deviation

from the "anatomic and genetic norms of bodily functioning" (Szasz 1972, p. 10). If that is correct, then the term "mental illness" must be self-contradictory and so logically absurd. This argument from common usage is questionable, however, since it is far from clear that common usage does restrict the application of the term "illness" to the outcomes of bodily dysfunction; and, even if it did, Szasz provides no reason for supposing that it would not be perfectly legitimate to extend common usage to include mental disorders in the class of illnesses. Szasz is also assuming that what we call "mental illness" *cannot* be the outcome of one kind of bodily dysfunction, namely, a deviation from norms of *brain* functioning. In a later work, he explicitly says, "However, diseases of the brain are brain diseases; it is confusing, misleading, and unnecessary, to call them mental illness" (Szasz 1997, p. 49). This statement is merely an expression of dogmatic dualism, however, unless it is supported by argument: one possible argument for it will be considered later, in section "The Brain and Mental Disorder."

A more subtle argument can be found in philosophers such as Christopher Boorse. In the 1970s, Boorse published a number of journal articles (Boorse 1975, 1976, 1977), in which he attempted to define concepts of health and illness in general. He later summarized his position in the light of further reflection and responses to criticism (Boorse 1997). At the heart of his view is a distinction between the concepts of disease and illness. "Disease," he claims, is a purely objective, value-free concept, which makes it the primary focus of medicine as a *science*. To call a state or process in any living organism "diseased" is, he argues, to say that, as a matter of fact, it deviates from the normal functioning of organisms of the relevant species. The normal functioning of an organ or a system is that which is in accordance with its "design." The use of the term "design" does not, he says, carry with it any evaluative connotations: to function in accordance with its design is neither good nor bad but simply to proceed in ways required for the pursuit of the goals which the organism happens to have. (In Darwinian evolutionary theory, e.g., organisms are treated as metaphorically "having the goals" of survival and reproduction.) A scientific medicine, however, is also one which seeks to *apply* science for practical human benefit (promoting human well-being), and this is where the concept of illness has its home. A disease is called an illness, Boorse then said, only if it is serious enough to be incapacitating and so undesirable for the person who suffers from it.

The concept of illness, as used in a genuinely scientific medicine, is thus logically dependent on that of disease: it is simply a kind of disease which we find undesirable. From this Boorse draws skeptical conclusions about the idea of *mental* illness. The reason for doubt seems to be that mental disorders like schizophrenia or depression, while they may qualify as "illnesses," in the sense of being conditions which we find undesirable, do not seem to be "diseases," in the sense of being objectively determinable deviations from the normal functioning of an organism in accordance with its design. Logically, not all members of a species can deviate from the design of that species, because of what "design" means. Psychological norms, however, may vary between different groups within the species: beliefs which are

considered bizarre in one society, for example, are considered perfectly normal in another. Nothing objective, like the possibilities of survival of a species, seems to depend on whether people, for instance, deludedly believe in witchcraft. If correct, this argument implies that what we call "mental disorder" is determined not by scientifically establishable facts but by purely subjective value judgments – by what people in any particular society generally regard as "bizarre" beliefs, or behavior, for instance.

Another possibility is to say that at least some recognized mental disorders are illnesses in exactly the same sense as bodily disorders, because they represent biologically harmful dysfunctions. One much discussed version of this view was proposed in the 1970s by the psychiatrist Robert Kendell (1975). Kendell defined an illness as a deviation from normality which conferred "biological disadvantage" (Kendell 1975, p. 310). Obvious examples of biological disadvantage, he thought, were increased mortality and reduced fertility, but "other impairments" (loc. cit.) might also be included. So any failure of normal functioning, bodily or mental, which could be shown to confer such disadvantages constituted an illness. To call something a "biological disadvantage" is, of course, to make a value judgment, though one with which most human beings in all cultures would probably agree. To say that some dysfunction confers biological disadvantage is thus to say something which is objectively (scientifically) verifiable. The rest of Kendell's argument consists in giving examples of recognized mental disorders which can allegedly be shown to confer biological disadvantage and so to qualify as "illness" in a straightforward medical sense. Schizophrenia, for instance, can, he claims, be shown to lead to reduced fertility. One problem with this kind of argument, however, is that it is easy to find examples of recognized mental disorders which do not seem to confer any of the *biological* disadvantages which Kendell lists, though they do confer what may be regarded in some cultures but not others as *social* or *psychological* disadvantages.

A more philosophically sophisticated attempt along the same lines is to be found in a number of articles published in the 1990s by the American philosopher Jerome Wakefield. Wakefield defines a disorder as a "harmful dysfunction." "Functions" here mean "biological functions," which are said to be "designed by nature" in a sense determined by Darwinian natural selection: that is, an "internal mechanism" is said to perform its function in so far as it tends to ensure individual and species survival. Success or failure in performing functions can thus be verified by objective scientific evidence (see, e.g., Wakefield 1992, 2000, 2009).

The question then is whether recognized *mental* disorders can be fitted in to this analysis. To do so would involve showing that generally accepted examples of such disorders can be *explained* by the failure of some internal mechanism(s) to perform their functions as designed by nature (in the sense explained above). To do this, however, would involve abandoning mind-brain dualism of the Cartesian kind, since that is incompatible with the existence of any kind of mental "mechanisms" which might be accounted for by natural selection. To proceed any further, therefore, we need to return to the general philosophical issue of the nature of our mental life and its relation to the brain.

The Brain and Mental Disorder

An objection to dualism which emerged very early (and was in a way admitted by Descartes himself) is that we cannot ultimately separate "mind" and "brain," because many of our mental operations depend in one way or another on the normal functioning of the brain. Brain damage, for instance, leads to loss of memory or even a change in personality. Many of the symptoms of mental illness, such as delusions, can equally be the result of problems in brain functioning. As scientific understanding of the brain has developed over the last three centuries, more and more such interactions between brain functioning and the character of our thoughts, feelings, desires, and behavior have been discovered. The very possibility of such reactions, however, seems to be ruled out by the dualist view that our "minds" are a separate "substance," with a distinct essence, from our "bodies," including our brains. To account for them, therefore, seemed to require a philosophical shift from thinking of a human being as composed of two substances, mind and body, to thinking of ourselves as composed of a single substance: our mental functioning, it seems, must be just part of our bodily or biological functioning – in particular the operations of our *brains*. This is the position known as "classical materialism."

One advantage of materialism was that it seemed to eliminate the dualist problem of interaction: the influence of mind on body and body on mind became not the unintelligible influence of one substance on another but simply the influence of one part of a substance on another part of the same substance. Our "minds" could then be liable to disease or illness in exactly the same sense as, say, our hearts or livers. In effect, so-called "mental illness" would be just one kind of bodily illness. This view is made explicit in the Introduction to the Fourth Edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*: "anachronism of mind/body dualism unfortunately implies a distinction between "mental" disorders" and "physical" disorders that is a reductionistic anachronism of mind/body dualism" (American Psychiatric Association 1994, p. xxi).

A connected benefit of abandoning dualist conceptions of mental disorder was believed to be that it made it possible to bring psychiatry into line with the scientific approach of the rest of modern medicine. If the mind is equated, not with some immaterial substance distinct from anything else in the created universe but with the brain and its operations, then it seems possible to explain mental disorder as the harmful outcome of a dysfunction in "internal mechanisms" operating in the brain and nervous system. The philosopher Dominic Murphy even defines psychiatry as "a branch of medicine dedicated to uncovering the neurological basis of disease entities" (Murphy 2006, p. 10). This approach to psychiatry is often labeled "biological psychiatry." These brain dysfunctions can then be related, in much the same way as heart dysfunctions, to biochemistry, genetics, and other sciences, to form part of the unified scientific picture of the world which modern science is believed to aspire to. Descartes had proposed such a unified science for the *physical* world, but excepted the mental sphere from it: materialism goes further, to include the mental sphere in the physical world. The philosopher Paul Churchland, an advocate of "eliminative materialism," according to which our common-sense (essentially

dualist) account of psychological phenomena will eventually be displaced by "a completed neuroscience," argues that such materialism offers a more coherent and effective account of mental phenomena in general and mental disorder in particular. We can do far better, he argues, in understanding and so dealing with psychological problems by linking them to the structures, physiology, chemistry, and genetics of the brain than by thinking of psychology in "common-sense" terms ("folk psychology") (see Churchland 1981).

Especially in American psychiatry, the attraction of this materialist philosophy was that, in the ways which Churchland suggests, it seemed to restore psychiatry to a scientific approach which was thought to have been abandoned during the long dominance in the USA of Freudian psychoanalysis, which was regarded as unscientific. Thus, the historian of psychiatry, Edward Shorter, wrote that "Biological psychiatry ... became able to investigate the causes and treatments of psychiatric illness by using the scientific method, a method other psychiatrists had virtually abandoned for half a century" (Shorter 1997, p. 272). Although Freud himself does seem to view his work as "biological psychiatry," he does not in practice explain mental disorder in terms of the kinds of brain abnormalities which Churchland refers to and thus has come to be regarded by many psychiatrists and philosophers as essentially unscientific. The implication of materialism in this sense is that the so-called "mental" disorders can be treated not by Freudian "talking cures" but by manipulating their physical and chemical causes.

Opponents of biological psychiatry include Thomas Szasz, who, as quoted earlier, wanted to distinguish brain disease from the so-called "mental illness." Such a distinction requires the support of a philosophical argument against the materialism which identifies the two kinds of disorder. One argument of this kind which is often used goes as follows. Mind cannot be identified with the brain, nor can talk of mental phenomena be replaced without loss by talk of brain states and processes, because thoughts, emotions, desires, and other mental phenomena have certain essential properties which brain states and processes cannot have. The two properties are *subjectivity* and *intentionality*. Thoughts, emotions, etc., must be "subjective," in the sense that they must be *someone's* thoughts (my thoughts, your thoughts, his or her thoughts). They are necessarily identifiable, therefore, not only by their content (e.g., "that dualism is false") but also by the "subject" or person who has a thought with this content: if you and I both think that dualism is false, then there exist two thoughts with the same content. This is not the case with brain states: a brain state, such as the firing of certain neurons, is identified completely by its physicochemical properties and not by which brain it occurs in. The same brain state may thus occur in two (or a million) different persons or even, in the right circumstances, such as a laboratory experiment, when the neuron in question is not part of *any* living human brain.

The other feature of mental states and processes which is emphasized by this kind of anti-materialist argument is "intentionality." This medieval term was reintroduced into the philosophical vocabulary above all by the nineteenth-century Austrian philosopher Franz Brentano (see Brentano 1973). "Intentionality" essentially refers to the relation of consciousness to what it is about (its "intentional object"). The

claim is that anything mental is necessarily directed toward an object – it is "about" that object. Thus, one cannot think without thinking *of* something or *that* something: a thought might be identified, for example, as being a thought *of* Paris or *that* Paris is a beautiful city. Similarly, any emotion must be directed *toward* someone or something: I love my wife; I am afraid of terrorism; I admire bravery. Again, one cannot desire without desiring *someone* or *something*: I want that picture, I long for your return, and so on. The argument of the anti-materialists assumes that brain states and processes are not intentional in this sense: again, a brain state like the firing of a neuron is defined entirely by its physicochemical properties and is not "intentional," that is, it is not defined by being *about* anything. We cannot identify my thought that dualism is false, for example, with any particular set of occurrences in my brain (even though I can't have this, or any other, thought *unless* something goes on in my brain).

What is held to follow from this, if it is correct? From the present point of view, the most important conclusion is that mental disorders (disorders of thought, emotion, desire, etc.) cannot be completely or satisfactorily explained by brain dysfunction. Brain dysfunctions do not involve subjectivity or intentionality and so cannot explain these essential features of mental disorder. To use a particular example, the mental disorder agoraphobia consists in fear of open spaces, and that fear is necessarily experienced by someone (it is subjective), and it is defined by what it is fear *of* (its intentional object is "open spaces"). It may be the case that someone experiences agoraphobia only when their brain is in a certain state, but being agoraphobic involves more than being in that brain state: it also requires that the *person* as a whole experiences certain emotions about his or her environment. What is required for an explanation of how someone comes to be in any mental state, including one which is disordered in the psychiatric sense, is precisely something which will answer the question why the person is in the relevant subjective state, defined by a certain intentional relation to their world. The state of her brain cannot by itself answer that "why" question. We need also to know about the person's *reasons* for having these fears about being out of doors.

This has implications for psychiatric treatment, since appropriate treatment must depend on the way in which we explain the occurrence of the disorder being treated. Altering the patient's brain state (for instance, by administering medication) cannot target the irreducibly subjective aspect of her disorder: only engaging with her reasons for having those problems can hope to do that. This is the essence of the case made by some opponents of biological psychiatry, such as the psychiatrist R. D. Laing (1971, 2010) and the clinical psychologist Richard Bentall (2004, 2009). Laing was mainly concerned with schizophrenia, which he approached from the direction of existential phenomenology, rather than that of clinical psychiatry. His approach concentrates on a sympathetic understanding of the subjective personal experience of the patient, rather than on the biological or chemical causes of the current state of the patient's brain. The patient's condition is seen as expressing an individual response to the problems of his or her existence as a human being, rather than as symptoms caused by dysfunction in his or her brain. Bentall's approach is somewhat different. He is happy to accept the relevance of biological (e.g., genetic)

factors in predisposing individuals to mental abnormality of various kinds; but he attaches more importance to the "environmental" or "psychological" elements in the etiology of mental disorder. The assumption that "mental illnesses are genetically influenced brain diseases" has been, he argues, "a spectacular failure" (Bentall 2009, p. 264). It has failed, in that it has contributed very little to relieving the suffering of those with the severest forms of mental disorder. A psychological approach, by contrast, would recognize, Bentall argues, "that distress in human beings is usually caused by unsatisfactory relationships with other human beings" (Bentall 2009, p. 265). That is, Bentall, like Laing, maintains that successful treatment of the mental distress of human beings must be based on a conception of that distress as a subjective response to problems which those human beings experience, rather than as caused by a breakdown in their internal brain mechanisms.

An Alternative Account of Brain-Mind Relations

It can be argued that Cartesian dualism and classical materialism, despite their obvious differences, have something important in common and that this common element is responsible for the problems in thinking of the relation between bodily and mental illness which have been raised. Putting it briefly, the common element is a conception of the question to be asked. The question is taken to be this: in saying that human beings have a mind, are we saying that this "mind" is a thing (or "substance") distinct from and independent of the brain, or are we saying it is identical with the brain? Whichever we say, we are assuming that the term "mind" refers to a substance. Descartes's formulation of dualism makes this assumption explicit. Classical materialism is less explicit but clearly implies that "mind" refers to a thing, in identifying the mind with the brain. For this reason, some recent philosophers have called classical materialism "Cartesian materialism."

Such critics of a medicalized psychiatry as Laing and Bentall can be seen as dualist in spirit, even if they officially reject Cartesianism and accept that brain states and brain processes have some relevance to the explanation of mental disorder. Their conception of mental disorder, however, rules out attributing any *central* importance to brain dysfunction. Laing, for example, argues that the issues lived through by people with schizophrenia "cannot be grasped through the methods of clinical psychiatry and psychopathology" but "require the existential-phenomenological method to demonstrate their true human relevance and significance" (Laing 2010, p. 18). In similar vein, Bentall criticizes "biological investigators" for failing to consider the possibility "that their findings might reflect the tribulations of life, rather than some lesion or genetic scar carried by the victim from birth" (Bentall 2009, p. 152). If mental disorders are seen in this way, as human responses to certain kinds of problems in life, then it seems we must explain them in terms of patients' *reasons* for finding certain situations insuperable problems, rather than in terms of the failure of their brains to function in biologically appropriate ways – that is, to offer a dualist or "mentalistic" explanation rather a materialist or "physicalistic" one.

The arguments for one alternative in the debate between dualism and materialism largely consist in objections to the other. As seen earlier, materialism is argued to be superior to dualism, for example, because it makes "mental" operations like thinking identical with brain operations, which are objectively accessible, and so that mental disorder is a neurological problem which can be dealt with by a scientific medicine. On the other hand, the brain-mind identity thesis seems incompatible with the widely held view that minds have properties (subjectivity, intentionality) which brains do not have, and so that mental disorder needs to be empathically understood, as a human problem, rather than causally explained as a breakdown in brain mechanisms. The conclusion seems to be, then, that neither dualism nor materialism is entirely satisfactory.

Is there another possible way of approaching the relation of mind and brain, which might offer the prospect of a more satisfactory conception of the role of that relation in thinking about health care? This essay will conclude by considering one such alternative. Some philosophers have suggested recently that the problems arise only because we start from an unexamined assumption: putting it at its simplest, it is the assumption that "mind" refers to a thing, so that the only question is, is "mind" a *separate* thing from "brain," or are they one and the same thing? Perhaps, if we abandoned that assumption, we could also avoid the difficulties just mentioned.

One way of arriving at this position is *linguistic*, to be found, for instance, in Gilbert Ryle's book *The Concept of Mind* (Ryle 1949). Ryle proposes that we avoid asking abstract questions like "What kind of thing is a mind?" and instead examine how we use words like "mind," "mental," etc., in ordinary language. If we do this, he argues, we will conclude that "mind" is not the name of a thing, but a way of classifying a wide variety of human activities, capacities, and other dispositions. An alternative approach to a similar position is by means of *phenomenology* – the philosophical method introduced and developed by Edmund Husserl (1859–1938). Central to this method is the attempt to avoid all "presuppositions" (e.g., that a mind is a "thing" which human beings have) and instead to describe phenomena as we actually *experience* them, avoiding as far as possible assumptions derived from science or previous philosophy. Husserl himself laid some of the foundations for a new approach to mind-body relations along these lines, most notably in his posthumously published work, *The Crisis of European Sciences* (Husserl 1970). The phenomenological approach was most fully developed, however, by the French philosopher, Maurice Merleau-Ponty (1908–1961). Merleau-Ponty's approach to these issues has also influenced some recent philosophers in the analytic tradition, such as Shaun Gallagher (see 2005; Gallagher and Zahavi 2008) and Andy Clark (see 1997). Matthew Ratcliffe's work is also relevant, especially his (2008).

A phenomenological approach, as said above, seeks to set aside all assumptions, for instance, about what a "mind" is, and to consider our actual experience, as far as possible in a "presuppositionless" way. We experience minds, both our own and those of other human beings, in our dealings with other people. For instance, we converse with others, expressing our own thoughts and hearing and responding to theirs. So we experience minds in experiencing ourselves and others as *subjects*. To

experience someone as a subject, however, is necessarily to experience them as *embodied*. A subject is a being who relates to the world both in the way material objects do (being spatially and causally related to other objects) and also in *experiencing* the world – relating to objects in finding them *meaningful* to him- or herself. A simple example would be that a human being may relate to, say, an apple, not only because the light reflected from the apple causes him/her to see the apple but also in that he or she perceives the apple as having such meanings as "good to eat," "esthetically attractive object," and so on. Being a subject and being embodied are two sides of the same coin: we can only experience the world subjectively because we are embodied in a particular way (we have senses and have a physical location in space and time from which we perceive things) and the way in which we are in the world objectively is not like the way an inanimate (i.e., "subjectless") object is, because, as active subjects, we find *meaning* in the objects (including other people) around us. We are thus essentially *embodied subjects*. Merleau-Ponty's fullest and clearest development of this view can be found in his major work, *Phenomenology of Perception* (Merleau-Ponty 2012: especially Part One and references in Index to "embodiment/incarnation").

Starting with the notion of human beings as embodied subjects offers the possibility of a totally different way of thinking about mind-brain relationships and their relevance to the treatment of both mental and physical disorders, from that implicit in either Cartesian dualism or Cartesian materialism. In this understanding, thinking, feeling, desiring, wishing, intending, hoping, remembering, and the behavior, which is explained by them, are activities of neither "minds" nor "brains" but of *human beings*. Because human beings are embodied, their responses to their environment necessarily involve bodily reactions, especially brain processes, changes in brain chemistry, etc., but these bodily reactions themselves can be fully understood only as part of the human response. A human being may, for instance, feel suicidally depressed: if so, the serotonin levels in his or her brain would characteristically be lowered. This change in serotonin levels, however, does not *explain* the depression, on this view: rather, it is part of what has to be explained. To explain why someone feels in such a mood, or any other mental state, requires us to explore what it is about the situation which leads them to see it as they do (e.g., depression may be a response to a dramatic breakdown in a close relationship). Psychotherapeutic modes of treatment would thus be central. At the same time, however, we cannot ignore the fact that depression, in a human being, necessarily involves changes in serotonin levels, so that medications which affect those levels may alleviate depressive mood. This is also why, it might be suggested, in some cases of what we should normally call "bodily" disorders, we can, for the same reason, cite psychological responses to difficult human situations as playing a significant role, because of our embodiment, in leading to the physical problems involved. An example might be paralysis, as part of a response to psychological trauma. Merleau-Ponty discusses a number of spatial and motor disorders along these lines in Merleau-Ponty 2012, pp. 100–148.

References

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, fourth edition, (DSM-IV). American Psychiatric Association, Washington, DC
Bentall RP (2004) Madness explained: psychosis and human nature. Penguin, London/New York
Bentall RP (2009) Doctoring the mind: why psychiatric treatments fail. Allen Lane, London/New York
Boorse C (1975) On the distinction between disease and illness. Philos Public Aff 5:49–68
Boorse C (1976) What a theory of mental health should be. J Theory Soc Behav 6:61–84
Boorse C (1977) Health as a theoretical concept. Philos Sci 44:542–573
Boorse C (1997) A rebuttal on health. In: Humber JF, Almeder RF (eds) What is disease? Biomedical ethics reviews. Humana Press, Totowa, pp 1–134
Brentano F (1973) Psychology from an empirical standpoint (trans: Rancurello AC, Terrell DB, McAlister LL). Routledge/Kegan Paul, London
Churchland PM (1981) Eliminative materialism and the propositional attitudes. J Philos 78:67–90
Clark A (1997) Being there: putting brain, body and world together again. The MIT Press, Cambridge, MA/London
Descartes (1984) The philosophical writings of Descartes, Vol II (trans: Cottingham J, Stoothoff R, Murdoch D). Cambridge University Press, Cambridge/London/New York
Descartes (1985) The philosophical writings of Descartes, Vol I (trans: Cottingham J, Stoothoff R, Murdoch D). Cambridge University Press, Cambridge/London/New York
Gallagher S (2005) How the body shapes the mind. Oxford University Press, Oxford/New York
Gallagher S, Zahavi D (2008) The phenomenological mind: an introduction to philosophy of mind and cognitive science, 1st edn. Routledge, Oxford/New York
Husserl E (1970) The crisis of European sciences and transcendental phenomenology: an introduction to phenomenological philosophy (trans: Carr D). Northwestern University Press, Evanston
Kendell RE (1975) The concept of disease and its implications for psychiatry. Br J Psychiatry 127:305–315
Laing RD (1971) Self and others. Pelican Books, Harmondsworth/Baltimore
Laing RD (2010) The divided self: an existential study in sanity and madness. Penguin, London/New York (Penguin Classics edition)
Merleau-Ponty M (2012) Phenomenology of perception (trans: Landes DA). Routledge, London/New York
Murphy D (2006) Psychiatry in the scientific image. The MIT Press, Cambridge, MA/London
Ratcliffe M (2008) Feelings of being: phenomenology, psychiatry and the sense of reality. Oxford University Press, Oxford/New York
Ryle G (1949) The concept of mind. Hutchinson, London
Shorter E (1997) A history of psychiatry: from the era of the asylum to the age of prozac. Wiley, New York
Szasz T (1972) The myth of mental illness: foundations of a theory of personal conduct. Granada Publishing, St. Alban's
Szasz T (1997) Insanity: the idea and its consequences. Syracuse University Press, Syracuse/New York
Wakefield JC (1992) Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder. Psychol Rev 99:232–247
Wakefield JC (2000) Aristotle as sociobiologist: the 'function of a human being' argument, black box essentialism, and the concept of mental disorder. Philos Psychiatry Psychol 7(1):17–44
Wakefield JC (2009) Mental disorder and moral responsibility: disorders of personhood as harmful dysfunctions, with special reference to alcoholism. Philos Psychiatry Psychol 16(1):91–99

Memory, Identity and Dementia

24

Simon Walker and John McMillan

Contents

Abstract

People with dementia often undergo profound changes in attitudes and behavior. Following such changes it can be difficult for carers to know whether or not to apply advance directives: should the person's apparent current choices be respected or should they be overridden by choices that were made before the dementia developed? This question raises further questions about the nature of autonomy and identity and how these are affected by the various kinds of functional damage caused by dementia. Ronald Dworkin answered this question by arguing that identity is based around what he terms "critical interests" and that these interests are formed prior to the onset of dementia, which means that prior decisions should be applied. Agnieszka Jaworska has criticized this view and argued that people with dementia retain the capacity to form new critical interests because they retain the capacity to value. Christine Korsgaard's account

S. Walker (✉)
Bioethics Centre/Te Pokapū Matatika Koiora, University of Otago, Dunedin, New Zealand
e-mail: simon.walker@otago.ac.nz

J. McMillan
Bioethics Centre, University of Otago, Dunedin, New Zealand
e-mail: john.mcmillan@otago.ac.nz

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_64

of personal identity shows that both Dworkin and Jaworska are partly right, but also that judgments about whether or not a given decision is in keeping with a person's identity may be yet more difficult than either suggest. This indicates the need for further philosophical and ethical analysis of this area of medicine.

Introduction

The term dementia is used to describe a spectrum of neuropsychological disorders characterized by a significant loss of cognitive functions, particularly memory, planning attention, and judgment. As these disorders are more prevalent among older people they present a growing challenge to health-care systems in developed societies where the proportion of older people is steadily increasing (Nuffield Council on Bioethics 2009, esp. 4–5). Within this challenge are a range of philosophical difficulties, several of which center on questions about the nature of personal identity and how it is affected by the loss of memory. While the significance of these questions is perhaps most clearly evident in the ongoing debate regarding the status of decisions (past or present) expressed by patients who are suffering profound memory loss, they also bear upon the everyday duties of those caring for such patients and the grief and estrangement these carers often feel in relating to a friend or family member who seems in many ways radically changed (Hope and McMillan 2011). The discussion in this chapter sets out a broad philosophical response to these questions and indicates some ethical implications and directions for further research.

Dementia, Interests and Agency

The progression of dementia can cause profound changes in attitudes, behavior, and quality of life, and these changes can generate a range of ethical questions, the starkest being whether to withhold or withdraw medical treatment for the purpose of not prolonging life. The problem of personal identity becomes relevant to these questions when there is a disagreement between the patient's prior choices and the "decisions" they seem to be making now. The well-known case of "Margo," described by a medical student named Firlik and recounted by Ronald Dworkin, illustrates this tension well.

> The apartment had many locks to keep Margo from slipping out at night and wandering in the park in a nightgown, which she had done before. Margo said she knew who Firlik was each time he arrived, but she never used his name, and he suspected that this was just politeness. She said she was reading mysteries, but Firlik "noticed that her place in the book jumps randomly from day to day; dozens of pages are dog eared at any given moment. . . Maybe she feels good just sitting and humming to herself, rocking back and forth slowly, nodding off liberally, occasionally turning to a fresh page." Firlik was confused, he said, by the fact that "despite her illness, or maybe somehow because of it, Margo is undeniably one of the happiest people I have ever known." (Dworkin 1993, 221)

Margo suffers a moderate form of dementia which has caused substantial cognitive impairment, to the extent that she requires daily monitoring and care. Dworkin asks us to consider what should be done if before developing dementia Margo had written an advance directive refusing medical treatment if she were to be in such a condition (Dworkin 1993, 226). However, there is a question about whether this advance directive should be applied, and this question is sharpened by the fact that despite her profound cognitive decline (or perhaps somehow because of it), Margo seemed to Firlik to be living well and contented.

It is not immediately clear how current well-being should factor in decisions about applying advance directives. There is a sense that given Margo is currently happy it is wrong not to make efforts to extend her life. If this were the sole source of uncertainty, it would follow that if Margo was especially unhappy, there should be no hesitation in applying the advance directive (see, e.g., Dresser 1995). If there is hesitation (as for many there is), then it must be because there is something other than present happiness that is relevant to these decisions. Dworkin's view is that the advance directive should be applied irrespective of Margo's happiness, because the advance directive reflects what he describes as her "critical interests," which he relates to the "shape and character" of a person's life and to her "sense of integrity" (Dworkin 1993, 213). Though this is a complex idea that requires careful analysis, it highlights at least two useful and comparatively uncontentious ideas: firstly that some interests (or values) are more important to a person than others and secondly that certain important interests are identity defining, i.e., they somehow constitute the person as the person who she is. In other words, a critical interest is an interest that one would not willingly or perhaps readily forego and which if unmet may make the person "less herself." Hence, Dworkin contrasts critical interests with "experiential interests": interests that are worth meeting where possible but which are transitory (not identity defining) and which therefore occupy a second order of priority in a person's decisions. The main implication of Dworkin's position is that when thinking about respecting a person's autonomy critical interests must take priority over experiential interests. So, if the decision about whether or not to apply the advance directive is based on the need to respect the patient's autonomy, then, argues Dworkin, it is necessary that one identifies what the person's critical interests are, and if these align with the advance directive, then the advance directive should be applied, irrespective of her current happiness (which is conditioned by transitory experiential interests).

While Dworkin's distinction between critical and experiential interests is clear in some respects, it raises a number of important questions. To begin with, it is not immediately obvious how, or to what extent, a critical interest is identity defining. Related to that is the question of why a critical interest is more important than a non-critical interest. Also closely related is the question of how critical interests are established. Dworkin does not offer a theoretical explanation of the link between critical interests and identity, but rather describes several examples that illustrate the apparent link (Dworkin 1993, 209–212). Critical interests are important, he says, because they most closely reflect a person's autonomous choices and are the primary indication of what is in a person's "best interests" (given that they express

the person's important values) (Dworkin 1993, 226–229). They are established, he argues, through a person's rational assessment of what gives integrity to her life – what makes her life "successful as a whole." From this position, Dworkin concludes that as dementia advances a person becomes unable to form new critical interests. This is because such people have:

> . . .lost the capacity to think about how to make their lives more successful on the whole. They are ignorant of self – not as an amnesiac is, not simply because they cannot identify their pasts – but, more fundamentally, because they have no sense of a whole life, a past joined to a future, that could be the object of any evaluation or concern as a whole. They cannot have projects or plans of the kind that leading a critical life requires. They therefore have no contemporary opinion about their own critical interest. (Dworkin 1993, 230)

Hence in a case like Margo's, one should abide by the advance directive unless there are good reasons to think that she has changed her position sometime between her writing the advance directive and losing her capacity to think of her life "as a whole," as this respects her autonomy and serves her best interest, irrespective of her current quality of life.

An important challenge to Dworkin's position has been put forward by Agnieszka Jaworska. She agrees with Dworkin's claim that one should respect the important values of dementia sufferers but rejects the idea that those values (critical interests) can only be generated through the kind of rational deliberation Dworkin describes (Jaworska 1999, 116). She argues that so long as a person continues to hold values, they should be regarded as expressing critical interests. She defines a value as a view about how one should act (the view that the action is "correct" for oneself), which is related to one's self-esteem (in that how one feels about oneself is tied to whether or not one performs this action), and which is grounded in reasons that hold independently of one's own experience (i.e., one does not consider the action correct simply because it makes one feel good). She illustrates this position through several examples in which patients with dementia hold firm views about how they should act and who link these ideas to their own self-worth (and hence their sense of who they are). To support her claim, she points out that the neurological and cognitive deficits suffered by those with dementia are not centrally related to the capacity to value, but rather to the ways that one knows *how* to realize values (Jaworska 1999, 122). Knowing how to do what one thinks one should do is, she points out, something that everyone struggles with by degrees; the difference for people with dementia is that they have more difficulty than others (not that they are unable to form and express such views). People with dementia, Jaworska concludes, should be supported in enacting their contemporary critical interests, just as is often done for people in a variety of other disabling conditions.

In her argument Jaworska emphasizes the link between valuing and agency. Moral philosophers have long discussed this link and its relationship to identity – a connection that is partly reflected in the value that is placed on autonomy (see, e.g., Korsgaard 1996a, 128–131). On this view, to negate agency is to negate the person. The question is: what is human agency and how is it sustained or altered over the

course of a person's life and through the kind of changes brought about with dementia? It seems that for Dworkin agency is expressed through rational deliberation on the overall purpose of one's life, i.e., it is based on ideas that a person has about who she is and what her life is for. Jaworska's view is simpler: in her view agency is expressed simply through the sense that some actions are "right" for that person. The depth or coherence of the ideas supporting that sense is of secondary consideration. So, when Margo chooses a certain kind of sandwich, to participate in an art class, or to sit and "read" a book, she is exercising her agency, and hence these choices should be regarded as expressing important values and as reflecting what is centrally important in Dworkin's notion of "critical interests."

Two Examples

An appealing aspect of Jaworska's position is that it seems to "humanize" those with dementia, in that it affirms the value of the lives they are evidently still "living." In doing this, it appears to resolve the tension many feel about disregarding the contemporary happiness of a demented person on account of what may be seen as a heartless commitment to formalized choices. However, the implications of her position are not always favorable in this way, as sometimes the contemporary "values" of the demented person are sharply at odds with his or her prior values or seem simply "wrong" (c.f. Jaworska 1999, 136). This point may be illustrated through two comparable but contrasting cases. The first is from Jaworska:

> Mr. O'Connor was a deeply religious man for whom thoughts of taking his own life or of withholding lifesaving measures for whatever reason were completely unacceptable. In his seventies he developed Alzheimer's disease. He lost his ability to do many of the things he used to enjoy, such as playing the piano; soon he could no longer take care of himself. With the loss of capacity for complex reasoning, most of his religious beliefs gradually faded away. Then came a terrible emotional blow: the death of his wife. He has now begun saying that he does not want to go on, that he does not want to live. His daughters no longer know what would be best for him: Should they make decisions for him based on his earlier life-philosophy or should they respect and take seriously his current wishes? (Jaworska 1999, 107)

The second case is from an article by Tony Hope:

> Mr D taught classics at a boy's preparatory school. He loved music and played the piano. He married at 25 years of age and has had two daughters. He retired at age 63 because his workload was beginning to become too much for him. Soon after retirement he was diagnosed with Alzheimer's disease. At first he enjoyed retirement, but life gradually became more difficult. After about a year he became less affectionate to his wife. Things deteriorated to the point where, for the first time in their married lives, they slept in separate beds. As time went on he started to do less and less. His indifference to his wife started to become active hostility and this hostility became directed to his daughters as well. About 4 years after retirement he wandered out of the house when his wife was out shopping.

> The police brought him back after complaints that he was 'molesting little girls'. Apparently he had said some things to a group of schoolgirls in a shop and the shopkeeper was concerned and called the police. His physical aggression has never extended beyond pushing his wife away. He can feed himself and is fully continent. There are times when he appears to know who she is, but for much of the time he does not appear to know her.
>
> His wife's attitude towards him has changed radically over the years. At first she did not see the changes as being the result of an illness – she thought that he had ceased to love her. The second stage was when she accepted he was ill and that he needed her help. But over the last year her attitude has changed again... She says that he must go into a home permanently. 'I don't see why I should have him in the house at all. It's like living with a stranger. He's not the man I married – that man has been dead for at least two years. (Hope 1994, 133)

These cases are similar in that they both involve radical changes in beliefs and feelings and yet also differ in several significant respects. In the first case Mr. O'Connor retains a deep emotional commitment to his wife despite the loss of his cognitive faculties. Jaworska argues that the grief he feels and his consequent unwillingness to accept medical treatment indicate contemporary values that should be respected over his prior beliefs (Jaworska 1999, 120–121). Like Mr. O'Connor, Mr. D expresses views about what is right for him to do, but these bear even less connection to his previous life: he retains no emotional connection to his wife, and his behavior has changed to the extent that it seems to her as though he is an altogether different person. If this was to be taken literally, then the question of applying the advance directive would not arise; one would rather have to assess the autonomy and interests of this new "Mr. D." This conclusion is problematic at many levels: ethical, legal, and ontological (McMillan 2005; Buchanan 1988). A simpler interpretation would be that Mr. D's volitional capacity has somehow been overthrown by his disease and that in his case an advance directive should be applied – an idea that is also expressed by his wife. However, this conclusion does not easily fit with Jaworska's view of autonomy, as Mr. D seems to exhibit the same volitional capacity as Mr. O'Connor. If Mr. O'Connor is capable of some form of autonomous choices, then it appears that Mr. D is too. To avoid this conclusion it would be necessary to show that there is some significant difference between the two men. This requires a closer examination of the nature of agency and its relationship to identity.

Agency, Identity, and Coherence

Focusing simply on the superficial features of human agency, one might conclude that there is no significant difference between Mr. O'Connor and Mr. D. Both are expressing a form of choice, and so Jaworska's criteria may be said to be autonomous. If there is a perceived difference, perhaps it is based on the fact that the actions of Mr. D seem distasteful or even wrong, whereas in the case of Mr. O'Connor, one can more readily sympathize with his choices. There is a risk, in other words, that Mr. O'Connor's values are given more weight merely because

most people agree with them and want them to be the values that he has. These differences – the argument might go – may be ethically significant, but they are not relevant when assessing decision-making capacity. The problem with this is highlighted in Dworkin's initial conception of critical interests: people (consciously or not) endeavor to impose some degree of coherence over the course of their lives, and as part of this, they hope that the final stages of life will be somehow in keeping with what has proceeded. "Death" writes Dworkin, "... is not only the start of nothing but the end of everything, and how we think and talk about dying – the emphasis we put on dying with "dignity" – shows how important it is that life ends *appropriately*, that death keeps faith with the way we want to have lived" (Dworkin 1993, 199). Mr. D's behavior having developed dementia is sharply at odds with his previous life, which is of course why his wife describes him as a "different" person and why one might be inclined to negate his apparent agency and, in so far as possible, steer him in a different direction. To accept his current behavior as even minimally autonomous means accepting that his current life is a betrayal of the value that his life has attained, and this would for many people be a kind of moral tragedy. One may perhaps avoid this bleak conclusion by relating identity to the notion of moral coherence, a relationship that is developed in Christine Korsgaard's moral philosophy.

Korsgaard argues that the self is primarily constituted through agency and that the hallmark of human agency is the ability to order one's life according to an understanding of what it is that gives one's life value (Korsgaard 1996b, 120–125). She describes such an understanding as a "practical identity." According to this view, a person acts *as herself* when she acts out of a conscious endorsement of the reasons underpinning the given action and not simply on blind impulse or desire (see also Korsgaard 2009). When these conditions are met, a person places a kind of ownership over his actions. He will accept credit or blame (if he is honest) for the consequences of the action, and if he is asked why he did what he did, he will identify the reasons in his explanation. By contrast, a person fails to act – and fails to be "herself" – when she acts without any sense of why she does what she does or when she is compelled to act but cannot endorse the reasons she is acting upon. There are a variety of factors that might cause a person to fail in this sense, such as ignorance, or because none of the possible actions align with the person's practical identity. A child, for instance, may act angrily because she is tired without being aware of this. In such cases one does not usually regard the child as wholly autonomous, but instead attempts to steer her toward a more appropriate behavior. In order for the child to become autonomous in such situations, she must learn to understand how tiredness affects her thoughts and perception and to not falsely attribute the cause of her "bad feeling" (i.e., her tiredness) to those others with whom she was angry. This is a simplification of a developmental process everyone must go through. It is of course not likely to be only the feelings of tiredness that are causing the child's anger, and simply being aware of the cause is usually not enough to achieve self-mastery (most would also need to have a rest). Attaining this kind of awareness and appropriate self-management is usually a lifelong task. Similarly, if a person is required to make a decision without knowing all the

information, or when all the available options seem bad to her, her action would be at best only partly autonomous, because the decision she makes is not one that she can consciously endorse given her particular practical identity. Such situations can arise in a number of ways, e.g., through coercion (a person being forced to do something with a gun pointed at her head) or through an unrealistic assessment of how the world is (a person making "unreasonable" demands of a doctor). It is generally thought that when a person is coerced or disempowered in some way, efforts should be made to change the external factors; by contrast when a person is unreasonable or ignorant, it is generally thought that the internal factors (i.e., the person) should change. The broad point is that in order to become more autonomous, the person needs to either find or be shown another option that she can endorse or modify her practical identity. The goal is not to dismiss the person's status as a valuer, but is rather to enable her to make a choice that is in some way consistent with her agency.

When applying Korsgaard's account of agency and identity to dementia, one might think that the dementia patient fails to act as himself because she is either not conscious of her practical identity (i.e., she cannot articulate who she is or what makes her life worth living) or not conscious of the reasons why she is doing what she is doing (e.g., Margo's claim that she is reading mystery novels). This is something like the view taken by Dworkin. Jaworska's view on the other hand is that this sets too high a standard for autonomous action. In her account, people are often only partly conscious of why they do what they do, and this fact does not in itself render them non-autonomous. However, this point does not negate Korsgaard's analysis of agency and identity. Though people are not always consciously aware of the reasons underpinning their actions, such reasons are typically still present – engrained as it were in the structure of the self (Gillett and McMillan 2000, 222–235). Though practical identities are developed through conscious deliberation, once established they operate preconsciously in framing our perception of the world and ourselves within it and setting our patterns of thought and behavior. Thus, the question of whether or not an action was autonomous can be considered retrospectively. An "instinctive" action may be autonomous, particularly if it is the result of a habit that one has trained oneself into, for reasons that one has reflectively endorsed (c.f. Aristotle 1985, 1095a). The same point may be applied to people with dementia: though their present reasoning as expressed may be inadequate or spurious, if the action coheres with their engrained practical identity, it may still be thought of as their own action.

Jaworska notes the importance of coherence in relation to a number of the cases she describes. Dr B, for example, is an Alzheimer's sufferer who chooses to participate in a study because he sees the study as a significant contribution to science. In this case it is clear how this decision coheres with his life as a medical professional. Moreover, the participation has a significant effect on his self-esteem. Comparing the study to the "filler" group activities at the care facility he is staying at, he says to the researcher: "If I'm working with you, I can-look, I can work in here for 30 times and all that, but in this group, *I'm nothing*" (Sabat 1998, 41, cited in Jaworska 1999, 118). This stark assessment of a life spent in what is regarded as

meaningless activity fits squarely with Korsgaard's link between agency and identity: being unable to endorse the activities that are made available to him, Dr B is unable to act as himself, and this literally negates who he is. In Jaworska's view this kind of coherence may be taken as confirming that the values expressed are in fact "authentic" values and not caused merely by "brain pathology" (Jaworska 1999, footnote 41). However, she also claims that Alzheimer's patients are unlikely to form new values and that for this reason their values are typically coherent in the manner described.

This link between identity and coherence provides a way of differentiating the cases of Mr. O'Connor, and Mr. D. Korsgaard's account of identity provides grounds for concluding that Mr. D is in fact not acting autonomously. He has – perhaps through some disorder of perception or volition caused by the neurological damage – become disconnected from his practical identity and is no longer capable of acting on reasons that are his own. Mr. O'Connor's values, on the other hand, are traceable to his pre-dementia life, and so one may view him as acting autonomously (albeit in a minimal sense). However, it must be noted that this conclusion is based on possibility and presumption. It is also possible that Mr. D's "practical identity" was not as it seemed or at least that he was more conflicted in his values than he revealed. Perhaps while maintaining a respectful family life he carried unspoken resentment or disappointment which is now being expressed because his ability to understand and inhibit his feelings is impaired (McMillan 2005, 68–70). This too is only a conjecture – the broader point is that the grounds for declaring the person autonomous or non-autonomous are not definitive. This uncertainty can be a cause of great pain to those caring for a family member with dementia. The sense that there may be some "truth" in the contrary behavior now exhibited may cause the carers to doubt the authenticity of the life that has been lived, and which they have shared in.

The possibility of an identity being conflicted raises further problems, applicable to cases besides that of Mr. D. Korsgaard's account of identity starts with the observation that one needs to unify one's actions through practical reasoning. This need arises in response to some kind of discordance in our minds. While the primary structures of the mind enable an infant to move, to eat, and to begin relating to others, as consciousness grows the child becomes subject to a multitude of feelings which may move her in contrary directions. Being conscious of this discordance, the child becomes aware of the need to "pull herself together" through identifying with reasons (Korsgaard 2009, 126). This work of self-unification is, as Korsgaard puts it, the "task of life": the task of becoming a person – *oneself* (Korsgaard 2009, 130). It is only ever achieved by degrees, and most people experience thoughts and feelings that they would not make the basis of their actions, but rather try to expunge from their minds. Given that this is the case, it is reasonable to ask why Mr. O'Connor's feelings of grief over the loss of his wife should be regarded as more coherent with his prior values than his religious convictions, which he no longer carries in his demented condition. It is possible to argue that his dementia is preventing him from enacting a central part of his identity – i.e., the person he endeavored to be – and that it is therefore appropriate to apply his advance directive to continue receiving life-extending treatments.

Persistent Uncertainties Regarding Identity and Autonomy

It may be noted that the argument just outlined leads to a conclusion very similar to that arrived at via Dworkin's notion of critical interests. The difference is that Korsgaard's account of identity provides a more developed theoretical account of what critical interests are, how they are formed, how they might conflict, how they may be distinguished from other kinds of interests, and how they might be variously affected by the loss of particular cognitive faculties. This explains more clearly why one cannot – at least on the basis of the information that has been given – say definitively whether a patient like Mr. O'Connor is making an autonomous choice or not. To make this assessment one would need a better understanding of who he was, the life that he has lived, and in particular how he related to the religious beliefs mentioned. Practically speaking, one might think that this theory makes the task of carers more difficult, and in a sense this is true. However, while it provides no easy solutions, it does at least validate and partly explain the tensions and uncertainties that people in such situations are facing. Given such uncertainties, in some cases it may be more responsible to admit that one simply does not know whether an advance directive should be applied and revert instead to some kind of substitute decision-making, perhaps based on an assessment of the patients current "best interests." This may be the most ethically honest approach.

To go beyond this position of uncertainty, it would be necessary to develop a more detailed account of how a person's "true" identity can be discerned, and the conditions under which such a judgment can be made. For example, if it could be shown that close personal relationships and the feelings associated with them (e.g., the grief experience by Mr. O'Connor) were in fact more indicative of the authentic self than cognitive beliefs (e.g., the religious beliefs attributed to Mr. O'Connor), then this would support Jaworska's view that Mr. O'Connor's current choices are sufficiently autonomous. There is also scope for further ethical analysis of situations where a judgment about authenticity seems impossible because of the fragmentary nature of the person's life. Along with this, it may be useful to have a more extensive neurophysiological account of the brain functions required to form and retain values, as this could support judgments about whether a change in attitudes or behavior is directly attributable to disease or trauma, rather than, say, a loss of inhibition. All these questions will only become more pressing as the prevalence of dementia increases.

Definition of Key Terms

Identity	The properties, qualities, or characteristics that make a person who he or she is.
Dementia	Refers to a spectrum of neuropsychological disorders characterized by a significant loss of cognitive functions including memory, planning, attention, and judgment. Alzheimer's disease is a well-known form of dementia.

Critical interests	An action, capacity, or state of affairs that a person is concerned about because it in some way significantly affects that person's identity.
Experiential interests	An action, capacity, or state of affairs that a person is concerned about because it generates positive experiences, but which does not significantly affect that person's identity.
Agency	The capacity to act so as to give effect to something. When applied to person's, agency usually refers to the person's ability to act in ways that express that person's identity.

Summary Points

- In many cases dementia causes profound changes in attitudes and behavior, and these changes raise questions about the nature of identity and how it is affected by reduced neuropsychological function.
- These questions about identity become pressing when considering whether or not to apply an advance directive for a patient who exhibits apparent changes in identity following the onset of dementia.
- Ronald Dworkin has argued that identity should be understood in terms of "critical interests," which he takes to be acquired through a rational awareness of oneself and one's life considered as a whole. On this view most people with dementia are incapable of forming new critical interests, and hence their pre-dementia wishes should be regarded as expressive of their identity.
- In response to Dworkin, Agnieszka Jaworska has argued that critical interests are primarily grounded in the capacity to value and that many people with dementia retain this capacity. On this view, one may regard a person with dementia as capable of expressing genuine choices that may override prior choices (advance directives).
- Christine Korsgaard's account of identity fits with and extends upon both Dworkin and Jaworska's ideas and indicates that the critical interests are yet more complex and nuanced.
- This more developed understanding of critical interests shows that there is scope for further philosophical work on the nature of identity and its application to these areas of medicine and that this work could be usefully supplemented by a better understanding of the neurophysiology of valuing.

References

Aristotle (1985) Nicomachean ethics (trans: Irwin T). Hackett Publishing Company, Indianapolis/Cambridge

Buchanan A (1988) Advance directives and the personal identity problem. Philos Public Aff 17 (4):277–302

Dresser R (1995) Dworkin on dementia: elegant theory, questionable policy. Hast Cent Rep 25 (6):32–38
Dworkin R (1993) Life's Dominion: an essay about abortion euthanasia and individual freedom. Alfred A Knopf, New York
Gillett G, McMillan J (2000) Consciousness and intentionality. John Benjamins, Amsterdam/ Philadelphia
Hope T (1994) Personal identity and psychiatric illness. In: Griffiths P (ed) Philosophy, psychology and psychiatry. Cambridge University Press, Cambridge, pp 131–143
Hope T, McMillan J (2011) Advance decisions, chronic mental illness, and everyday care. Lancet 377(9783):2076–2077
Jaworska A (1999) Respecting the margins of agency: Alzheimer's patients and the capacity to value. Philos Public Aff 28(2):105–138
Korsgaard C (1996a) Creating the kingdom of ends. Cambridge University Press, Cambridge/New York
Korsgaard C (1996b) The sources of normativity. Cambridge University Press, Cambridge
Korsgaard C (2009) Self-constitution: agency, identity and inte. Oxford University Press, Oxford
McMillan J (2005) Identity, self and dementia. In: Hughes J, Louw S, Sabat S (eds) Dementia; mind, meaning and the person. Oxford University Press, Oxford, pp 63–70
Nuffield Council on Bioethics (2009) Dementia: ethical issues: 45. http://nuffieldbioethics.org/project/dementia/. Accessed 15 Dec 2014
Sabat S (1998) Voices of Alzheimer's disease sufferers: a call for treatment based on personhood. J Clin Ethics 9(1):35–48

Children Are Not Small Adults: Significance of Biological and Cognitive Development in Medical Practice

25

Vic Larcher

Contents

Abstract

Conceptions of childhood have undergone continuous and historical evolution; children can no longer be regarded as small adults. Most contemporary views on the nature of childhood are derived from Aristotelian concepts; they stress its developmental nature and the role of adults in guiding and facilitating children's development. Transformation to adulthood occurs by a process of biological, cognitive, and moral development in which distinct stages can be identified. Children's portrayal in art, literature, and the media has largely mirrored evolving concepts of childhood and increasingly takes account of children's voices.

V. Larcher (✉)
Honorary Consultant in Paediatrics and Ethics, Great Ormond Street Hospital, London, UK
e-mail: vic.larcher@yahoo.co.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_16

Biologically, children's anatomy, physiology, and pathology go through significant changes. Infants and very young children show an increased vulnerability to environmental and other harms, because of their size, immature anatomy and physiology, and differing pharmacodynamics. Organ development determines the patterns of disease that occur in childhood and affects drug treatments and responses to them.

Cognitive development (the child's evolving ability to think and understand) is classically understood as following a series of stages in which genetically determined characteristics interact with events and experiences to influence patterns of cognition. A similar staged pattern is seen in moral development in which the child's capacity to make decisions about what is right or wrong gradually evolves. Contemporary neuroimaging techniques have provided structural and functional evidence to underpin this process. A child's stage of cognitive and moral development is used to determine what legal status they may have and what weight should be placed on their desire to determine their own best interests.

Concepts of children as rights holders allow them to have moral agency, irrespective of their cognitive ability or capacity to exercise choices. A staged acquisition of liberty rights (i.e., right to self-determination) seems plausible, as does the universal possession of welfare and protection rights.

The interests of children who lack the capacity to exercise liberty rights can be protected by the application of the best interests or welfare principle. Determination of best interests includes consideration of the harms and benefits involved; the likely outcome of various options; the wishes, preferences, beliefs, and values of the child insofar as she/he is able to express them; the wishes, beliefs, and values of the family and other relevant individuals; and which option is least restrictive of future choices for the child.

Children are not small adults who can be treated as though they were, and neither are they uniformly vulnerable beings who need protection; rather they are individuals in transition whose growth into adulthood should be supported, encouraged, and facilitated.

Introduction

All readers of this chapter will have shared at least one common experience; all have made the transition from childhood to adulthood. During that time they will have accumulated knowledge, values, and beliefs that have shaped who and what they are. The means by which they have done so will have varied, in part at least conditioned and influenced by parents, teachers, and the psychosocial and cultural milieu in which they were raised. The way in which they were treated as children will undoubtedly have had some influence on who they are today, which in turn will influence how they treat children and their perception of childhood.

It is often stated that "children are not small adults." (World Health Organization) Unpicking this simple assertion raises questions as to the status of children in

society, the duties that owed to them, and how they should be treated. There is a need to understand what children are, what properties distinguish them from adults, and how those properties relate to their status in society. These physical, mental, and moral properties alter at different rates for different individuals as part of a continuous process of development, though the latter is often categorized in stages. For a variety of reasons, not all children will attain the neurodevelopmental status associated with adulthood.

If children are *not* small adults, it is necessary to understand why they were once so regarded, what has altered to change these perceptions, and how this relates to the contemporary biopsychosocial concept of childhood. How childhood is conceived is crucial for addressing philosophically interesting questions about children, including their symbolic, sociological legal and moral status in society.

This chapter is concerned with these questions, and in keeping of the theme of this book, the focus will be on the narrow area of children's health rather than the wider social context.

The Nature of Children and Childhood

Preliminary Considerations

In biological terms, a child is a young human being below the age of maturity. Policy and legal documents commonly define childhood and its subdivisions in terms of age. The World Health Organization defines newborns as being aged 0–28 days, infants from 28 days to 12 months, children from 1 to 10 years, and adolescents from 10 to 19 years, but all fall within the United Nation's definition of children. The notion of a staged or gradual transition from the vulnerability and dependency of infancy to the maturity and autonomy of adulthood is key to understanding the concept of childhood.

Concepts of childhood have shown historical evolution (Ariès 1962) and are socially and culturally conditioned. Advances in biology, psychology, and neuroimaging have provided valuable insights into how children grow, develop, and acquire knowledge. In pre-Aristotelian times, and as late as the Middle Ages, children appear to have been regarded as small adults, sometimes on the basis of their depiction in visual arts. But depictions of children, especially in religious paintings, may have more to do with depiction of the infant Christ's deity rather than a depiction of childhood per se. Some medieval thinkers seem to have understood childhood as a staged process (Shahar 1990), even in the context of a preindustrial society with high childhood mortality and lower overall life expectancy.

Aristotle (Aristotle's *Nicomachean Ethics* 1934, *The Politics* 1932) challenged the notion that children were small adults, regarding them as part of a community whose members shared common ends and worked together to achieve the higher good. He argued that children, unlike adults, are not capable of true happiness because they have not developed the ability to use their intelligence to guide their

actions. Equally their inability to realize the implications of their actions may lead them to impulsive and uncontrolled behaviors that have harmful consequences. Aristotle did not believe children should be left to make moral decisions until their intellect had developed sufficiently to enable them to decide what to do in order to achieve moral and social ends and to achieve the best conduct of which they were capable. Education directed at this purpose was to be delivered by intelligent teachers of high moral character, who should have a sense of what children can and should do. All activities of childhood, including games and imitative play, should be directed at preparing children to become responsible adults with high moral principles.

The Aristotelian concept of childhood regards children as immature human organisms having the natural potential to mature into their normal or standard adult forms. This concept determines the responsibilities, duties, and relationships that adults have toward children, namely, an obligation to provide the supportive environment that children need to evolve into responsible, functional members of society. This process of development guided by adults toward societally determined goals underpins many contemporary perceptions of children and childhood.

The actual process whereby this transformation happens has been the subject of further discussion and debate. The "recapitulation" theory, that the development of the individual recapitulates the evolution of the race or species (Spock 1968, 229), has largely been rejected in favor of stage theory, which explains apparent recapitulation by reference to general principles of age-related structural changes in cognitive development (see, e.g., Piaget 1968, 27). Descriptions of staged development can be found from medieval times onwards (Shahar 1990, 21–23; Rousseau's *Emile*). In the twentieth century, Piaget formulated a highly sophisticated version of stage theory that remains the dominant paradigm for conceiving childhood (see, e.g., Piaget 1971). Versions of stage theory are consistent with clinical observations and scientific studies of physical and neurological development of children in health and disease

Humanitarian Considerations: Children in Visual Arts Literature and the Media

The depiction of children in art, literature, and visual media over time has enabled a further move away from the concept of children as small adults. Early artistic representations of children gave them the proportions of small adults; even when these were corrected, children were often dressed in adult style clothes and adopted adult style poses. The development of later forms of visual imagery did allow more realistic representation of the child and his/her world, albeit viewed with an adult's eyes.

Early literature was more involved in providing educational material. As the concept of childhood as an age of innocence and immaturity developed, so did the moralizing nature of children's literature, as fulfillment of the need to provide moral education through literature. In the later twentieth century, children have been depicted as moral protagonists in the struggle between good and evil.

Even later in the twentieth century, the growth of other forms of visual media and its relationship with every element of children's lives has become important in defining society's concept of children and childhood. For example, exposure of children to violence on the media has adverse behavioral effect, yet protection of children from these influences has become more difficult. Media portrayal of children also affects how they are perceived, provides role models for children, and may reinforce perceptions of the children as ignorant, self-centered individuals, as victims, as cute kids, as little angels, or as commodities or accessories. Coverage of issues that affect children may be poor, and there may be limited opportunities for children to express their views, and when this is possible, the approach is adult centered. New media has been a vehicle in which both good and bad, e.g., pornographic images of children, can be stored and distributed.

However current media portrayals of children may underplay the similarities with adults that do exist while overemphasizing children's vulnerabilities and powerlessness. To redress this imbalance, representation of children should assist in the understanding of what it is to be a child and to view the world from their unique biopsychosocial perspective.

Biological Considerations

Human babies are vulnerable; they cannot feed independently, move out of danger, or maintain their temperatures. Children's anatomy and physiology differs from adults, but over time develops toward that of adults. The pattern of children's diseases also differs: some conditions are unique to childhood; others are genetically determined; others have their major impact in childhood; and others differ from those of adults in their expression, severity, impact, and outcome.

Environmental Factors

Children are at greater risk from environmental hazards than adults because they have different and unique exposures, and their changing physiology means they have different responses to risks that are exacerbated by longer life expectancy; they have critical windows of vulnerability that have no parallels in adult physiology. These vulnerabilities are enhanced because young children lack the physical, cognitive, or linguistic abilities to avoid dangers or articulate concerns.

Babies can be harmed before birth in ways that may affect their future lives. Assisted reproductive technology (ART) is increasingly used to conceive children, yet the outcomes for future development may be unclear and difficult to ascertain. Fetuses can be harmed by agents that cross the placenta; the latter may be chemical (drugs such as thalidomide and stilbestrol, alcohol, tobacco; pollutants such as polychlorinated biphenols), physical (radiation, heat), or biological (viruses such as rubella, cytomegalovirus; parasites such as toxoplasmosis). After birth babies can be uniquely harmed by toxic materials transmitted in breast milk.

Small children are also at increased risk; their exploratory behaviors lead them to consume pollutants; they are nearer ground level where pollutants are concentrated; they have higher surface area to body mass than adults, placing them at greater risk of skin exposure; they may be unable to recognize or escape from dangers.

Preadolescent children may take unreasonable risks because of cognitive immaturity and risk-taking behaviors that exist even when there is cognitive awareness of dangers.

Physiological Aspects

Children have a dynamic anabolic physiology which permits growth and results in increased energy, water, and oxygen consumption. In consequence food toxicants are delivered at two to three times the adult rate and fluid-delivered toxicants at five to seven times the adult rate. Air pollutants are delivered at greater rates because the infant has three times the breathing capacity (by weight) of adults and the child two times that of adults. Distribution of toxicants is also affected by differences in body composition, availability of binding proteins, and an immature blood-brain barrier. Detoxification is compromised by immaturity of enzyme systems or their abnormal functioning. Elimination is also compromised since kidney function in neonates is <40 % of adult values.

All of these physiological processes may differ unpredictably between children and adults. Children's physiological systems continue to develop through adolescence and undergo significant change at time of puberty. Growth of vital organs occurs at different rates, but if disrupted during critical periods, damage may be severe and lifelong. Environmental hazards may operate to harm a developmentally dynamic child by mechanisms that do not operate in adults.

Respiratory Development

Respiratory system development continues throughout children's linear growth; the number of alveoli (air sacs) increase from 10 million at birth to 300 million at age of 8. Exposure to toxic agents (tobacco smoke, particulates) has adverse effects on lung structure and function, both in childhood and beyond. Exposure to dirty air is related to development of asthma and emphysema and may adversely affect overall growth and development.

Neurodevelopment

Brain development extends continuously into adult life, though the major changes in growth occur early in childhood. Brain connections (synapses) form and reorganize, while neurotransmitters are redistributed. Vulnerable periods of nervous system development are sensitive to environmental insults that can be associated with later acceleration of age-related decline in function. Brain development can also be partially determined by the interaction individuals have with their environment; identifying and "closing" windows of exposure may be important in the prevention of adult onset disease.

Throughout adolescence similar brain areas undergo significant restructuring in both sexes, although the rate is dependent on sex. Development of the prefrontal and temporal parietal cortical areas is particularly protracted. Since these regions are involved in a number of cognitive functions, including decision-making and social cognition (the understanding of other people), their protracted development might contribute to teen-associated behaviors, e.g., increased risk taking and

reduced self-control. In medical practice adverse adolescent behaviors include poor compliance with treatment regimes and even outright refusal of treatments.

The Nature of Childhood Diseases

Diseases that occur in children may have similar clinicopathological features to those occurring in adults but may be modified as a result of the child's developmental status.

Extremely premature babies have specific age-determined defects in lung maturation that can be treated by surfactant. Their immature intestines can develop conditions, e.g., necrotizing enterocolitis that have no adult counterpart; their eyes can be affected by retinopathy of prematurity. Extreme prematurity is also associated with a range of neurological and cognitive impairments that may profoundly affect future life.

Genetically determined conditions may present in early childhood (e.g., metabolic storage disorders) and affect a number of organ systems. They can have a profound effect on the child's future life, but also pose problems for families, because of their inherited nature, the possibility of genetic screening, prenatal testing, and embryo selection in future pregnancies. The social and economic costs of these and other childhood illnesses may be considerable and raise philosophical questions about the nature and treatment of disability.

Other conditions show the phenomenon of programming, where permanent effects result from a stimulus applied at a "critical window" of development in fetal or neonatal life. Undetected and untreated congenital hypothyroidism produces lifelong typical phenotypic changes and permanent learning difficulties, whereas adults who develop hypothyroidism do not suffer the same sequelae because their central nervous systems have ceased developing.

A number of congenital malformations affecting the major organ systems (e.g., heart, kidney, brain) of children arise in embryological development and have no adult parallel. Increasingly children survive, with treatment, into adult life; their condition and its sequelae play a role in determining their identity and authenticity. Similar constraints may apply to those with chromosomal abnormalities (e.g., trisomies 18 and 21, translocations, etc.) that may have long-term somatic and neurodevelopmental sequelae.

Increasing numbers of children with diseases that were previously fatal in childhood now survive into adult life (e.g., cystic fibrosis; sickle cell disease) often with increasing comorbidity. Other illnesses (e.g., childhood cancers and leukemia), with a better response than their adult counterparts, may show significant late effects such as growth suppression or infertility. Attempts to prevent the adverse potential impact of these late effects may involve children considering options for their future adult selves (e.g., the desire to have children and hence to consent to harvest and store gametes) at a time when they may lack the capacity or experience to do so. In effect they are being treated as small adults.

Finally, some adult onset conditions have their origin in fetal life or early childhood (the Barker hypothesis). Adult coronary heart disease, stroke, diabetes, and hypertension have been associated with lower birth weights and placental

weights. As a corollary there are concerns that changes in lifestyle in children have, by promoting obesity, produced a rising incidence of adult-type disease, e.g., type 2 diabetes and hypertension.

Pharmacology

Most cellular processes are similar in children and adults so that extrapolation of data on children's medicines from adult studies, or from small-scale studies, is possible. Some disease processes, e.g., celiac disease or Crohn's disease, share the same pathology so that treatment regimes can be extrapolated from adults.

However, there are significant differences which mean that it would be unwise and potentially or actually unsafe to treat children as small adults (see Stephenson 2005). Children may have different rates of handling medicines, because of differing gastrointestinal function, limited protein binding of drugs, differing body composition (lipid-water ratio), elevated blood-brain barrier to drug penetration, and differing rates of liver metabolism and renal clearance. The toxic or therapeutic effects of some drugs may be unpredictably enhanced or diminished.

Identical drugs may be used to treat different diseases in adults and children, e.g., proton pump inhibitors to treat peptic ulcer in adults and gastroesophageal reflux in children. Similar drugs (e.g., theophylline and caffeine) may behave in different age-dependent ways. Some disease processes show differences between adults and children in their response to therapy, e.g., asthma and depression, while in other cases diseases that carry the same name, e.g., migraine and epilepsy, show variability in their clinical features. Drugs can have age-related adverse events, e.g., valproate-induced liver disease in young children with learning difficulties and on multiple anticonvulsants and chloramphenicol-induced cardiotoxicity. In other illnesses individual host responses to medication vary, e.g., pneumonia and leukemia.

The pharmacodynamics of drugs can alter with development resulting in enhancement of their desired action or adverse effects, e.g., warfarin and cyclosporin. Drugs given in fetal life or in childhood can also temporarily or permanently program future outcomes.

Drugs may have different actions on individuals within an adult population because of genetic polymorphism. Mutations for enzymes or receptors explain some of the variation in response to medicines seen in adults. Less attention has been paid to the phenomenon in children, despite the fact that some genes (e.g., that for fetal hemoglobin) are expressed much more in early life than in adults. Such gene switching could account for age-dependent changes in drug efficacy.

Cognitive Development

Cognitive development is the emergence of the child's ability to think and understand. Studies have focused on aspects of children's development, e.g., information processing and language learning, compared to those of adults.

A major controversy has been whether cognitive development is mainly determined by an individual's innate qualities ("nature"), or by their personal experiences ("nurture"). Plato regarded learning as a recollection of previously known (innate) material; Descartes postulated that a clear and distinct knowledge of the world can be constructed from resources innate to the human mind (Descartes [PW] 1985, 131). In contrast John Locke regarded the human mind as beginning as a "tabula rasa," a "white paper, void of all characters, without any ideas" (Locke [EHC] 1959, 121). All the "materials of reason and knowledge" can be postulated as coming from experience.

Behaviorism, concerned with observable behaviors of individuals rather than unobservable mental processes, provides some support for an empiricist or "nurture" view. However, Chomsky (1959) has argued that no purely behaviorist account of language learning is possible and has provided a viable alternative to a purely empiricist conception of language development.

Few contemporary theorists of children's cognitive development completely accept either extreme empiricism or strong innatism and regard the "nature vs nurture" dichotomy as a false one. There is evidence (from biological and behavioral sciences, e.g., twin studies) that gene activity interacts with events and experiences in the environment to influence cognitive development (see Carlson et al. 2005). Although a number of more contemporary theories of cognitive development have been generated, Piaget's work (Piaget 1968, 1971) has remained strongly influential on current educational and psychological thinking about the cognitive development of children. By recording children's intellectual growth, Piaget formulated theories to explain children's cognitive development. He discussed the nature of thinking, the "location" of dreams, what it is to be alive, the philosophy of language and concepts of space, time, and causality, setting out to identify the discrete stages in which children come to understand what these topics are. However underpinning this is the assumption that there are satisfactory responses to the philosophical questions that such topics raise. Piaget's theory of cognitive development postulated four stages of cognitive development.

The **sensorimotor stage** extends from birth to the acquisition of language (ca. 2 years). Infants progressively acquire knowledge and understanding of the world by coordinating experiences with physical interactions with objects. In this stage the reflex actions of newborns give way to the beginnings of symbolic thought. The child comes to learn that he/she is separate from the environment, in which objects continue to exist even if they cannot be seen or heard (object permanence), and hence develops a distinct notion of self.

The **pre-operational stage** lasts from 2 until the age of 7 years. During this stage children's play and pretending increases although they cannot understand logic and have difficulty in seeing the viewpoints of others because their thinking is still egocentric. The pre-operational stage is split into substages. In the symbolic function substage, children can understand, represent, remember, and picture objects in their mind without having the object in front of them. In the intuitive thought substage that follows, children ask "why?" and "how so?" type of questions.

The **concrete operational stage** occurs between the ages of 7 and 11 years (preadolescence). Children's thought processes become more mature, "adult like" and logical, although they have not developed abstract hypothetical (what if?) thinking. Children can reason by induction but struggle with deductive reasoning. During this stage children develop the ability to distinguish between their own thoughts and those of others, classify objects by certain physical characteristics, think logically about objects and events, and successfully complete addition and subtraction problems.

The **formal operational stage** spans from adolescence to adulthood (roughly ages 11 to approximately 15–20 years). Children become capable of hypothetical and deductive reasoning and develop the ability to think about abstract concepts; they are able to consider possible outcomes and consequences of their actions. Adolescents and adults also develop the abilities to reason about their thought processes and monitor them and to systematically solve problems in a logical and methodical way. Not all persons in all cultures reach the formal operational stage, and people do not use formal operations in *all* aspects of their lives.

Some of the theoretical claims made by Piaget, and their implications for how children are treated, have been questioned, but the general identification of changes in cognition with age and the need to explain them are still valid. According to Matthews (2008, 2009), stage theories of development tend to support a "deficit conception" of childhood, in which children are perceived as individuals lacking the capabilities of normal adults. This ignores or undervalues the fact that children are able to outperform their future adult selves in some domains and affects the relationships adults think they can have with their children.

In recent years, other models, for example, information-processing theory and alternative theories of cognitive development, have been proposed to integrate Piaget's ideas with more recent concepts in developmental and cognitive science and theoretical cognitive neuroscience.

Moral Development in Children

The ability to make decisions about what is right or wrong, to attempt to answer philosophical questions, and to debate philosophical topics is closely related to the development of the ability to reason and of self-awareness. This process can be linked to cognitive development if an account of moral behavior linked to rationality can be accepted. Society regards the moral development of children as important in determining their moral status, how they may be treated and what weight is attached to their views.

Interest in the moral development of children parallels the recognition of childhood as an important and expanding stage of human development. Traditional philosophy has been mainly concerned with those who have achieved moral maturity, based on rationality. Nevertheless philosophers have devoted their attention to issues of moral development of children. Rousseau (1979) described a sequence of five age-related stages through which an individual must pass to

reach moral maturity: (i) infancy (birth to age 2), (ii) the age of sensation (3–12), (iii) the age of ideas (13 to puberty), (iv) the age of sentiment (puberty to age 20), and (v) the age of marriage and social responsibility (age 21 and above). In early life a child's behavior may be modified by parental reward of good behavior and punishment of bad behavior. Rousseau believed that, in general, children could not respond to moral reasoning before they had reached the third stage (age 13+); he believed that attempts to reason with a child younger than 13 years of age were developmentally inappropriate, a proposition which later educationalists have disputed.

It seems logical that a child's acquisition of moral reasoning, as opposed to moral intuition, does broadly follow a similar pattern to that of cognitive development. Kohlberg's psychological research with young people identified six discrete sequential stages of development, each enabling the child to give a more adequate response to moral dilemmas. He postulated that moral development continued throughout life and was strongly associated with respect for the principle of justice and with the importance of social cohesion. Kohlberg (1981–1984) grouped his six stages into three levels, with the final level going beyond what normal adults could or did achieve. Identification of an individual's stage of moral development depends on the form of reasoning used in resolving a moral dilemma. Higher-ranking individuals were more likely to exhibit consistent and predictable moral behaviors than those with lower scores. Each stage was regarded as cumulative, sequential, necessary, and irreversible. The rate of progression through the stages varies between individuals and cultures.

Level 1. Premoral

The morality of an action for the individual is determined by its immediate consequences for that individual alone; it is egocentric and not related to social convention.

Stage 1: Punishment and obedience orientation

At this stage the level of punishment or praise is directly linked to how bad or good the action was.

Stage 2: Naive instrumental hedonism

The rightness or wrongness of an action is determined by whatever the individual believes to be in his/her own best interests, rather than any direct consideration of the interests of the wider population.

Level 2: Morality of conventional role conformity

This level, typical of adolescents or adults, is characterized by obeying social conventions of right or wrong. An individual follows socially determined rules even when there are no personal consequences for obedience or disobedience and rarely questions their fairness or appropriateness.

Stage 3: Morality of maintaining good relations and approval by others

To be good is to conform to society's expectations. The morality of an action may be determined by its consequences on social relationships rather than purely

on itself but may include consideration of values such as respect, gratitude, and reciprocity.

Stage 4: Authority-maintaining morality

The key drivers of this stage are obedience to authority and maintenance of the social order. Obeying laws and observing conventions is important because it maintains social cohesion, function, and stability which are seen as desirable ends. Violation of a law is morally wrong and deserves punishment. Most members of the society and cultures remain at stage 4 and as such their morality is determined by the outside influences.

Level 3: Morality of accepted moral principles (postconventional)

At this level an individual's views, not formed out of self-interest but from the rational application of the individual's ethical principles, may run counter to those of society. Rules that are accepted as maintaining a general social order may not be held to be binding if they conflict with justifiable principles, e.g., rights, liberty, and justice.

Stage 5: Morality of contract, of individual rights, and of democratically accepted law

Although the individual regards the outside world as having certain perspectives and values that should generally be respected, rules can be changed democratically if a case can be made to do so, e.g., if they do not promote overall welfare. Some (inevitable) moral trade-offs are recognized and accepted.

Stage 6: Morality of individual principles

Moral reasoning becomes based on abstract reasoning and the discernment or application of universal moral principles, of which Kohlberg regarded justice as being the most important. Decisions are not reached conditionally but in an absolute fashion that has resonance with Kantian philosophy in that an action is intrinsically right regardless of its consequences. Actions are perceived as ends in themselves, and the principles underpinning them are universal in application.

Kohlberg can be seen as regarding the final stage of moral development as achievement of an austere Kantian philosophy with an emphasis on rationality and respect for the principle of justice, but acknowledged that this stage was reached by relatively few and was difficult to apply consistently. Some individuals and cultures progress through the stages at different rates and to different levels for reasons that remain to be defined.

Kohlberg's stages of moral development presuppose that humans possess the potential to develop certain characteristics which have intrinsic worth, e.g., an ability to communicate, the capacity to reason, and a desire to understand one another. His theory has been criticized because of its emphasis on justice and the claimed exclusion of other values that people regard as important and because it has been regarded as androcentric (Gilligan 1982). Controversially, Kohlberg placed women at a lower stage of moral development than men, because he claimed that women focus on social and family relationships as a means for solving moral

disputes rather than the higher level application of abstract reasoning. However it seems counterintuitive to propose that abstract moral reasoning is the only characteristic that should be valued and encouraged.

Paradoxically adults often demonstrate significant inconsistency in their moral judgments and may appear to function at lower levels than they seem to have achieved. Kohlberg's theory may actually take insufficient account of the complex influences that determine how most individuals make moral decisions in their everyday lives. Individuals often appear to make moral judgments without weighing concerns such as fairness, law, human rights, or abstract ethical values.

Moral reasoning may involve developing other perspectives, for example, the staged acquisition of empathetic feelings and responses that can be observed in young children. If this is so, it would seem to allow the ascription of genuine moral feelings, and so of genuine moral agency, to very small children. This seems at variance with Kohlberg's categorization of such individuals as premoral.

Contemporary neuroimaging techniques have provided structural and maturational evidence that underpins the concept of moral development but without delineation of the precise neurological processes by which it occurs. Explicit making of right and wrong judgments seems to link to the activation of parts of the brain that can be identified on functional MRI studies in adults. The dual-process theory of moral reasoning, arising from the neuroimaging work of Greene et al., claims that certain kinds of moral dilemmas activate brain regions specific to emotional responses, while others activate areas specific to cognition (Greene et al. 2004). This appears to indicate dissociation between different types of moral reasoning. Personal moral dilemmas (where the subject had to imagine intentionally acting to cause serious harm directly to another person) activated regions of the brain that had been associated with emotional processing in previous studies. Impersonal dilemmas, by contrast, created more activation in areas previously associated with working memory, a presumably more cognitive process. In other studies counterintuitive moral judgments were associated with greater difficulty and with activation in the rostral anterior cingulate cortex, suggesting that such judgments may involve emotional conflict; intuitive judgments were linked to activation in the visual and premotor cortex. However others have claimed that the identified cortical areas are not functionally specific, since they may be active in a wide variety of cognitive and emotional tasks. There may be a continuity between moral decision-making and other kinds of complex social deliberation, but further work will be necessary to determine the basis for this and its relevance for moral development and clinical practice in children.

The formal stage of moral reasoning that a child may have reached does not of itself determine how that individual should be treated or what respect should be given to their wishes, beliefs, preferences, and values. It would seem counterintuitive to require that a child reaches a particular stage of moral development in order to make decisions, or for their views to be accorded some weight, when adults are not required to reach the same stage or demonstrate consistently that they have done so. However in medicine great emphasis is placed on a child's level of cognitive

ability and moral development in determining the weight that might be given to their expressed wishes over the care and treatment they will receive. Inter alia this involves striking a balance between protecting children from unwise choices or harmful situations and granting them the right for self-determination which would follow if they were truly mini-adults. Such considerations determine the level of moral agency that might be ascribed to a child in which the language of rights is increasingly used.

Children's Rights

Children have moral status that stems from their very nature as young human beings. However, there are certain things that they may not do that adults are allowed to do, e.g., vote, marry, buy alcohol, and serve in the armed forces. One way of balancing possibly conflicting perspectives of what children are entitled to do, to be or to have, is to grant them rights that carry correlative duties on others. It is common to distinguish legal rights from moral rights and most jurisdictions accord children legal rights. The possession of moral rights, and if – and how – these should differ from those held or claimed by adults, is more conjectural and has been subject to different kinds of philosophical criticisms (see Archard 2004).

Historical Perspectives

Although the notion of children as rights holders dates from the eighteenth century, the first international declaration of children's rights of the child was adopted by the League of Nations in 1924 (Geneva Declaration of the Rights of the Child 1924; in Buck 2014). It provided physical and spiritual development rights, welfare and protection rights, and the aspiration that the child "must be brought up in the consciousness that its talents must be devoted to the service of its fellow men." In accordance with the views of childhood at that time, there was no emphasis on participation rights. The document was accepted by the United Nations in 1959; it was updated to the UN Convention of the Rights of the Child (UNCRC) in 1989 and has been ratified by most countries. The Convention grants children a wide range of rights (in 42 articles) including the right to have their "best interests" be "a primary consideration" in all actions concerning them (Article 3), the "inherent right to life" (Article 6), and the right of a child "who is capable of forming his or her own views . . . to express these views freely in all matters affecting the child" (Article 12) (United Nations 1989). The document therefore extends the range and scope of children's rights, sets out the importance of the best interests (welfare) principle, and includes participation or "will" rights to reflect the growing acceptance of the capacity of those under 18 years (as the Convention defined children) to be actively involved in decision making in matters that concern them.

Children and Rights

In the interest theory of rights, possession of a right confers protection of an interest of sufficient importance to impose duties on others that allow the right holder to enjoy that interest in question. On the will or choice theory of rights, to have a right is to have the power to enforce or waive a duty which correlates to the right. On this basis to have a right is the ability to choose and to act purposefully that many children (and indeed some adults) lack. According to the "will" theory of rights, children cannot have rights because they lack the necessary capacity for informed choice; according to the "interest" theory, they can. To deny children some rights on this basis seems counterintuitive. Adults have duties and responsibilities to children, their number and scope depending on the adult's relationship to the child, e.g., parent, teacher, trustee, legal representative. It seems possible on the "will" theory that children could have rights that are exercised or held in trust by responsible adults, e.g., parents, trustees, or legal representatives.

Child liberationists claim that children have all the moral rights that adults do. Others are skeptical, for theoretical and political reasons, about attributing rights to children. Archard (2004) has summarized their arguments as:

(a) Children are not qualified as right holders because they lack capacity.
(b) Ascription of rights to children is inappropriate because it displays a misunderstanding of what childhood is, what children are like, or what relationships children stand in to adults. Ascription of rights may be harmful because of the possible adverse effect that their ascription might have on the family dispositions and emotional and psychological bonds that render the need for rights and rules of justice unnecessary in the first place.
(c) Children can be assured of adequate moral protection by other means.

These arguments seem to run counter to contemporary concepts of childhood with its concomitant obligation to recognize and enhance the status and interests of children. The language of rights provides a critical reminder of the extent to which children are maintained in an artificial condition of dependence and vulnerability by the actions of adults, without the opportunity to make their own choices. Granting children rights promotes the development of capacity, while their denial maintains this state of dependence and vulnerability. In response O'Neill (1988, 459–463) has asserted that rights talk misses what is distinctively different about children as a group. Childhood is a dynamic stage of human development in the transition from childhood into adulthood through which all must pass, rather than a permanently maintained status of oppression or discrimination. Denying that children have moral rights does not remove the strong obligation upon adults to facilitate this transition for *all* children in their care. Those who deny children all or some of the rights possessed by adults nevertheless believe that children, as humans, have a certain moral status that ought to be protected.

If adults have some rights that children do not possess, it should be possible to draw a line that distinguishes them. If the distinction depends on having the

capacity to exercise meaningful choice, it is obvious that many older children have the same levels of capacity as adults. Evidence from sociological and psychological research (Weithorn and Campbell 1982; Kuther and Posada 2004; Matthews 2009), together with the data from neuroimaging studies demonstrating maturation of brain areas concerned with rational decision-making, supports this proposition. Lack of opportunity to exercise choice may also limit the acquisition of relevant life experience which may also be necessary for the exercise of informed choice.

If possession of some rights depends on the capacity to make choices, then a low level test of capacity will mean that all adults and many children have rights; a high level test will mean that many children and a significant proportion of adults do not have moral rights. Societies habitually distinguish between adults and children by age since the latter broadly correlates with capacity. According to Cohen, "[a]ny line which uses age to distinguish people with rights from people without can be shown to be arbitrary" (Cohen 1980, 48). This applies to many dividing lines that are used to delineate gradual transition between stages. Objections to the use of age as a dividing line between childhood and adulthood may be that use of *any* age is wrong or that use of this *particular* age, e.g., 18, is wrong. Age-related thresholds clearly do exist when age disparities are considerable; to deny some rights on grounds of age is not to deny all rights. However to deny that different capacities are progressively acquired at different ages is implausible. Decisions to vote, marry, or join the armed services presuppose different levels of understanding and autonomy. It is true that the differences between 17¾-year-olds and 18-year-olds are small, but are an essential price to be paid in accepting any age-related threshold and that a distinction needs to be made.

If the capacity to choose is important in allocating some rights, a test of individual competency seems a defensible alternative to age in determining who should have a particular right. But applying such tests has administrative, methodological, and logistical difficulties that make their routine use inappropriate and impracticable.

If the capacities needed for exercising different rights are themselves different, and capacity is related to age, it is reasonable that different rights should be acquired at different ages and in proportion to the degree of competence required. Once a child has acquired the level of competency to achieve a particular right, she/he should not be refused another kind of right that presupposes the same level of competence. This is consistent with Kohlberg's theory of stages of moral development.

On the standard view, children have welfare but not liberty (choice) rights, whereas adults have both. Parents can and do overrule a child's choices; they can make decisions about a child's education, domicile, and general upbringing. Separate children's rights are those possessed by reason of the properties they have as children, e.g., inability to provide their own food and shelter, vulnerability to abuse and exploitation, and a need to be loved. In the health context, children's patterns of illness and disease susceptibilities and their different physiological, pathological, and pharmacological characteristics seem to require specific welfare rights in addition to general protection rights.

Developmental Potential and Rights

Since children develop into adults, the question arises as to what sort of adults they have the right to become and what rights are necessary to support this process. Feinberg used the term "right to an open future" to describe the anticipatory autonomy and welfare rights given to the child in the person of the adult she/he will become (Feinberg 1980). Eekelaar used the term "developmental" rights (Eekelaar 1986) as being the rights of a child to develop her/his potential so that she/he enters adulthood without disadvantage. An important consideration is how wide the scope of a child's future choices should be set. A maximal interpretation in which the child has the greatest opportunity to exercise the maximum amount of choices seems overambitious and a minimalistic version too restrictive; a better alternative is to suggest that a child should have enough autonomy to be able to make reasonable future life choices. Some children by reason of physical or mental impairments may never develop into autonomous adults and as such may lack developmental rights. Nevertheless they do have welfare or protection rights, even if their interests in becoming an adult are restricted.

Best Interests and Rights

Adults who have the capacity for decision-making have the right to decide what is best for them because they have liberty rights. Children do not have the same liberty rights or at least the capacity to claim them that adults have (for general discussion on ethics of surrogate decision-making, see Buchanan and Brock 1998). Conventionally their interests are protected by application of the best interests principle, which may also been formulated as the [less onerous] welfare principle (see Archard 2004, also UK Children Act 1989), since the latter does not seem to carry the maximizing obligation that *best* interests does. The weight which attaches to the principle varies according to this formulation: Article 3 of the United Nations Convention on the Rights of the Child states that "In all actions concerning children .....the best interests of the child shall be *a primary* consideration" and the Children Act 1989 [S1(1)] of the UK states that "the child's welfare shall be the Court's *paramount* consideration."

There are no clear ethical or legal definitions of the principle. Philosophically it is closely associated with concepts of welfare or well-being of which several theories exist (see Parfit 1984, Crisp 2015):

(a) Mental state, e.g., pleasure
(b) Desire fulfillment, in which a person's best interests are served by them getting what they want – irrespective of the pleasure/happiness it brings, e.g., wealth
(c) The objective list, which ordains that certain things improve an objective view of a person's welfare, irrespective of whether or not they are desired or lead to pleasurable mental state

The latter version is most likely to be applied by independent third parties who have to make decisions on behalf of another, but there are difficulties with it.

Legally the best interests principle was derived from the parens patriae jurisdiction, when courts made judgments about the affairs of those who could not decide for themselves. It was applied to health cases in US and UK case law and underpins decision-making in the UK Mental Capacity Act of 2005.

In clinical medicine an action or intervention is in a patient's best interests if overall benefits outweigh overall burdens, there is acceptable minimization of harm, and there is respect for as much autonomy (capacity for self-directed, informed choice) as the individual is capable of exercising. In the UK and elsewhere, both statute and professional guidance list the factors that need to be taken into account in determining best interests of children. These are in addition to clinical considerations given above and require consideration of:

- The ascertainable wishes and feelings of the child concerned in the light of their age and understanding and their previously stated wishes
- Their physical, emotional, and educational needs
- The likely effect of change of circumstances for the child and their family
- The cultural, religious, or other beliefs and values of the child or parents.
- The views of parents and others close to the child and their ability to care for the child
- The views of other healthcare professionals involved, whether or not directly providing care
- Which choice will least restrict the child's future options produce less harm

The weight given to each of these factors depends on circumstances. Most guidance specifies the need to consider other relevant information in a way that is nondiscriminatory, not based on unjustified assumptions, is consistent with human rights, and that encourages children's participation and enhances competence.

There are problems with the best interests principle and its application. Firstly, a maximal view of welfare is unfeasibly demanding. Moreover it takes insufficient account of the interests of others who are involved, e.g., parents, siblings, and exceeds the duties required of them in everyday life. A family welfare view seems a better option, because family and child are interdependent and have interlinked interests and shared values that society acknowledges. On this basis there should be at least equal consideration of the interests of relevant others; the interests of a sick child should not trump all other interests. Secondly, an objective view of what is best for any given child, in terms which of several different kinds of life would be best for them, is difficult to reach – even in clinical terms. It becomes even more so when values and cultural and other circumstances are taken into account. What is best for a child may depend on the circumstances in which they find themselves. There may be limited cross-cultural agreements as to what is best for any child and even within a culture views will differ, e.g., on limited physical punishment. Thirdly, a more subjective view of best interests in which "We must choose for others as we have reason to believe they would choose for themselves if

they were at the age of reason and deciding rationally"(Rawls 1999, 183) is also problematic. When adults make choices for other adults who have lost their capacity, they may have a good idea as to what the adult would have chosen, based on the knowledge of that person's wishes and beliefs etc.; this is not usually the case with children. It seems impossible to know what a child would choose if she/he had rational powers of choice because the essential nature of childhood is not to have these powers. It is not possible to ask adults what they would choose in the child's situation because an adult would not be in that situation. Choice should be based on what is best for the child rather than what an imagined adult version of the child might choose.

Best Interests and the Child's Voice

All the checklists used to determine best interests emphasize the importance of listening to children and ascertaining their views and giving them appropriate weight. Article 12.1 of the United Nations Convention on the Rights of the Child asserts that "*States Parties shall assure to the child who is capable of forming his or her own views the right to express those views freely in all matters affecting the child, the views of the child being given due weight in accordance with the age and maturity of the child*" (United Nations 1989). However this participation right is not the same as the liberty right to choose and is not determinative. The developmental stage a child must reach before their views will be determinative and trump parental and societal views remains open to question.

The UK Gillick ruling is widely cited as providing some guidance as to the point at which this occurs. The most cited element of the judgment was made by Lord Scarman as

> The underlying principle of the law ... is that parental right yields to the child's right to make his own decisions when he reaches a sufficient understanding and intelligence to be capable of making up his own mind on the matter requiring decision.
>
> I would hold that as a matter of law the parental right to determine whether or not their minor child below the age of 16 will have medical treatment terminates if and when the child achieves a sufficient understanding and intelligence to enable him to understand fully what is proposed. (Gillick [1986], 186, 188–189)

The philosophical basis of the judgment has been debated. It appears to create a defined threshold beyond which the child can make important medical decisions without recourse to parents, provided she/he had sufficient understanding and intelligence to make the decision in question. However, it can also be interpreted as providing a proportionate or gradualist solution, as the child's views carry gradually greater weight as his/her understanding increases.

Central to the debate is the quantum of intelligence and understanding regarded as sufficient. The criterion is surely not based on a decision that coincides with a professional or parental view of best interests. A child's poor choice is not a conclusive evidence of a general incapacity to choose.

To demonstrate competence, a child must show they have knowledge of the facts, an understanding of what will be the likely consequences for themselves of any act or omission, and the ability to appreciate the significance of the act or omission and the relevant consequences for themselves and their family. In addition the level of the test for competency seems to depend on the seriousness and complexity of the question in hand in a way that does not apply to adults. Subsequent judgments have appeared to indicate that levels of competency required to consent to, or to refuse, particular forms of treatment in children differ. This seems illogical since if a child is competent, then she/he is in all significant and relevant respects the equal of an adult and should be able both to choose and to refuse treatment. The Gillick judgment has arguably set a threshold for determining a child's competence to make a decision that is higher than required for adults in whom capacity is assumed.

Children cannot exercise choice if they do not receive factual information in a form and at a pace that they can comprehend and in circumstances that optimize this process. Notions of what a child is and at what level adults regard them as operating have influenced the amount of information that children have received. But children can absorb complex information, make perceptive comments, and are able to pose philosophical questions and enter into a dialogue that their later adult selves might not. Adults, in their desire to protect children, may seek to filter out information that they consider harmful to the child, even though the child may already be aware of it. Adults may also make unwarranted assumptions about a child's level of understanding and knowledge, especially if there is physical rather than mental impairment, e.g., cerebral palsy affecting only motor function. They may also attribute a lower quality of life to the child than the child would do so for himself/herself. The UK legislation places a duty on professionals to enhance the capacity of those adults who appear to lack it. Philosophically there seems no reason why this obligation should not apply to those who care for children, so that children can exercise as much choice as they are capable.

Best interest is therefore an evolving concept; although it remains the standard for decision-making for those lacking capacity, it is subjective and individualistic. Furthermore it is unclear in the current economic and political climate how much – if at all – wider societal factors might be involved in the determination of best interests for individuals. Any widening of the scope of the best interest considerations would have important consequences for children who have only a limited political voice which is often only expressed by adults.

Conclusions

If children are small adults, there are clear implications for how they should be regarded and treated in the healthcare setting. Visual arts, literature, and the media have all had influences in determining how a given society perceives children and in particular whether they are regarded as small adults. However there are sound historical, social, biological, developmental, and philosophical reasons why they should not be regarded as small adults. This is not to say that their interests can be disregarded or that they should be treated as vulnerable, immature subjects who

need protection from both physical harms and their own impulsive unwise decisions. This obligation for protection of the vulnerable and weak does not mean that those who lack the capacity to make their own free choices are excluded from the consultation or decision-making process. There is a need to balance recognition of the child's vulnerabilities against their positive attributes – to focus on what they can do – and the need for protection from hazards, abuse, exploitation, and neglect against the need to recognize and respect their growing capacities to absorb information, make moral judgments, and take responsibility for their own actions. Over time and in response to social attitudes – influenced by increased understanding of neurodevelopmental, psychological, and educational research and media representations – the pendulum has swung between under- and overprotection, liberalism, and paternalism. Clearly, younger children do have greater but differing capacities for philosophical inquiry and understanding of themselves than had previously been supposed. It is likely that greater insight derived from personal accounts of childhood and other sources will continue to provide richness to the debate. It is no less than children deserve and no more than they need.

Definitions of Key Terms

Child	A form of human being having specific biological and developmental characteristics (of which some are immature versions of those of adults) but with their own unique interests.
Childhood	A phase in the transition between the dependency of infancy and the maturity and self-determination of adulthood, often divided into stages.
Stage theory	A theory of development in which the transition from one entity to another occurs in a series of stages that can be recognized and tested.
Cognitive development	The emergence of the child's ability to think and understand; it may be determined by innate qualities and personal experiences.
Moral development	The acquisition of the ability to make decisions about what is right or wrong.
Rights	Justifiable moral claims that entail on others the obligation to act or forebear; may be legal or moral and of various types, e.g., concerning liberty or welfare.
Best interests	The best interests principle is used to determine what actions are appropriate for those who lack the capacity to exercise liberty rights. Best interests may lack precise definition but their determination involves consideration of harms, benefits outcomes, an individual's wishes and preferences (insofar as they are able to express them), and those of relevant others.

Summary

- Children are not small adults; the Aristotelian developmental concept of childhood from immaturity to full moral capacity underpins much contemporary thinking.
- Depiction of children in art, literature, and the media has mirrored changing public perceptions of childhood.
- Increased understanding of the biological characteristics of children enables the identification of points of differences and similarities between adults and children and identification of patterns of vulnerability in the former.
- Children undergo a process of physical, cognitive, and moral development whose stages can be identified and may be relevant to determining their moral status.
- Children who lack the capacity of adults to exercise their liberty rights may still have welfare rights, and their interests can be protected in law by determination of the best interests using defined criteria.

References

Archard D (2004) Children: rights and childhood. Routledge, London
Ariès P (1962) Centuries of childhood. Random House, New York
Aristotle (1932) The politics (trans: Rackham H). Harvard University Press, Cambridge, MA
Aristotle (1934) Nicomachean ethics (trans: Rackham H). Harvard University Press, Cambridge, MA
Buchanan AE, Brock DW (1998) Deciding for others: the ethics of surrogate decision making. Cambridge University Press, Cambridge
Buck T (2014) International child law. 3rd edition Routledge London and New York
Carlson NR et al (2005) Psychology: the science of behavior, 3rd Canadian edn. Pearson Education. Pearsoned.com New York ISBN 0-205-45769-X
Chomsky N (1959) Review of B.F. Skinner's verbal behavior. Language XXXV:26–58
Cohen H (1980) Equal rights for children. Littlefield, Adams, and Co., Totowa
Crisp R. Wellbeing. In: Zaita EN (ed) The Stanford encyclopedia of philosophy, Summer 2015 edn. http://plato.stanford.edu/archives/sum2015/entries/well-being/
Descartes R [PW] (1985) The philosophical writings of descartes, vol 1 (trans: Cottingham J, Stoothoff R, Murdich D). Cambridge University Press, Cambridge
Eekelaar J (1986) The emergence of children's rights. Oxford J Legal Stud 6:161–182
Feinberg J (1980) A child's right to an open future. In: Aiken W, LaFollette H (eds) Whose child? Parental rights, parental authority and state power. Littlefield, Adams, and Co., Totowa, pp 124–153
Gillick v. West Norfolk and Wisbech Area Health Authority [1986] AC 112
Gilligan C (1982) In a different voice. Harvard University Press, Cambridge
Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. Neuron 44(2):389–400
Kohlberg L (1981–1984) Essays on moral development, vols I and II. Harper & Row, San Francisco
Kuther TL, Posada M (2004) 2004 children and adolescents' capacity to provide informed consent for participation in research. Adv Psychol Res 32:163–173
Locke J [EHC] (1959) An essay concerning human understanding, vol I. Dover, New York

Matthews GB (2008) Getting beyond the deficit conception of childhood: thinking philosophically with children. In: Hand M, Win Stanley C (eds) Philosophy in schools. Continuum, London, pp 27–40
Matthews GB (2009) Philosophy and developmental psychology: outgrowing the deficit conception of childhood. In: Siegel H (ed) the Oxford Handbook of the Philosophy of Education,; Oxford handbooks OUP Oxford pp 162–176
O'Neill O (1988) Children's rights and children's lives. Ethics 98:445–463
Parfit D (1984) Reasons and persons. Clarendon Press, Oxford
Piaget J (1968) The mental development of the child. In: Elkind D (ed) Six psychological studies. Vintage Books, New York, pp 1–73
Piaget J (1971) The child's construction of quantities. Routledge/Kegan Paul, London
Rawls J (1999) A theory of justice, Rev edn. Oxford University Press, Oxford
Rousseau J-J (1979) Emile or on education. Basic Books, New York
Shahar S (1990) Childhood in the Middle Ages. Routledge, London
Spock B (1968) Baby and child care, 3rd edn. Hawthorn Books, New York
Stephenson T (2005) How children's responses to drugs differ from adults. Br J Clin Pharmacol 59 (6):670–673
United Nations (1989) The convention on the rights of the child. In: Alston P, Parker S, Seymour J (eds) Reprinted in Children, rights and the law. Oxford University Press, Oxford, pp 245–264; and in LeBlanc L. The convention on the rights of the child. United Nations Law making on Human Rights. University of Nebraska Press, Lincoln, pp 293–316
Weithorn T, Campbell SB (1982) The competency of children and adolescents to make informed treatment decisions. Child Dev 53:1589–1598
World Health Organization. Children are not small adults. http://www.who.int/ceh/capacity/Children_are_not_little_adults.pdf

Extending Human Life as an Aim of Medicine

26

Søren Holm

Contents

Abstract

This chapter considers whether life extension per se can be an aim of medicine. It first provides an initial tentative analysis of what it means for something to be an aim or goal of medicine in general and whether there are any normative restrictions on what can be an aim or goal of medicine. It then applies this analysis to moderate and radical life extension. It is concluded that moderate life extension has many affinities with other aims and goals that are accepted as uncontroversial goals of medicine and that there are no conclusive arguments showing that it falls

S. Holm (✉)
School of Law, The University of Manchester, Manchester, UK

Center for Medical Ethics, HELSAM, University of Oslo, Oslo, Norway

Department of Health Science and Technology, Aalborg University, Aalborg, Denmark
e-mail: soren.holm@manchester.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_18

outside the scope of proper medicine. It is further concluded that radical life extension falls outside the current scope of medicine and that medicine as a normative practice would have to be significantly reconfigured if radical life extension is to become an aim of medicine.

Introduction

Saving life that is threatened by disease is one of the traditional aims of medicine, and given that saving a life that would otherwise have ended is by definition extending that life, there can be no doubt that extending life is an aim of medicine. This however leaves open the question whether extending life in itself is a goal of medicine in the absence of any disease process or debilitating condition – what we could call "pure life extension" – or whether all forms of medically mediated life extension fall within the aims and goals of medicine. It is this latter question that will be the focus of this chapter.

The reasons that these questions have become topical are (1) the rapidly expanding possibilities for biological life extension and (2) the equally rapidly expanding health-care costs associated with interventions that have only a marginally life-extending effect. Average, median, and modal life expectancy is continually increasing in the developed world, partly as a result of medical progress, and the gradual increase in life span is likely to continue. There are also groups that are predicting that we will soon be able to radically extend life. It has, for instance, been claimed that the first person to live to 1,000 years of age may already be alive today (de Grey 2005, 2006). Significant life extension has already been demonstrated in a range of animals from nematodes (especially *Caenorhabditis elegans*) to rodents. It is, however, also important to note that the quest for radical life extension and some form of immortality is not a new quest but a long-standing human ambition (Gray 2011).

The chapter falls in three sections. It first provides a tentative analysis of what it means for something to be an aim or goal of medicine in general and whether there are any restrictions on what can be an aim or goal of medicine. This analysis is then sequentially applied, first to moderate and then to radical life extension.

The Aims and Goals of Medicine

What does it mean for something to be an aim or goal of medicine? Historically there have always been a set of activities that the medical profession engaged in and goals that it aimed for and a set of activities and goals that were seen to fall outside the scope of proper medical practice. In the Hippocratic Oath, the Hippocratic physician, for instance, solemnly pledges himself to both positive and negative obligations. He will:

> . . . apply dietetic measures for the benefit of the sick according to my ability and judgment; I will keep them from harm and injustice.

But he will not:

> . . . give a deadly drug to anybody who asked for it, nor will I make a suggestion to this effect. Similarly I will not give to a woman an abortive remedy. In purity and holiness I will guard my life and my art.
>
> I will not use the knife, not even on sufferers from stone, but will withdraw in favor of such men as are engaged in this work. (Edelstein 1943, p. 3)

This indicates that in the Hippocratic tradition, something could fall outside of the scope of medicine, either because it was not a proper thing to do or not related to a goal that one should aim for (prescribing poisons and abortifacients), or because it fell outside the set of skills possessed by, or below the dignity of physicians (using the knife).

The set of skills possessed by members of the medical profession is malleable over time by changes in education and training. Today, the medical profession encompasses both physicians and surgeons, and most of the status differences between physicians and surgeons have disappeared. The Hippocratic view that using the knife is outside the scope of medical practice is therefore now obsolete. So the mere fact that some kind of life extension requires skills that are not currently possessed by the medical profession cannot definitively rule it out as a possible aim or goal of medicine.

For the following analysis, it is useful to make a distinction between the fundamental goals of medicine, i.e., the ultimate aims that medical practice is directed at achieving, and the operational goals such as diagnosis, caring, curing, and preventing that are pursued as means to achieving the fundamental goals (Fleischhauer and Hermerén 2006). A similar way of dividing up the goals of medicine is in final goals and instrumental goals. Liss, for instance, argues that the final goal of medicine is a nonmedical goal, i.e., welfare and that health is only an instrumental goal (Liss 1996).

A number of goals have been proposed as the goals of medicine, and there is quite considerable overlap between the lists produced by different authors who all accept that medicine has fundamental goals that are internal to the practice. An international project led by the Hastings Center in the 1990s produced a list that is fairly characteristic stating that the goals of medicine are (Allert et al. 1996):

- The prevention of disease and injury and promotion and maintenance of health
- The relief of pain and suffering caused by maladies
- The care and cure of those with a malady and the care of those who cannot be cured
- The avoidance of premature death and the pursuit of a peaceful death

However, the fact that life extension per se is not on this list or on similar lists produced by others cannot in itself be taken as conclusive proof that life extension could not fall within the fundamental or operational aims or goals of medicine.

The interesting question in the present context is therefore whether there are fundamental or operational aims or goals that are normatively outside the scope of proper medicine and whether life extension belongs to the class of excluded goals. A goal could be normatively outside the scope of proper medicine in at least four different ways:

1. It could be a goal that no one should pursue, making it incidental and trivial that the medical profession should not pursue this goal.
2. It could nontrivially be a goal that is outside the scope of medicine specifically, but which is not generally normatively prohibited. Such a goal need not be uniquely excluded for medicine; it can also be outside the scope of other practices or professions, e.g., nursing. It has, for instance, been argued that purely aesthetic surgery falls outside of medicine proper, because it solely pursues a nonmedical goal (i.e., beauty).
3. It could be a goal that the medical profession pledges itself to not pursuing, either individually or collectively.
4. It could be a goal that some legitimate normative authority inside or outside of the profession proscribes.

We can briefly dispose of the last possibility because there is no plausible candidate for being the legitimate normative authority that can proscribe and/or prescribe the aims or goals of medicine. There are national regulators that have legal authority to regulate the medical profession in the particular jurisdiction, but there is no plausible candidate for a global normative authority. The two most plausible candidates are perhaps the World Health Organization (WHO) and the World Medical Association (WMA), but neither of these organizations possesses the de facto and de jure legitimacy necessary to proscribe and/or prescribe the aims or goals of medicine. In relation to the development of WMA declarations and guidelines stating or implying goals of medicine, it is, for instance, evident that what really decides whether or not a guideline will be followed by the majority of doctors is not that it has been issued by the WMA but that it accords with what most doctors find is a reasonable stance.

The third possibility is what could be called the "traditional model," following on from the Hippocratic Oath and the many other oaths and pledges that doctors have sworn or are swearing now when they are being inducted into the medical profession. On this model the aims and goals enunciated in WMA declarations and guidelines (and any similar documents) obtain their normative force, not by being issued by the WMA or by Hippocrates but because doctors pledge themselves to the guidelines. This approach to defining the aims and goals of medicine is however deeply problematic in a number of ways. There is the practical problem that doctors are not directly members of the WMA but of national organizations who are members of the WMA, and many of these national organizations do not have 100 % of a nation's doctors registered as members. For instance, in 2011 only about 15 % of US doctors were members of the American Medical Association (Collier 2011). So, in what way are doctors who are not members of their national

medical association, or doctors whose national medical association is not a member of the WMA, bound by the aims and goals enunciated by that organization? If it is the individual voluntary acceptance of the obligations that is doing the normative work here, we have to conclude that nonmembers are not bound at all. And, if it is an idea of collective acceptance, then only doctors in countries that fulfill two conditions can be normatively bound, i.e., doctors where (1) the national medical association can be legitimately viewed as representing all doctors, including nonmembers collectively, and (2) the medical association is a member of the WMA. Furthermore the oaths, pledges, declarations etc. are not univocal as to the limits of proper medicine, so a doctor may normatively be committed to two (or more) inconsistent set of aims and goals. And, again if it is merely the voluntary individual or collective acceptance that is doing the normative work, then it is not clear how to resolve conflicts between the equally binding normative commitments. Finally it is unclear within this model whether doctors are bound by the first oath or pledge they affirm (either explicitly or implicitly) if that oath or pledge is later changed or a completely new one is introduced. There is also a more fundamental problem with the traditional model. It is simply not clear why anyone outside of the medical profession should accept that the profession itself sets its aims and goals and consequently the limits of proper medicine and/or should feel bound to respect those limits. Doctors can pledge themselves to what they want, but that binds no one else, least of all the state.

This leaves the second possibility as the only viable option. For the second possibility to obtain, there must be something specific about medicine as a practice and/or medicine as a profession that generates the normative restriction in scope, i.e., there has to be some internal normativity.

In this connection it is important initially to see that the content of any internal normativity cannot be decided by the medical profession alone. It might be argued that the profession is in a special epistemic position in relation to discerning what the proper normative scope of the profession is. But unless a very strong version of standpoint epistemology is adopted, claiming that only doctors can understand what medicine is about, the conclusion that only the medical profession can define the scope of medicine cannot be secured. And such a strong version of standpoint epistemology is implausible. It may well be the case that doctors are in a better position to understand what medicine is about, because they practice and experience medicine every day and still be the case that a determined medical sociologist could also understand what medicine is about. It is also important to note that if a strong version of standpoint epistemology is used to advocate for the special role of the profession, then it directly undermines any claim the profession might want to put forward in relation to being able to speak for patients (Holm 2011).

How can an internal normativity of medicine as a practice or a profession be generated and justified? An internal normativity of medicine could be generated in a number of different ways; it could be a consequence of the historical development of the profession; it could be a result of the current configuration of medical practice and the position of the profession vis-à-vis other professions, clients, and society; or it could be based on necessary features of the relation between a healer and a person who wants to be healed (Brody 1993).

The two first options both entail, although in two different ways, that the internal normativity of medicine is historically contingent either directly or through the fact that the current position of medicine is historically contingent. They also entail that it is difficult to define exactly what kind of normativity that is generated; is it ethical, professional, social, societal, etc. or all of these at the same time?

In order to produce an argument for a general and noncontingent restriction of the aims of medicine, it is therefore necessary to pursue the third option, i.e., the aims can be derived from the relation between a healer and a person who wants to be healed. This is somewhat circular since it defines medicine as a healing profession, but it is not viciously circular. Even if it is accepted that the final goal of medicine is welfare, what sets the medical profession apart from other professions is that it pursues welfare in a specific way, i.e., through healing.

Some of the features of medical practice that have been claimed to be relevant in relation to defining the proper scope of medicine in other contexts do not seem to be even prima facie relevant to the context of life extension. In discussions of abortion and euthanasia, it is sometimes claimed that medicine is a healing profession and that this rules out direct killing of human beings by doctors, and also in relation to euthanasia, it has been argued that medical involvement in euthanasia will necessarily undermine trust in the profession and that trust between doctors and patients is a necessary requirement for proper medical practice to take place. But life extension does not involve killing, and it is difficult to see that medical participation in life extension could undermine trust in the profession, if the extension is wanted by the patients and if it does not mean that other valued activities of the profession are displaced (see also the last section on how life extension should be balanced against other goals of medicine).

However, it is relevant to an evaluation of life extension as a potential goal of medicine that pure life-extending interventions are not directly related to health or disease. It may seem trivial, but it is important to remember that it is possible to pass from life to death in a very short span of time, without any appreciable intervening period of disease or unhealth.

Pure life-extending interventions are not directly contrary to health-related aims, they are clearly not comparable to, for instance, medical participation in torture where a goal of harming is pursued, but they may tentatively be defined as being outside the scope of medicine as a healing profession. The next sections will explore the extent to which this tentative conclusion can be sustained for moderate and radical life extension.

Moderate Life Extension

Could moderate life extension nevertheless be an aim of medicine? That is, could it be legitimate for doctors to perform medical interventions with the sole or primary aim of securing moderate extensions of life span, i.e., extensions within or somewhat beyond the current human maximal life span? It is at present difficult to conceive of an actual intervention that has life extension as its sole result, because all current

life-extending interventions also at the same time either prevent or treat disease or form part of caring for persons with disease or disability. But this does not mean that there could not be pure life-extending interventions. There could, for instance, be genetic changes with the sole effect of life extension.

Life Extension as a Side Effect

Life extension may come about as a side effect of medical procedures that have a primary aim that falls squarely within the traditional goals of medicine as discussed above. Vaccination against childhood diseases has as a primary aim to prevent disease, but a predictable side effect of a comprehensive vaccination program is that it will also lead to life extension. Not having had serious childhood diseases will mean that people are slightly more healthy in adulthood than they would otherwise be and will live slightly longer on average. Or to give another example, osteoporosis in the elderly is treated in order to prevent fractures, and this has life extension as a predictable side effect.

In the near future, it is likely that treatments will be developed for other conditions that affect the old, e.g., the characteristic slowly progressing myopenia (loss of muscle mass) (von Haehling et al. 2012), and these will also have life extension as predictable side effects.

However, if life extension is a predictable effect of a particular medical intervention, it raises the question whether life extension should not be included as one of the aims of that intervention. It is uncontroversial that a medical intervention can have more than one aim. Stabilizing a bone fracture by internal fixation, for instance, reduces pain and increases the chance of healing in a good position, and both of these effects are aims of the intervention. So what about predictable life extension? Two situations can be distinguished: (1) intervention I has life extension as a predictable effect, but we would pursue I in order to achieve its primary medical aim A, even if there was no life extension and (2) intervention I has life extension as a predictable effect, and we would not pursue I in order to achieve its primary medical aim A, if there was no life extension, only the combination of A and life extension makes I worth pursuing.

In the first situation, it is possible to persist in ignoring life extension as an independent aim of the intervention, by only focusing on the doctor's intention when prescribing/performing the intervention. It is of course an empirical question what intention a particular doctor has in a specific situation, and she/he may well intend both the primary medical aim A and life extension, even if A is sufficient for prescribing. This would simply be a case of overdetermination of intention. But the intention in the second situation cannot be adequately analyzed or described unless the intention to aim for life extension is identified and stated as part of the overall intention.

On the basis of an analysis of the question of responsibility, it is obvious that the doctor is causally responsible for the life extension in both situations. The intervention is a necessary causal factor for the occurrence of the particular instance of life

extension. And, if a consequentialist account of moral responsibility is accepted where responsibility tracks foreseeable consequences of actions (Harris 1980), it is impossible to ignore the side effect of life extension in any of the two scenarios. The doctor will be causally and morally responsible for the (positive) consequence of life extension in both.

It is therefore difficult to escape the conclusion that medicine is already pursuing moderate life extension as one of its many aims and goals, simply by virtue of the fact that life extension is a known side effect of many bona fide medical interventions and that, because it is a positive side effect, it is a side effect that doctors deliberately aim at producing.

Is Aging a Disease?

Another possible argument for life extension being a legitimate aim of medicine is based on the premise that aging in itself is a disease and that since treating diseases is one of the uncontroversial aims of medicine, treating aging and thereby achieving life extension is also a legitimate aim of medicine. The soundness of this argument pivots on the truth value of the premise "aging is a disease." If this premise is true, the argument is valid and sound.

Following the WHO's all-encompassing definition of health as "A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity," the aged state can undoubtedly be characterized as a state which is not completely healthy (WHO 1948). It is usually not a state of complete physical, mental, and social well-being, and it may include conditions that uncontroversially count as either diseases or infirmities. Old age is, thus, for most a stage of unhealthy life, although not necessarily a state of disease and being unhealthy is linked to loss of welfare and suffering. There may, thus, irrespective of any life-extending effects be good reasons to develop anti-senescence interventions that would allow us to increase the proportion of healthy life span and "square the curve," i.e., ensure that morbidity in old age was compressed to a very small timespan prior to death (Fries 1980; Bostrom 2005; Andersen et al. 2012).

The proper analysis of the concept of disease is one of the central questions in the philosophy of medicine, and this is not the place to provide an exhaustive account of the debate and the many positions taken (Hofmann 2001). However, a distinction between putatively value-free analyses, analyzing disease as necessarily involving a deviation from what it species-typical or normal, and normative accounts, analyzing disease primarily as any biological condition leading to loss of welfare or restriction of action, is useful here.

Aging is undoubtedly species typical for humans (as for most other organisms). All humans age not only chronologically but also biologically. This seems to speak strongly against defining aging in itself as a disease (Schramme 2013). There may also be reasons of ontology that speak against defining aging as a disease, since it is far from clear that we can individuate a particular biological condition as "aging."

What is perceived as a single process of aging may be simply an outward appearance that supervenes on a large number of distinct biological processes.

On the other hand, from the perspective of a theoretical stance seeing disease as unlinked to conceptions of normality, aging very much looks like a disease as Caplan very succinctly argues in the abstract to another chapter in this handbook:

> Unless one is so concerned about the social or economic consequences of doing so, it is hard to see why aging ought not be characterized as a disease. The changes associated with aging, unlike those associated with growth and sexual maturation, are manifestly dysfunctional. The causes of the dysfunctional changes that fuel senescence are clearly rooted in the loss, collapse, or deterioration of cellular functions.
>
> [. . .]
>
> The fact that they occur for almost all people at advanced ages does not make them any less dysfunctional relative to the experience of the individual in terms of "symptoms" or the overall ability of the person beset by these changes to flourish and survive. Aging is a disease. (Caplan 2015)

The debate concerning the correct analysis of the concept of disease is not likely to be concluded, since (1) both sides of the argument have an array of good arguments and counterarguments and (2) the main areas of contention are connected to deep underlying theoretical disagreements about metaphysics and ontology (Hofmann 2001). It is therefore not possible to state conclusively whether or not aging is a disease, a cluster of diseases, or merely an epiphenomenon supervening on a complex set of more basic biological conditions.

Radical Life Extension

Let us roughly define radical life extension as any life extension that increases the median life expectancy to 200 years or more. Could the pursuit of radical life extension be an aim of medicine? Radical life extension differs from moderate life extension in that it is unlikely to come about as a by-product or side effect of interventions primarily aimed at preventing or curing disease. Radical life extension in a biological body is likely to require significant reorganization of many aspects of basic human cellular biology (de Grey et al. 2002). This means that the arguments discussed above in relation to moderate life extension as a side effect and life extension as treatment of the disease of aging are irrelevant in relation to an assessment of radical life extension as an aim of medicine.

The set of skills necessary for applying the new life-extending interventions are likely to be extensions of the skill set already possessed by doctors, so could radical life extension nevertheless be or become an aim of medicine? One argument showing that radical life extension is a plausible aim of medicine would rely on Liss's identification of the final, extra-medical goal of medicine as welfare. Although radical life extension does not necessarily make people healthier, it does make it possible for them to have more welfare in their lives, simply because those lives are much longer. Further argument would still be needed to show that life extension

interventions should fall within the scope of the medical profession and not of some other, perhaps newly established profession. But, there would be no in-principle arguments for excluding radical life extension as a possible aim of medicine since radical life extension is compatible with the stated final goal of well-being or welfare.

Against this it can be argued, as it has in relation to WHO's definition of health, that if there are no restrictions on the kind of well-being or welfare that medicine should try to promote or create, then every problem and policy affecting welfare becomes a medical problem that ought to be the target of medical concern. The argument for accepting radical life extension as a legitimate aim of medicine will therefore commit us to an even more radical extension of the proper scope of medical concern in many other areas.

How Important Is Life Extension as an Aim of Medicine?

That something is a legitimate aim of medicine does not tell us how important it is or what weight we should attribute to it when balancing it against other medical or societal aims. If relief of suffering and life extension are, for instance, both legitimate medical aims, there will be many situations where we will need to be able to say how important each of them is in a particular context. This can be an important question in relation to a single patient, e.g., a patient who can only achieve life extension that entails some future suffering, and it can also be important in relation to resource allocation either between groups of patients where life extension for some is "bought" at the price of suffering for others or vice versa or in relation to the allocation of resources to different research programs, e.g., should life extension research have higher or lower priority than research aimed at developing new treatments for diseases.

In relation to the single-patient case, the balancing problem can in many cases be solved by simply asking the patient what she/he prefers or if that is not possible by referring to the patient's own values and choices. A general answer to the question of how important life extension is as a goal of medicine is thus not required in these cases.

In the resource allocation context, there is, however, a need for a general answer both to the question of how life extension is to be balanced against other medical goals and to the more specific question of how life extension is to be balanced against other medical goals for the young and for the old (see, for instance, Callahan 1995).

The most difficult of the questions is the question of research priorities because an answer must be based on a number of uncertain predictions, e.g., how likely is it that any given research program or group of research programs will be successful and on an answer to a very basic, but extremely controversial philosophical question, i.e., "how bad is it (ethically speaking) that we are mortal and have finite life spans?" That last question will not be answered conclusively here, but the two main possible positions will be outlined. It might be worth noting

initially that unless we achieve real immortality, the human death rate will continue to be 100 %, and life extension is not going to change that. Even under conditions of radical life extension, every single person who is born will also, eventually, die.

The positions on the importance of life extension research tracks the consequentialism/non-consequentialism divide in ethics to a significant degree. Those who are consequentialist argue that every extra year of (healthy) life is valuable and that we have no rationale for either focusing on relieving suffering and/or focusing on ensuring that as many people as possible attain a full, current life span. If life extension research is likely to produce most good in the long run, then it should be given the highest priority (Bostrom 2005). Those who are not consequentialist argue that we should focus on peoples' current suffering and problems and that while adding extra years of life to what is currently a standard life span may be a good thing to do, it is clearly outweighed in the priority stakes by the real health problems people have now (Callahan 1995). Life extension research is only an optional extra as long as there is real current illness and suffering.

Both positions make sense within their own specific framework of moral philosophy and little sense within the opposed theoretical framework. It is therefore difficult to see how the question can be answered definitively prior to a fundamental theoretical breakthrough (or rapprochement) in moral philosophy (Parfit 2011).

Definitions of Key Terms

Aim of medicine	A goal toward which medical practice is properly directed. A distinction can be made between the fundamental goals of medicine, i.e., the ultimate aims that medical practice is directed at achieving, and the operational goals such as diagnosis, caring, curing, and preventing that are pursued as means to achieving the fundamental goals. A similar way of dividing up the goals of medicine is in final goals and instrumental goals.
Internal normativity	Binding normative rules, principles, restrictions, etc. generated within a particular practice and justified by features of that practice.
Moderate life extension	Life extension leading to a predicted, increased life span within the current or slightly beyond the current possible range for human life (122 years).
Pure life extension	Life extension achieved by an intervention that is not a treatment for disease or infirmity and that does not improve health.
Radical life extension	Life extension leading to a predicted, average, increased life span greater than 200 years.

Summary Points

- Any attempt at answering the question of whether pure life extension can be an aim or goal of medicine actualizes important and fundamental disagreements relating to the possibility of an internal normativity within the practice of medicine.
- Any attempt at answering the question of whether pure life extension can be an aim or goal of medicine actualizes important and fundamental disagreements concerning the concept of disease.
- Any attempt at answering the question of how important pure life extension is as a goal, compared to other goals, actualizes important and fundamental disagreements in moral philosophy.
- It is arguable that moderate life extension is already an aim of medicine.
- The arguments that support radical life extension as an aim of medicine also entail a further significant reconfiguration of medicine as a normative practice.

References

Allert G, Blasszauer B, Boyd K, Callahan D (1996) The goals of medicine: setting new priorities. Hastings Cent Rep 26(6):S1–S27

Andersen SL, Sebastiani P, Dworkis DA, Feldman L, Perls TT (2012) Health span approximates life span among many supercentenarians: compression of morbidity at the approximate limit of life span. J Gerontol A Biol Sci Med Sci. doi:10.1093/gerona/glr223

Bostrom N (2005) The fable of the dragon tyrant. J Med Ethics 31:273–277

Brody H (1993) The healer's power. Yale University Press, New Haven

Callahan D (1995) Setting limits: medical goals in an aging society with a response to my critics. Georgetown University Press, Washington, DC

Caplan A (2015) How can aging be thought of as anything other than a disease? In: Schramme T, Edwards S (eds) Handbook of the philosophy of medicine. Springer, Dordrecht

Collier R (2011) American Medical Association membership woes continue. Can Med Assoc J 183: E713–E714

de Grey AD (2005) Like it or not, life-extension research extends beyond biogerontology. EMBO Rep 6:1000

De Grey AD (2006) We will be able to live to 1000. In: Moody HR (ed) Aging: concepts and controversies. Sage, London, pp 66–68

Edelstein L (1943) The hippocratic oath: text, translation, and interpretation. Johns Hopkins Press, Baltimore

Fleischhauer K, Hermerén G (2006) Goals of medicine in the course of history and today. Kungl. Vitterhets Historie och Antikvitets Akademien, Stockholm

Fries JF (1980) Aging, natural death, and the compression of morbidity. N Engl J Med 303:130–135

Gray J (2011) The immortalization commission: science and the strange quest to cheat death. Macmillan, London

Grey AD, Ames BN, Andersen JK, Bartke A, Campisi J, Heward CB, McCarter RJ, Stock G (2002) Time to talk SENS: critiquing the immutability of human aging. Ann N Y Acad Sci 959:452–462

Harris J (1980) Violence and responsibility. Routledge and Kegan Paul, London

Hofmann B (2001) Complexity of the concept of disease as shown through rival theoretical frameworks. Theor Med Bioeth 22:211–236

Holm S (2011) Who should decide the content of professional ethics? J Appl Ethics Philos 3:1–8

Liss PE (1996) On the notion of a goal – a conceptual platform for the setting of goals in medicine. In: Nordenfelt L, Tengland PA (eds) The goals and limits of medicine. Almqvist & Wiksell International, Stockholm, pp 13–31

Parfit D (2011) On what matters, vol 1–2. Oxford University Press, Oxford

Schramme T (2013) 'I hope that I get old before I die': ageing and the concept of disease. Theor Med Bioeth 34:171–187

von Haehling S, Morley JE, Anker SD (2012) From muscle wasting to sarcopenia and myopenia: update 2012. J Cachex Sarcopenia Muscle 3:213–217

World Health Organization (1948) Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p 100) and entered into force on 7 Apr 1948

Part III

Patients

Holism in Health Care: Patient as Person

27

Simon Woods

Contents

Abstract

This chapter explores the place of holism within the philosophy of medicine. It is organized into three sections which explore three related aspects of holism in the context of medicine and health care. The chapter begins with an introduction which places the holism debate within the wider history of ideas and the evolution of medicine as a scientifically grounded practice. An early consideration in section one is whether holism can be positively defined or whether it can be discussed only in terms of its relationship with reductionism – the holism/reductionism dyad. A key question is whether medicine's reliance upon the natural sciences requires the adoption of a form of reductionism that is

S. Woods (✉)
PEALS (Policy, Ethics and Life Sciences) Research Centre, Newcastle University, Newcastle upon Tyne, UK
e-mail: simon.woods@newcastle.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_21

incompatible with holism. Developing the analysis of reductionism the question of the compatibility of reductionism with holism is explored further. Section two deals with the question of whether holism is actually antireductionist. Here several strands are explored in order to show that the holism/reductionism dyad is not merely a scientific, philosophical debate but in addition has important normative implications for understanding the very raison d'être of medicine. Section three continues the analysis of the normative issues by exploring the holism/reductionism dyad in the context of medical ethics.

Introduction

Any account of holism invites an expansive and complex discussion, and so inevitably, a chapter length essay will be inadequate on many fronts. Holism has an ancient and complex history, and several strands of this history interweave with the history of medicine. The concept of holism is also relevant to the *philosophy* of medicine broadly construed. There are several aspects to holism, and each has different – if sometimes overlapping – significance for medicine. Implicit within different accounts of holism is the potential for a tension, if not an outright antagonism between reductionism and holism such that one is incompatible with the other, what Woods describes as "hard holism" (Woods 1998b). The normative implication of this tension is that medicine ought to be practiced under one model, holism, rather than the other; reductionism though the contrary claim is sometimes made. Supporters of holism usually do so because holism is regarded as more compatible with the humane practice of medicine which ought to be focused upon the most meaningful wholes, namely, the person or patient, or the community of persons which makes up society and whose interests form part of the telos of medicine (Pellegrino 1976).

Concerns about reductionism as a theoretical model usually focus upon what might be *reduced out*. By reducing a sick organism to its component parts medicine loses focus on the most important entity, the person to whose interests medicine is essentially directed. However this analysis is too naïve; presuming that holism and reductionism are mutually exclusive. What will emerge in this chapter is that many philosophers of medicine who are also doctors conclude that medicine requires both concepts and must therefore adopt a compatibilist stance.

The term "holism" is derived from the Greek term "Ολος-holos" meaning all, whole, entire, and total. A simple definition of holism is that "the whole is greater than the sum of its parts" (Smuts 1926, 98). In many ways this definition, taken out of the context in which the statesman-biologist Smuts used it, makes this too simplistic a claim. It does not alone capture the various and significant implications of holism for medicine. Nevertheless it cannot be ignored as a place-marker within the debates, since if it is contrasted with its antithesis "the whole is nothing more than the sum of its parts," then the basis of the essential debate is stated.

Phillips' (1976) analysis of holism within social science identifies several meanings and applications of the concept of holism, for example, as a critical concept used to emphasize the inadequacies of the reductionism of the natural sciences and holism as a dynamic emergent property of the relationships between the parts and the whole of an entity; holism, on this view, is necessary to understand the parts. Phillips shows that the holism/reductionism debate is relevant within social science but beyond that it is a debate that is germane across several domains. The debate is seen within philosophy, especially within metaphysics, philosophy of science and its various applications to the philosophy of biology, and the philosophy of medicine with regard to both scientific and normative aspects. The challenge of discussing holism is that it is almost impossible to give an account of holism without referring to reductionism because what is substantive about holism is most effectively described in contrast to reductionism.

Not a New Debate

The holism/reductionism debate is not new. Aristotle in the Metaphysics subscribes to a form of holism that rejected pre-Socratic atomists, thus establishing one of the possible earliest associations of holism as a response to, and rejection of, reductionism. However Aristotle qua biologist also contributes to the history of ideas on holism via his claim that, in biology, organic parts cannot be understood apart from the whole organism and, arguably, the environment within which the organism flourishes (Montalenti 1974). This further strand of thought within positive accounts of holism is similar to the point argued by Smuts (1926) that wholes are essential to understanding the parts of complex systems. Thus holism understood as a critical response to reductionism and holism as a theoretical approach which emphasizes the necessary connectedness of phenomenon can both be regarded as early, important, and continuing leitmotifs within the holism story.

Medicine's concern with healing, curing, and palliating has been characterized within the earliest accounts of medicine. These accounts include the idea that illness is deficit or imbalance in the components that allow the whole to function well. Such well functioning might arguably be regarded as synonymous with health and the still broader concept of flourishing. Placing health as *a* if not *the* central component of the good life is something which is clearly echoed in the much criticized but still extant World Health Organization's definition of health (WHO 1946).

Ancient medical interventions usually required whole lifestyle changes alongside any specific treatment in order to achieve the rebalance necessary for health. Similar ideas are evident in contemporary accounts of holistic medicine, something which is addressed in the section "Holism as Antireductionism." The section "Medicine's Moral Purpose" of this chapter addresses the normative accounts of holism as an approach to the practice of medicine with resonance for contemporary debates within medical ethics.

Science and Medicine

The history of medicine is also intimately connected with the history of science both in terms of developing theory and with the scientific technologies which were adapted or directly developed for medical purposes. Thus, medicine requires theories of causation coupled with techniques and technologies of enquiry such as magnification, amplification, and other means of observing the otherwise invisible aspects of the human organism. This combination of theory, technique, and technology resulted in a practical empirical medicine. Medicine's evolution therefore relied upon the reductionist, atomistic, and mechanistic approaches which characterized the successful evolution of Western science, a science that has contributed so significantly to medicine's success (Foss 1989). Therefore it is impossible to explore the meaning of holism for medicine without discussing the relationship between holism and reductionism.

Holism/Reductionism: Epistemology and Ontology

Epistemology is the branch of philosophy concerned with the nature and scope of knowledge, the *theory* of knowledge. The major questions addressed within epistemology, what is knowledge, how is it acquired, and how it relates to truth and belief, have been fundamental to philosophical enquiry for as long as human memory extends. Epistemology is not only concerned with *how* knowledge is gained but also with what it is possible to know and what there *is* to know. Therefore epistemology is also concerned with the stuff of knowledge or ontology, the theory of what there is. The preoccupation of Western philosophy with the nature of knowledge and truth has presumed that human perceptions are vulnerable to deception. This, in turn, has influenced the development of the scientific method. This method is essentially premised on the idea that a trained and disciplined system of enquiry is needed in order to uncover the truths about the world. Medicine has been one of the beneficiaries of this influential approach to the acquisition of knowledge, the disciplined application of "rational thought" or reason, leading humankind into the "light of knowledge."

Reason, applied through empirical enquiry, characterizes the development of science as is evident in the Renaissance, a period when pioneers such as Vesalius and Leonardo da Vinci challenged the accepted wisdom of Galen by performing human dissection (Quigley 2012). In doing so they demonstrated some of the fundamental errors in knowledge of human anatomy and helped to establish the, still vital to medicine, science of anatomy and physiology. Anatomical dissection is a very practical application of reductionism, with significant advantages for medicine. The challenging of received wisdom and the commitment to a careful scientific enquiry set the tone for how medicine progressed. Benefitting from the parallel developments in science and technology, medicine evolved through such landmark discoveries as the circulation of the blood, germ theory, human immunity, and the role of genes in disease. Of course this rather one-sided picture of the

evolution of medicine does not tell the whole story; there were certainly many blind allies as well as blind attitudes to overcome.

The symbiotic relationship between medicine and science, especially the natural sciences, is one of the contexts in which the holism or the holism-reductionism dyad is located. The fact that medicine, as an evolving discipline, sought out a sound knowledge base, finding that to lie essentially with the natural sciences, suggests that medicine, like these sciences, ought to be inclined toward reductionism. This provokes two key questions. The first is what does "reductionism" mean in this context? The second is what implications does such a commitment have for holism?

What Is Reductionism?

Reductionism has several meanings, but at its most general, it stands for the process of distilling complex entities into their simpler component parts. At one level this may be seen as a mere method of practical enquiry. As a method of enquiry, reductionism has been utilized by such familiar enlightenment thinkers as Descartes. Descartes's second precept in his *Discourse on Method* (1637) where he advises the reduction of complex propositions to their simpler components, characteristic of his system of analysis. Although Descartes is usually identified as a rationalist philosopher and not associated with the experimental methods of the British empiricists, he was also an empirical experimenter as can be seen in the *Optics* in the *Fifth Discourse*. Descartes's mechanistic and reductionist methods have been influential on the development of science especially his method of hypothesis and experiment with a view to establishing generalizable laws. Descartes was hopeful that his work might have benefits for medicine as he writes "The preservation of health has always been the principle end of my studies" (Descartes 1637, 275), and he hoped to conceive "a system of medicine which is founded on infallible demonstrations" (Descartes 1637, 17). A more detailed consideration of Descartes's work is beyond the scope of this chapter; however, one can detect a Cartesian echo in the work of much more radical reductionists such as the biologist Jacques Loeb of the Rockefeller Institute. Loeb's work on the physicochemical basis of biology was accompanied by a more radical normative program to rid science of teleo-mechanistic claims and to rid biology of the scourge of vitalism (Loeb 1912).

Thinking of reductionism as a practical method of scientific enquiry then it could be argued that reductionism is compatible with holism. Reductionism at this level might be seen as merely entailing a commitment to a certain kind of empirical method of epistemological enquiry. On this view complex wholes like human bodies, or their parts, hearts, eyes, and so on, can be understood by reducing them to their component parts. Reductionism understood in this way does not necessarily pose a threat to the commitment medicine has to the complex wholes such as patients. If the body can be understood under generalizable laws, then one body is much like another, and thus, the general efficiency of medical science is improved by employing general laws leading to protocols for the treatment of disease. In

principle a doctor can still direct his or her attention to the individual patient while understanding the patient's problems in terms of, or, *reduced* to, a chemical imbalance, a deranged physiology, or a genetic defect. This compatibilist view of reductionism with holism is expressed by Tauber:

> Medicine is, by its very character, holistic in orientation. Endeavouring to address all systems at once and to effect full function in each. This requires a global view of function from molecule to intact organism. However medicine is more than a science of an organic entity, and ultimately must be judged as how effectively it addresses the person, the individual with illness. (Tauber 2002, 270)

However it could equally be argued that medicine's commitment to natural science entails a far more radical form of reductionism such as that advocated by Loeb for whom medical science, if not a misnomer, should be reduced to the biochemistry of proteins (Loeb cited in Zucker 1981).

The Compatibility of Holism and Reductionism

A key question therefore is whether medicine's reliance upon the natural sciences also entails reductionism that is incompatible with holism. As van Riel and van Gulick (2014) in their discussion of scientific reductionism note, a general formula for reductionism can be given in the ontologically neutral form that:

> . . . if an entity x reduces to an entity y then y is in a sense prior to x, is more basic than x, is such that x fully depends upon it or is constituted by it. Saying that x reduces to y typically implies that x is nothing more than y or nothing over and above y.

Filling out the missing values in ways which imply a materialist or physicalist reductionism renders such a reductionism potentially incompatible with holism. So, in terms more allied to medicine, if reductionism, in which biology is reduced to microbiology and ultimately physics, or classical genetics is reduced to molecular genetics, results in a very lean ontology, then reductionism and holism are incompatible. An example of a reduction relevant to medicine is given by Ernest Nagel with regard to the phenomenon of headaches:

> [T]he detailed physical, chemical, and physiological conditions for the occurrence of headaches are ascertained . . . an explanation will have been found for the occurrence of headaches. (Nagel 1961, 366)

Nagel's example can be read as *ambiguously* lean in the sense that it could be read in agreement with Zucker's characterization of reduction in medicine which is:

> that all disease is physiology gone astray. Where there is truly no physiological problem, there is no disease. . . The ideal goal of reductionist medicine would be diagnostics (and treatment) accomplished by a biochemical-biophysical survey of the patient's body. Ideally

> psychological problems would be captured by this technique. It is part of the assumptions of reductionist medicine that, at the very least, mental states have clinically useful physical correlates. (Zucker 1981, 150)

A reductionist model of this kind has been applied at different times for different ends. Thomas Szasz in his *The Myth of Mental Illness* (1961) attacked psychiatry for its misuse of this physicalist reductive model. Szasz argued that the label "mental illness" is a misnomer because it does not meet the criteria for a disease and constitutes a mistaken application of the physicochemical model of disease. Szasz's attack on psychiatry notwithstanding the debate remains a live one with a powerful contingent within psychiatry convinced by the biological model of mental illness continuing the quest for genetic markers for disease and biomarkers for drug therapies. Psychiatrist Tim Cantopher's description of depression in his book aimed at lay readers puts it like this: "depressive illness is not a psychological or emotional state and is not a mental illness. It is not a form of madness. *It is a physical illness.* This is not a metaphor; it is a fact" (Cantopher 2012, 1). He goes on to describe the physical evidence of serotonin and noradrenaline deficiencies detectable in the cerebral spinal fluid of people who have depression as well as employing a range of reductive and mechanistic descriptions in his account of depression. Cantopher describes the limbic system of the brain as being like a "thermostat" and operating a "reverberating circuit" to be "found at the core of any complex machine." However, Cantopher is not a lean reductive physicalist; far from it, his emphasis on the biological aspects of depression is an attempt to avoid the denial of mental illness as *real*. Attempts to prove the biological basis of mental illness have had some partial successes thereby showing that Szasz's critique of psychiatry is not entirely vindicated. Nevertheless mental illness is still diagnosed by symptoms, and attempts to collapse the distinction between organic and functional mental illnesses have not been successful.

The compatibility of reductionism with holism is very much dependent upon the status of the reductionism adopted. Reductive approaches might be further divided between those that have an ontologically *descriptive* agenda and those that have an ontologically *revisionary* agenda. For the former the fact that a migraine is reducible to certain brain states does not question the reality of the migraine. However for those who are committed to a more radical revisionary ontology, then reduction is equivalent to *elimination*. On this view, if there is no physical correlate, then there is no condition, as Szasz argued regarding the status of mental illness. Whereas the first approach might be broadly compatible with the practice of medicine, the second approach might be regarded as seriously problematic for medicine since certain kinds of "whole" such as "health," "community," "person," and "pain" run the risk of elimination by reduction.

Holism: Some Practical Puzzles

The history of the understanding of infection provides an interesting point for reflection about the place of holism in medicine. It took a long time, and attacks on several fronts, before doctors were ready to accept that the connection between

their behaviors (moving from postmortem dissection to the maternity ward) was the cause of the frequently deadly puerperal fever. The case of Ignaz Semmelweiss and his efforts to persuade his profession of the need to change their behavior in order to limit the risk of puerperal fever is an interesting example from which to explore one dimension of the relationship between holism and reductionism. One could say that Semmelweiss took a holistic approach to the problem of puerperal fever because he witnessed a complex whole in which many different factors came together to create the risk. Even though he was unable, prior to germ theory, to demonstrate that *Streptococcus pyogenes* was the causative agent of the fever, he was able to bring systematic observations to bear on the context and identify the complex of factors in a holistic relationship. Of course, although the pragmatic measures he recommended proved effective, the relevant knowledge was not complete until germ theory was able to demonstrate the role of pathogenic bacteria in infection. Thus, extrapolating from this seeming triumph of reductionism, one might argue that medicine *requires* a reductionist approach. The identification of *s. pyogenes* as the *true* cause of the problem *reduced* a complex phenomenon, puerperal fever, to the presence of *s. pyogenes*.

A related example explores the point further. A diabetic patient attends her family physician with a painful sore throat and fever. The physician performs a rapid strep test and confirms the presence of group A streptococci. The physician prescribes a course of antibiotics and the patient's sore throat is treated. Is this an example of reductionism in action? On the basis that reductionism is the diminution of a phenomenon to its component parts, then the second example can be seen as an example of reductionist medicine because the complex phenomenon of the presenting patient, fever, painful throat, swollen lymph nodes, and so on, can be reduced to the presence of a pathogen in the throat. However what is really going on here? To address this question it is useful to distinguish between *ontological* and *epistemological* reductionism. Ontological reductionism is the view that complex entities are mere collections of simpler more fundamental entities. Epistemological reductionism is the view that, for the particular purposes of a specific knowledge enquiry, a complex phenomenon can be divided into its component parts. There is no doubt that medicine has evolved in close proximity to the development of the natural sciences. Medicine is not only substantially grounded upon the same theoretical premises of the natural sciences but utilizes the experimental methodologies and applied technologies from the natural sciences, for example, as demonstrated by medicine's fundamental reliance upon molecular biology and physics. This may be taken as grounds for medicine to align itself with the ontological reductionism adopted by some within the natural sciences. This step might be reasonable if medicine were only science but as Pellegrino puts it "medicine qua medicine transcends science" (1926). In other words, medicine might *use* science within its methods, but it is not itself a science.

For medicine to adopt ontological reductionism is therefore regarded as an error by several commentators. Pellegrino (1976) makes the point that medicine is knowledge applied for human ends, what Thomasma describes as *technê iatrikê*, a technical art (1990: 248). Both see medicine as standing between the sciences and

the humanities and thus requiring a holistic approach that is compatible with a scientifically informed practice. The mere presence of *S. pyogenes* does not explain the problem of cross infection within the maternity ward, since many people are entirely healthy carriers and do not exhibit the pathology of an infected person. Nor does the presence of *S. pyogenes* determine the strategies for managing the problem of cross infection. To understand the problem of cross infection, and how to deal with it, requires an insight into the whole system which sustains the process. This system is not reducible to the causative bacterium but also includes the culture of medicine including the beliefs that gentleman doctors are too socially superior to be considered "dirty." In the second example the strategy of using a form of epistemological reductionism to establish the particular cause of the presenting phenomenon does help to determine whether and which antibiotics to prescribe. The point is that medicine must be committed to strategies of *epistemological* reductionism because such knowledge is necessary for medicine to achieve the important goal of disease prevention and cure. As Zucker puts it reductionism is an "epistemological hypothesis" (1981, 149) for medicine but does not entail that medicine *is* premised upon reductive physicalism. It would be a rather odd thing to ascribe such a hard ontological commitment to something that is fundamentally a *practice* and which draws upon the epistemological utility of multiple disciplines. To put this differently, one might say that ascribing a reductive physicalist ontology to medicine is something akin to a category mistake.

Holism as Antireductionism

As has been indicated earlier, the relationship between holism and reductionism is an ancient one and one often characterized as a dispute between the two perspectives. If the dispute is considered within the context of *epistemology*, then the dispute is about rival accounts of how certain phenomenon within the world can be known. In the context of *ontology*, then it is a dispute between rival accounts of what there *is*. The holism/reductionism dispute takes on particular force in the early part of the twentieth century across a number of scientific and philosophical domains. Smuts's frequently quoted slogan that "the whole is greater than the sum of its parts" (Smuts 1926, 68) arose in the context of a debate within the philosophy of biology. Smuts and others wished to make an ontological claim that reductive accounts of biology could not account for the phenomenon that were the emergent properties of complex biological systems or wholes. There were several contexts in which the dispute was played out. One context was the nature of life itself. Scientists like Loeb (1912) proposed a biochemical biology that was ontologically reductive. Loeb's account of biology was not only used to progress the science but as an ideological weapon against *vitalism*, the belief that some special properties of "life" distinguished animate from inanimate things. Another was in the field of genetics in which Darwinian theory was being applied by scientists such as Galton whose attempt to reduce the complex human traits of character or intelligence to a form of simple genetic determinism became the basis of eugenics.

In the newly evolving discipline of psychology, behaviorism was an attempt to rid psychology of "mentalism" and eliminate the mental altogether.

From Epistemology to Ethics

The aim of the reducers was *eliminative* reductionism, to rid science of vitalism and other messy concepts, in the way that chemistry had rid itself of the phlogiston theory of combustion through systematic quantitative experiment. The equivalent debates within biology had particular resonance for medicine because of its reliance upon biological science as one of its foundational disciplines. However Brandon (1996) moves to dismiss both reductionism and holism as equally flawed theoretical perspectives; he notes in passing the ideological commitment of, at least some, within medicine to reductionism, quoting Henderson's introduction to a seminal text on experimental medicine: "man is by nature proud and inclined to metaphysics, but the practice of experimentation will cure these faults" (Henderson (1927) cited in Brandon 1996, 204). Just exactly what these "faults" included was and still is a concern of medicine as George Engel notes "How physicians approach patients and the problems they present is very much influenced by the conceptual models around which their knowledge and experience are organised" (1981, 101). Engel, in a series of papers from the early 1960s, had been working to reconcile medicine's dual commitment to being scientifically grounded and to resolving the problems of the patient or person. Engel noted that the tendency toward scientific reductionism had created a "biomedical model" as the modus operandi for medicine, but this resulted in an approach which tended to ignore the human component. Engel regards this as a "crippling flaw" of the model (1981, 103). Engel's response was to advocate an alternative model, the "biopsychosocial" model, drawing upon earlier systems theorists such as Ludwig von Bertalanffy (circa 1901) (Borrel-Carrió et al. 2004). It is beyond the scope of this chapter to give a fuller account of systems theory other than to say that it is premised upon the claim that within nature there are hierarchical relationships between its many component units. Simpler units are subordinate to more complex units within the hierarchy, and this hierarchy represents a dynamic whole. A molecule like water, to take one example, can be seen as nested within the higher-order complex that is the biosphere through a succession of hierarchical wholes. To use Engle's example, a man who has suffered a heart attack can be seen at a number of successive and interconnected levels, the cells that have been damaged by ischemia can be identified and characterized as *cells,* the cells are also part of the myocardium (tissues). This tissue forms an organ, the heart, and the heart is within the person; the person is part of a family, a member of a workforce, a part of a community, and so on. Engel comments:

> In scientific work the investigator is generally obliged to select one system level upon which to concentrate, or at least begin, his efforts. For the physician that system level is always a *person*, i.e. a patient. (1981, 106)

The same theme is taken up by Tauber (2002) who remarks on the challenge of defining holism except in contrast or opposite to the prevailing reductionism of the era, and he agrees with Rosenberg that the more one scrutinizes holism, "the more elusive it becomes, the more it dissolves and reconfigures itself into its opposite" (Rosenberg 1998, 348). The only positive definition that Tauber will commit himself to is "the general rubric of 'considering the patient as a person'" (Tauber 2002, 263). In this one modest phrase, Tauber captures the true import of the reductionism/holism dyad for medicine that it is not a debate about epistemology, ontology, or scientific method but rather a debate about the moral focus of medicine and that reductionism *eliminates* the moral. Holism in this sense is really a claim about patient or person centeredness. The point is made equally well by Engel who states:

> For medicine in particular, this neglect of the whole inherent in the reductionism of the biomedical model is largely responsible for the physician's preoccupation with the body and disease and corresponding neglect of the patient as a person. The widespread public feeling that scientific medicine is impersonal is consistent with how the biomedically-trained physician views the place of science in his everyday work. For him 'science' and the scientific method have to do with the understanding and treatment of disease, not with the patient and patient care. (Engel 1981, 107)

The theoretical obsession by scientists, especially biologists, with the reductionism/holism debate in the early decades of the twentieth century also had an influence on other areas of thought. Two are worth a brief mention because they enshrine a harder stance on holism, and they include developments within nursing theory and the genesis of the holistic medicine movement.

Arguably Florence Nightingale was the first systematic theorist of nursing; her *Notes on Nursing: What it is, What is not* (1860) and its emphasis on the environment of the patient certainly have a holistic tone. Subsequently a number of pioneering nurses took the theoretical ground of nursing one step further. Martha Rogers's Science of Unitary Human Beings (1970) continued the emphasis on the patient within the environment as well as the irreducibility of the human being. Hildegard Peplau (1952), who was strongly influenced by her work with the psychiatrist Erich Fromm, emphasized the interpersonal relationship between the nurse and the patient as the basis of the caring relationship. A later generation of nurse theorists emphasized the nature of nursing as a combination of science and art. Jean Watson (2011) developed the concept of nursing as a "caring science" with philosophical *value*-based foundations rather than reductive scientific ones. The idea that nursing, and implicitly health work, generally is best delivered through value-based practices of care has become a central pillar for many nurse theorists. Common threads through these theories are the holism of the patient and the holistic nature of the goals aimed at in health care including health, self-care, rehabilitation, and so on (Woods 1998a, b). Many of these nurse theorists had concerns in common with Engel and Tauber that medicine was failing patients because of the reductionism within the "medical model." In the UK one of the prime movers behind the hospice movement, Cicely Saunders, was working and writing as a nurse when she

wrote some of her seminal accounts of the values of hospice care with their emphasis on treating the whole of the patient (Woods 2007). Along with many others, Saunders was horrified that mainstream health care was failing dying patients, and she argued that dying patients needed to be taken out of hospitals in order to let values back in. In her later work as a palliative care physician, Saunders expressed the importance of scientific underpinnings of medicine. After all it was the scientific understanding and application of analgesia that enabled adequate pain relief to be consistently and effectively applied. However the science must be embedded within care, and her views can certainly be read as compatibilist between the reductive accounts of pain and pain relief and her holistic concept of "total pain" (Saunders 2005). She comments "Above all, my experience emphasises that the practice of medicine includes more than specific treatments" (Saunders 2005, 35), a point echoed by Kearney who claims that palliative physicians cannot be mere "symptomatologists" (Kearney 1992; Woods 2007).

Holistic Medicine

Perhaps the most radical antireductionist program, also a product of the disquiet directed at scientific medicine, is the holistic medicine (health) movement. There is no single source of inspiration for this movement; it came from within and from outside of medicine and certainly coincides with the popular cultural movements of the 1960s America and elsewhere which brought much of the establishment under critical scrutiny. One important influence here was the work of the social commentator and critic Ivan Illich whose book *Limits to medicine: Medical nemesis* (1974) included one of the earliest uses of the term "medicalization" – the epitome of medical control of people through the conversion of social and behavioral phenomenon into diagnosed conditions. Illich's line of attack was similar to those of other critics such as Foucault (1965) and Szasz (1961), but in addition Illich also took up the topic of iatrogenesis. Illich observed that medical interventions specifically were iatrogenic, the *cause* of illness. The thrust of his arguments was to show that medicine was concerned with its own power and autonomy and was failing the very people it was supposed to serve. These themes became the battle cry for many people with similar causes within the holistic health movement.

Holistic medicine (or health) is a broad banner under which "practitioners" of various healing arts, outside of mainstream medicine, come together to seek legitimacy. Across the spectrum are a variety of practices and treatment modalities, and some are more or less compatible with traditional medicine and some diametrically opposed. These include meditation, biofeedback, reflexology, and acupuncture (Koppelman and Moskop 1981). As Carlson, in an early commentary, notes that the evangelism for holistic health became a political movement in an attempt to rival the powers that had accrued to mainstream medicine (Carlson 1979). As a political force, health activism proved potent in getting patient's concerns onto the agenda. HIV/AIDS activism, women's health, and cancer movements set a very

evident trend that influenced both policy and practice. This activism was not solely directed toward the rejection of reductionism but was centrally concerned with the way the biomedical model disregarded the person and left the patient without a voice.

A full account of the holistic medicine/health movement is not possible in this chapter; however, a distinction that is significant for this chapter is that between the advocates of *complementary* and *alternative* medicine. Alternative medicine, as the term implies, advocates practices which claim to have the healing effects of medicine but are not based on evidence derived from scientific methods. These practices are not part of biomedicine and may be contradicted or even contraindicated by scientific evidence. Some practices may be biologically plausible such as naturopathy and chiropractic, yet others have no biological basis and draw upon the ancient folk practices of different cultures such as traditional Chinese medicine and Ayurvedic medicine.

Alternative medicine has been criticized by scientists and conventional medical practitioners as being quackery, fraudulent, and potentially dangerous. However many people turn to alternative medicine because of perceived inadequacies in what medicine provides. This has led some to advocate the use of complementary medicines in which alternative practices are utilized alongside conventional therapies. Complementary medicine is perhaps significantly evident within cancer and palliative care contexts. However mainstream medicine is often very hostile toward it, cutting across the distinction between "alternative" and "complementary" in order to condemn holistic medicine as anti-mainstream medicine. An example of this antipathy came to light following a Lancet publication of a study conducted by UK cancer charities. The paper reported interim results of a study of the treatment of women with breast cancer at the Bristol Cancer Help Centre, a renowned center for complementary therapies. The study allegedly found that the Bristol women were twice as likely to die when compared with women in conventional medical centers (Chilvers et al. 1990). However it was later revealed that the study was profoundly flawed and had made an invalid comparison of the Bristol patients with women who had less advanced disease. The damage was caused by an untimely press release of the interim report which made strong critical claims beyond the more cautious conclusions of the Lancet paper (Goodare 2007). The study was eventually discredited but not before international news headlines had condemned the practices of the Centre. A journalist Liz Hunt (1993) wrote about one woman's experience:

> . . .she sought help from Bristol because she needed more than the 'cut, burn or poison' approach of the NHS, which relied on surgery, radiotherapy or toxic drugs. The centre offered stress counselling, relaxation and visualisation techniques, healing and special diets. Suddenly, the women were being told that these therapies - these 'gentle therapies' - were actually harming them. When, two months later, the researchers admitted that their study was seriously flawed in several respects, and that the Bristol women in the study had more advanced disease than the Marsden women from the start, it provided little consolation - and received minimal publicity. The damage had been done; hundreds of cancer patients had been put through hell, and the centre's reputation was destroyed. (Hunt 1993 online article)

The place of holism, and the reductionism/holism debate, within medicine can now be seen to have evolved from the adoption of a particular practical scientific method, to a dispute within the philosophy of science, to an ideological dispute about goals and values. This chapter has demonstrated how critics of various kinds have argued that ontological reductionism is a too narrow philosophy for medicine because it too readily reduces out the very things to which medicine as a practice is directed toward: the health and care of sick persons. Holism that is posited as *anti*reductionism may also be regarded as problematic if it does not acknowledge the importance of epistemological reductionism as a method to be used within medical practice. These comments mark the transition to the final section of this chapter which explores the implications of the claim that, though grounded in science, medicine cannot be reduced to science and still fulfill its normative moral function.

Medicine's Moral Purpose

The final strand of this chapter will deal with the discussion about what is perhaps the worst indictment of medicine's temptation toward a reductionist model, the failure to regard patients as people. Pellegrino says "Medicine is, in short, a practical theory of human reality. It is a moral activity, since it operates through the interrelationship of persons in which physician and patient are coparticipants in defining the goal and achieving that goal – cure of illness or promotion of health" (1976, 15). Implicit within Pellegrino's account is yet another form of holism, a kind of *phenomenological* holism, which requires that the medical gaze toward the patient is not one of eliminative reductionism but keeps in focus the moral entity of the person. Medicine's repeated failure to achieve this has been recognized and has sometimes shocked the profession into a state of critical reflexivity that produced reforms to practice. The worst atrocities of the Nazi experiments are regarded as a graphic illustration of how far the wrongheadedness of zealotry could distort medical purposes. However it was perhaps the extremism of these events that prevented medicine from taking the lessons to heart. Beecher (1966) and Pappworth's (1967) exposure of unethical practices in medical research pointed out a problem within the routine core of medicine. Their exposé claimed that the application of scientific methodology to human subjects' research was neither good science nor moral conduct, reducing people to the objects of research. Though the ethics of medical research has become a major bioethics industry with a proliferation of international codes, professional guidance, and mechanisms for ethical scrutiny, there is still an ongoing discussion about a fundamental problem. The problem is that quantitative clinical research treats patients as objects. The concerns that were noted by Beecher and Pappworth are still evident. Faulder's *Whose body is it anyway?* (1985) is but one illustration of how the issues continue to be debated in relation to informed consent, placebo controlled studies, and a lack of effective engagement with the patients and publics for whom the research is intended to benefit.

So the patient as a *person* emerges as one of the important "wholes" to which medicine ought to be directed. The failure of medicine to maintain the patient-centered gaze and reduce the patient to an object for research or a mere pathology has been consistently criticized. Paul Ramsey's book *The Patient as Person* (1970) was one of the landmark contributions to the bioethical turn that showed a renewed interest by philosophers and theologians in the practical ethical problems facing medicine. There are bioethics innovators too numerous to mention who took up this mantle in the 1960s and 1970s, but the work of Beauchamp and Childress stands out for the abiding influence their book *The Principles of Biomedical Ethics* (1979) has had, and is having, on medical ethics as probably the most popular textbook on medical ethics used within the profession. However some commentators have brought the holism/reductionism debates to moral constructs such as autonomy (Clouser and Gert 1999; Tauber 2003).

Introducing the challenges of medical ethics may seem to be straying away from the discussion of holism and reductionism with which this chapter began; however, a brief excursion into medical ethics is relevant because there is a normative aspect to the reductionism/holism dyad which should now be clear. Whether medicine should adopt scientific reductionism becomes a normative question because it has implications for the very raison d'être of medicine, what medicine is about, and how it ought to be practiced. This point harkens back to Pellegrino's observation that "medicine has a telos which distinguishes it from its component sciences" (1976, 15) and to Saunder's concerns that medicine loses its moral compass when faced with the seeming medical "failure" of the dying patient. To see the patient as a diagnosis only (Tauber 2002) or to reduce the person with cancer to the collection of cancer cells in her larynx is, in the views of those critical of reductionism, to stray profoundly and dangerously away from the goals of medicine. Brandon's point that the reductionism/holism debate is now rendered redundant (Brandon 1996) is an important one especially as some of the archetypes of reductive sciences like biology and genetics have come around to recognizing the significance of holistic approaches. This is not to say that the old holism/reductionism dyad should be abandoned as the issues addressed in this chapter suggest; the dyad can be a useful critical framework with which to reflect upon the ongoing normative challenges within medicine.

Definition of Key Terms

Holism	A theory that parts of a whole are so interconnected that they cannot exist independently of the whole or cannot be understood without reference to the whole. The whole is regarded as greater than the sum of its parts. Holism has been applied to science especially biology, philosophy of mind, and language.
Reductionism	Can be defined as a process of distilling complex entities into their simpler component parts. At one level this is an epistemological or methodological approach. Ontological reductionism is the more radical view that a whole is nothing more than the sum of its parts.

Summary Points

- Holism is relevant to the philosophy of medicine because the goals of medicine are directed toward important holistic ends.
- Holism can be defined positively as a nonreductive property of complex systems.
- Most accounts of holism deal with the relationship between holism and reductionism.
- Holism can also be understood as an antireductionist strategy.
- Reductionism can be understood as compatible with holism.
- Reductionism ought to be adopted by medicine because medicine is dependent on sciences where reductionism has proven to be an effective methodology (according to some commentators).
- Reductive methods need not distract attention from the moral purpose of medicine (according to some commentators).

References

Beauchamp TL, Childress JF (1979) The principles of biomedical ethics. Oxford University Press, Oxford

Beecher HK (1966) Ethics and clinical research. N Engl J Med 274(24):1354–1360

Borrel-Carrió F, Suchman AL, Epstein RM (2004) The biopsychosocial model 25 years later: principles, practice, and scientific enquiry. Ann Fam Med 2(6):576–582

Brandon RN (1996) Concepts and methods in evolutionary biology. Cambridge University Press, Cambridge

Cantopher T (2012) Depressive illness: the curse of the strong, 3rd edn. Sheldon Press, London

Carlson RJ (1979) Holism and reductionism as perspectives in medicine and patient care, In Orthodox medicine, humanistic medicine and holistic health care – a forum. West J Med 131:466–470

Chilvers CED, Bagenal FS, Easton DF, Harris E, McElwain TJ (1990) Survival of patients with breast cancer attending Bristol Cancer Help Centre. The Lancet 336(8715):606–610

Clouser KD, Gert B (1999) A critique of principlism. J Med Philos 15(2):219–236

Descartes R (1637/1991) The philosophical writings of descartes, vol III (trans: Cottingham J, Stoothoff R, Kenny A, Murdoch D). Cambridge University Press, Cambridge

Engel GL (1981) The clinical application of the biopsychosocial model. J Med Philos 6:101–123

Faulder C (1985) Whose body is it?: the troubling issue of informed consent. Virago, London

Foss L (1989) The challenges to biomedicine: a foundations perspective. J Med Philos 14:165–191

Foucault M (1965) The birth of the clinic: an archaeology of medical perception. Vintage Books, New York

Goodare H (2007) The media and cancer survival. J R Soc Med 100(10):483–484

Hunt L (1993) News of death greatly exaggerated: the report on the Bristol Cancer Help Centre was damning. It was also highly inaccurate. These ex-patients want the world to be told, and they are tired of waiting. The Independent Newspaper. Monday, 15 Nov 1993. http://www.independent.co.uk/life-style/news-of-death-greatly-exaggerated-the-report-on-the-bristol-cancer-help-centre-was-damning-it-was-also-highly-inaccurate-these-expatients-want-the-world-to-be-told-and-they-are-tired-of-waiting-liz-hunt-reports-1504475.html#. Accessed 8 Aug 2015

Illich I (1974) Medical nemesis. Marion Boyars, New York

Kearney M (1992) Palliative medicine – just another speciality? Palliat Med 6:39–46

Koppelman L, Moskop J (1981) The holistic health movement: a survey and critique. J Med Philos 6:209–235

Loeb J (1912) Mechanistic conception of life. The University of Chicago Press, Chicago
Loewy EEH, Loewy Springer R. (2006) Textbook of healthcare ethics. 2d ed. Dordrecht: Kluwer Academic Publishers
Montalenti G (1974) Chapter 2, From Aristotle to Democritus via Darwin: a short survey of a long historical and logical journey. In: Ayala FJ, Dobzhansky TG (eds) Studies in the philosophy of biology: reduction and related problems. University of California Press, Berkley, pp 3–20
Nagel E (1961) The structure of science. Problems in the logic of explanation. Harcourt, Brace & World, New York
Nightingale F (1860/2006) Notes on nursing what it is, and what it is not. Hard Press, London
Pappworth MH (1967) Human guinea pigs: experimentation on man. Routledge and Kegan Paul, London
Pellegrino ED (1976) Philosophy of medicine: problematic and potential. J Med Philos 1(1):5–31
Peplau H (1952) Interpersonal relations in nursing: a conceptual frame of reference for psychodynamic nursing. Putnam, New York
Phillips DC (1976) Holistic thought in the social sciences. California University Press, Stanford
Quigley C (2012) Dissection on display: cadavers, anatomists and public spectacle. McFarland & Co, Jefferson
Ramsey P (1970) The patient as person: explorations in medical ethics. Yale University Press, New Haven
Rogers ME (1970) An introduction to the theoretical basis of nursing. Davis Company, Philadelphia
Rosenberg C (1998) Holism in twentieth-century medicine. In: Lawrence C, Weisz G (eds) Greater than the parts: holism in biomedicine 1920–1950. Oxford University Press, Oxford, pp 335–355
Saunders C (2005) Watch with me: inspiration for a life in hospice care. Observatory Publications, Lancaster
Smuts JC (1926) Holism and evolution. Macmillan, New York
Szasz TS (1961/1974) The myth of mental illness: foundations of a theory of personal conduct. Harper & Row, New York
Tauber QI (2002) The ethical imperative of holism in medicine. In: Van Regenmortel MHV, Hull DI (eds) Promises and limits of reductionism in biomedical sciences. Wiley, West Sussex, pp 261–277
Tauber AI (2003) Sick autonomy. Perspect Biol Med 46(4):484–495
Thomasma DC (1990) Establishing the moral basis of medicine: Edmund d. Pellegrino's philosophy of medicine. J Med Philos 15:245–267
Van Riel R, Van Gulick R (2014) Scientific reduction. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, Summer 2014 edn. http://plato.stanford.edu/archives/sum2014/entries/scientific-reduction/
Watson A (2011) Nursing: human science and human care, 2nd edn. Jones & Bartlett, New York
Woods S (1998a) Holism in nursing. In: Edwards SD (ed) Philosophical issues in nursing. Macmillan Press, London
Woods S (1998b) A theory of holism for nursing. Med Health Care Philos 1(3):255–261
Woods S (2007) Death's dominion: ethics at the end of life. Facing death. Open University Press
World Health Organization (1946/1958) Preamble to the constitution, adopted by the International Health Conference. New York 19 June to 22 July 1946. In: The first ten years of the World Health Organization. WHO, Geneva
Zucker A (1981) Holism and reductionism: a view from genetics. J Med Philos 6:145–163

Hope, Despair, and Other Strategies of Patients

28

Kenneth Boyd

Contents

Abstract

Is hope an emotion or a virtue? Is a patient's hope measurable? Philosophers and healthcare researcher have differed in their approaches and answers to such questions, but some areas of greater clarity and convergence seem to be emerging. The dynamics of hope, despair, and hopelessness have been more clearly delineated. Hope is neither simply an emotion nor a propositional attitude. It may be understood by patients either as determined by the doctor or in the power of the patient. The clinical and therapeutic relevance of hope and despair depend less on measuring hope in patients than on listening to the individual patient.

Introduction

"In the treatment of nervous cases, he is the best physician who is the most ingenious inspirer of Hope" remarked the English poet, philosopher, and professional patient Samuel Taylor Coleridge in January 1833 (Woodring 1990). While Coleridge was

K. Boyd (✉)
Biomedical Teaching Organisation, Edinburgh University Medical School, Edinburgh, Scotland, UK
e-mail: k.boyd@ed.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_25

speaking from personal experience of both his own and at that time his daughter Sara's partly psychosomatic and often opium-assisted illnesses, he was also expressing a long and widely held view about the therapeutic relevance of hope to illness generally. But what exactly is "hope" and what more precisely can be said about its therapeutic role?

Hope (together with related concepts such as fear, despair, and hopelessness) has been the subject of at least two kinds of scholarly discourse, not always apparently aware of each other. Firstly, from antiquity to the present, there is that of classical, modern, analytic, and other philosophers inquiring, for example, whether conceptually hope is an emotion or a virtue. Second, mostly from the mid-twentieth century onward, there is that of healthcare practitioners and researchers inquiring, for example, whether hope can be described and even measured psychologically.

This chapter will discuss what can be learned about hope, despair, and hopelessness from the diverse, but in some respects converging, views of philosophers and of healthcare professionals and researchers.

Philosophers

Is hope an emotion or a virtue? How this philosophical question is answered could have practical and ethical implications for the relationship between patients and health professionals. If hope is a therapeutically beneficial emotion, might deception to encourage it be justified? If a virtue, might moral exhortation be justified? The question of whether hope is an emotion or a virtue – or perhaps something else – has been answered in a variety of different ways by some of the most influential philosophers in the Western tradition.

Hope was not among the virtues specifically identified by *Aristotle* (384–322 BCE), the Ancient Greek philosopher most explicitly concerned with them. For Aristotle, as for many other classical philosophers, hope was an emotionally colored opinion about some future state of affairs: whether any particular hope was good or bad depended either on whether the person holding it was of good or bad character or on whether the elements of desire or fear mixed in with it seemed, to the wise and prudent, excessive or deficient.

In the European Middle Ages, the Aristotelian theologian and philosopher *Thomas Aquinas* (1226–1274) altered and expanded the classical understanding of hope. As an emotion, hope attracted you toward what you conceived as a future good that was "possible but difficult" to obtain (just as fear repelled you from what you conceived as a future evil). But in determining to overcome difficulty, hope could also become a virtue or settled disposition rooted in reasoned reflection and sustained by moral effort. And beyond this again, hope could become, but by God's grace alone, one (with faith and charity) of the three "theological" virtues, through the practice of which God was most deeply known.

Theological virtues however were of little interest to the radically innovative early modern philosophers of the seventeenth and eighteenth centuries: they reinterpreted the classical view of hope as an emotion in what they considered more scientific terms.

Benedict Spinoza (1632–1677), for example, considered the emotions to be aspects of nature's necessary causal system: as such, they could not be controlled by the rational human mind, but they could be understood by it, thereby reducing the intensity of the pleasures or pains they inflicted and leaving more room to be guided by reason. As emotions therefore, hope and fear, necessarily tied to the pain of doubt and uncertainty, could not be good in themselves – although in a socially useful fashion, they might moderate the conduct of most people, who in practice seldom were guided by reason.

David Hume (1711–1776) was less sanguine about reason as a guide. Reason, he claimed, not only is but ought to be the "slave" of the emotions: reason may work out the means, but our ends are determined by our emotions. For Hume as for Spinoza however, the emotions are ultimately related to the prospect of pleasure or pain. The emotions of hope and fear each represent differing mixtures of the emotions of pleasure-related joy and pain-related grief: hope rises as joy ascends over grief because some future good begins to seem probable; hope is eclipsed by joy when the good seems certain. (Fear rises with the probability of and grief over future evil and is eclipsed by grief when evil seems certain.)

Reason itself however was to be reinterpreted, and the moral significance of hope revived, by *Immanuel Kant* (1724–1804). Kant agreed that humans are subject to nature's necessary causal system and hence proper objects of scientific study: but these objects (including scientists) also experience themselves as subjects and as moral agents, benefited and burdened with the freedom and responsibility to envisage and prepare for future contingencies. To be human therefore was to ask not only scientifically "What Can I know?" but also morally "What ought I to do?" and (if I do what I ought) "What may I hope?" Kant's answer to this latter question was that in obeying the rational moral law (a version of the golden rule – "do not do unto others as you would not have them do unto you"), it was reasonable for individuals and societies to hope that they would be vindicated, not necessarily in the foreseeable future, but ultimately. "Ultimately," for Kant, was a claim based not on (scientific) knowledge but on experience interpreted by rational (not necessarily religious) faith. Hope therefore was more than a mere emotion and while not in Kant's terminology a virtue, a reasonable implication of one's duty to obey the moral law.

Against the background of this Western tradition, a number of contemporary writers have examined various philosophical aspects of hope.

The question of whether or not hope is an emotion has been examined in some detail by the analytic philosopher *J. P. Day* (1969, 1998). An objection to hope being an emotion, he argues, is that since hope involves not just desiring something but also estimating its probability (you may desire, but cannot hope for what you estimate to be impossible), this estimating aspect is cognitive and cannot be an emotion. The desiring aspect however also falls short of being an emotion, since it lacks two of the three characteristics of an emotion – a characteristic sensation and a characteristic physical symptom. Desire may possess the third – a characteristic behavior pattern, in the case of desire a tendency to try to bring about what is desired: but that characteristic is absent or irrelevant if what is desired is something (such as a

fine day tomorrow) that the behavior of the person desiring it, individually or with others, is unable to influence. In terms neither of its estimating nor of its desiring aspect therefore can hope be an emotion.

A more accurate way of understanding hope, Day argues, is as a "propositional attitude" – an attitude taken toward a proposition or statement about some future possibility. In the case of hope, for example, the attitude is positive, whereas in the case of fear, it is negative. A significant difference between hope and fear in this respect however, he continues, is that while both hope and fear are propositional attitudes, fear, unlike hope, is also an emotion. You may either hope or fear that tomorrow will be sunny (propositional attitude), but while you may fear a tiger (emotion), you do not hope it.

The definition of hope as a propositional attitude however leaves other philosophers unsatisfied that restricting hope to "hope that" does sufficient justice to the phenomenon of hope. The theologian *J. Macquarrie* (1978), for example, prefers to regard hope simply as an attitude, a stance, or a direction of the whole person's inseparable aspects of thinking, feeling, and willing. Its thinking aspect involves not only the creative imagination envisaging new possibilities but also the critical intellect assessing their feasibility. Its feeling or emotional aspect may lack any obvious characteristic physical symptoms, but as an aspect of the whole person, hope nevertheless may be felt as "bracing," while someone in whom it is acutely disappointed may feel "deflated." And the willing or volitional aspect of hope implies that the future can be influenced by the exercise of human freedom in moral choice and action.

Another writer on the subject, the philosopher *J. M. Waterworth* (2004), accepts Day's conclusion that hope is not an emotion. She then goes on to argue, however, that not all hopes take the form of hoping, for example, that tomorrow will be sunny or for a sunny day tomorrow. Such hopes, which she terms "direct hopes", certainly are or are akin to propositional attitudes in that what is hoped for can be described, if only in the most general terms (such as hoping "for a better world"). But there also exist what Waterworth terms "indirect hopes." These are more difficult to capture in propositions about the future, since they are not so much descriptive of future states of affairs as expressive of the values of the person who hopes, for example, to endure against the odds or to retain human dignity in outwardly degrading circumstances. All hope of its nature involves uncertainty about the future, but whereas the estimative aspect of direct hopes is concerned with degrees of probability, indirect hopes refer to what may just be possible, however unlikely. Waterworth distinguishes direct hope as "living in hope" from indirect hope as "living in the light of hope" when the "unreflective trust" that normally sustains direct hopes is no longer available.

Indirect hope, Waterworth argues, is particularly relevant to the experience of illness and suffering and the prospect of dying. In these contexts direct hopes, with their estimation of probability, clearly have a place – hope, for example, that a particular medical treatment will relieve pain, alleviate symptoms, or even cure. In this context, the estimative aspect of a patient's hope may largely have to rely on the degree of their trust in the doctor, and this may vary depending on the patient's

evaluation of the doctor's authority. A doctor's authority, Waterworth comments, may stem from the doctor's power, or from their knowledge and skill, or from a combination of the latter with professional and personal qualities, such as attentiveness to the patient as an individual, which attract the patient's respect and trust. If the doctor's authority stems simply from their power vis-à-vis the patient (including a general assumption about their medical knowledge and skill), the doctor's reassurances ("so that the patient may not lose hope") may put the patient at risk of entertaining direct hopes that eventually will be disappointed.

In the absence of indirect hope, Waterworth observes, the disappointment of direct hopes can lead to despair and at worst hopelessness. Despair can be distinguished from hopelessness by the fact that to despair is to do something, whereas hopelessness is to be in a passive state. Despairing is losing hope in something that is still in some sense valued, and psychic energy is still being used up in the process. In hopelessness all passion is spent and there is emotional indifference to what was once valued.

In circumstances where direct hope is disappointed, Waterworth argues, the importance of indirect hope is that its objectives concern what, for the patient, gives meaning in his or her life and what makes sense in his or her own particular living and their dying. Indirect hope, or "living in the light of hope," sustains the patient's perception of meaningfulness – often in relation to a variety of direct hopes concerning what were once the taken-for-granted activities of daily life.

In this respect (although not specifically mentioned by Waterworth), the concept of medical "futility" differs from that of hopelessness. As many experienced doctors are aware, while it may be physiologically futile to prolong a patient's suffering by continuing to keep them alive a few days longer by "artificial" means, these extra days may be highly meaningful for the patient, hoping (directly), for example, to see her grandson who is flying back to her from the other side of the world, and living in the (indirect) hope of a meaningful life and death. Dying in the light of hope, in the absence of despair, hopelessness, or fear, Waterworth observes, may even be considered a "healthy" approach to dying.

Waterworth's objection to limiting hope to a propositional attitude is shared by *A. Mittleman* (Mittleman 2009) and for similar reasons. Writing from a religious as well as philosophical point of view, Mittleman discusses not only the views on hope of the classical and early modern philosophers mentioned above but also references to hope, albeit often in passing, of others, including Plato and the Stoics from the Classical era; Hobbes and Locke from the early modern; Hegel, Schopenhauer, and Nietzsche from the post-Kantian; and more recently and at greater length (see below) the Marxist Bloch, the Catholic Marcel, and a number of contemporary theologians. While Mittleman valuably fills out the picture of how Western philosophers have regarded hope, his account confirms the impression that philosophical analyses of this subject have oscillated rather than developed over time, suggesting a certain ambivalence toward hope. Mittleman traces this ambivalence back to the conflicting pre-philosophical narratives concerning hope of the cultures from which Western philosophy has developed: on the one hand the ultimately positive future-oriented dimension of much of the Judeo-Christian scriptures and on the other the rather less

positive attitude symbolized in the Greek myths of Pandora and Prometheus, in which hope either flutters ineffectively or deceitfully blinds humans to their inexorable destiny of death.

While Mittleman is primarily concerned with hope in the area of politics and thus has much less to say about the medical context than has Waterworth, two of the philosophers whose views he examines, *Ernst Bloch* (1885–1977) *and Gabriel Marcel* (1889–1973), are discussed with more specific reference to hope in the healthcare context by *W. F. Stempsey* (2014). Stempsey is concerned to distinguish in this context between what he calls "deep" and "shallow" hope. A patient's (or their doctor's) hope may be shallow when, for example, it focuses on, and pins too much on, unrealistic expectations of innovative or alternative therapies, or more generally overestimates the probability or success or cure: optimism is not the same thing as hope, and deeper hope may well be shown by a patient who underestimates the probability of benefit, but nevertheless remains hopeful – a "hopeful pessimist." Deeper hope also is likely to be characterized in its estimative dimension by imagining rather than fantasizing about alternative possibilities, not necessarily of different therapies but of different and perhaps more creative perspectives on present and possible future circumstances.

Deep hope in the medical context, Stempsey claims, is ultimately about "healing in the fullest sense," not just cure or alleviation of symptoms. The nature of what is hoped for is illustrated by what the Marxist atheist Bloch and the Catholic existentialist Marcel, each in their different ways, write about hope. Bloch views hope as a basic human impulse toward an ideal way of life, expressed in the great variety of utopian visions entertained throughout history. Medicine's utopian vision ultimately is the abolition of death, but in practice medicine settles for postponing death by seeking to relieve suffering and remove illness. For Bloch, hope demands much more than this: hope's struggle is not just for individual bodies to be repaired but, more importantly, for society to be transformed into a radically new human future. While Bloch's hope is Marxist, whereas that of Marcel is mystical, Marcel agrees with Bloch that in the medical context hope must not be limited to what it is technically possible for medicine to achieve. Hope may be sustained by a sense of meaningfulness beyond the limits of what is conceivable and may thereby enable a patient, even while accepting that their prognosis is terminal, to accept this prospect without giving in to despair or hopelessness or giving up on themselves. Such "deep" hope, often but not necessarily expressed in the particular religious language most familiar to the patient, is important for health professionals to distinguish from the wide range of specific hopes and expectations patients have when seeking medical assistance.

Stempsey's "deep" hope corresponds in many respects with what Waterworth characterizes as "indirect" hope or "living in the light of hope." These concepts have particular relevance to the context of terminal illness. Awareness of the significance of something similar in relation to illness generally is demonstrated by the physician and philosopher *Georges Canguilhem* (1904–1995). Canguilhem writes (1991, 196f.), for example, that what "characterises health is the possibility of transcending the norm, which defines the momentary normal, the possibility of tolerating

infractions of the habitual norm and instituting new norms in new situations" or again that health "is a way of tackling existence as one feels that one is not only possessor or bearer but also, if necessary, creator of value, establisher of vital norms."

From this brief consideration of the views on the subject of some contemporary philosophers, we may conclude that while "direct" hopes, that a specific or general state of affairs may come about, can most appropriately be categorized not as emotions but as propositional attitudes, the phenomenon of hope cannot be fully explained in these terms. "Indirect" or "deep" hope involves not just attitudes but also agency on the part of those patients who remain within the moral community not as moral patients but as moral agents. In this respect, hope that creates and sustains the effort to exercise moral agency might well be considered to be a virtue.

Healthcare Practitioners and Researchers

Few if any healthcare practitioners seem to doubt the positive role of hope in medical care and treatment. Hope as well as evidence has so far sustained the enterprise of medical research, and in the case of individual patients, only "unrealistic" hope, when there is "no hope," may be discouraged. This much is part of medical tradition, confirmed by the anecdotes of everyday clinical experience. But until the mid-twentieth century, it was mainly on the basis of tradition, anecdote, and experience that this positive evaluation of hope was founded. Unlike the philosophers discussed above, few if any healthcare practitioners attempted to dissect the concept of hope, investigate its dynamics, or systematically inquire into what patients meant when they spoke of their hopes.

During and after the second half of the twentieth century however, such dissections, investigations, and inquiries concerning hope proliferated in the healthcare literature, especially in that of nursing and oncology. The academic professionalization of nursing, together with advances in the treatment of cancer and in the care of terminally ill patients, seems to have helped stimulate this new interest in hope in healthcare settings. Academic nursing, requiring a theoretical basis other than (or as well as) that of medicine, sought this in the psychological and social sciences, to which researchers (some themselves nurses) more concerned with the psychological than the physiological and pathological aspects of cancer also turned. Of particular theoretical interest in relation to hope (judging by the frequency with which they are cited) was the work of earlier twentieth-century social and developmental psychologists, psychiatrists, and psychotherapists such as *Jean Piaget* (1896–1980), *Erich Fromm* (1900–1980), *Karl Menninge*r (1893–1990), *Erik Erikson* (1902–1994), and *Viktor Frankl* (1905–1997).

As well as the existential psychotherapist Frankl, the existentialist philosopher Gabriel Marcel, mentioned above, is frequently cited in the nursing literature on hope, mainly as a philosophical source for distinctions similar to that between "direct" and "indirect" hope. In two papers by nursing researchers who cite Marcel, for example, distinctions are drawn between "hoping for something" and "living in

hope" (Benzein et al. 2001) and "having hope" and "being in hope" (Lohne 2008). That hope could be characterized as either "particularized" or "generalized" seems to have been widely recognized (Kylmä and Vehviläinen-Julkunen 1997), but the grounds for living in hope or being in hope (as opposed to having a particular hope for pain relief or a long-awaited visitor) in the final stages of a terminal illness, for example, were not necessarily clear to those in a secular society who rejected religious ideas of an afterlife. For some, illumination, if not yet clarity, was provided by Marcel's undogmatic, even poetic, account of hope as the mutual and enabling affirmation of existence as mysteriously meaningful.

It is possible that, as a Catholic philosopher writing in a phenomenological rather than theological idiom, Marcel offered a conceptual approach sympathetic to those in academic nursing who wished to retain, in a secular context, some of the religious values traditionally associated with nursing, but in a language that was neither sectarian nor confessional nor even overtly theological. Marcel's particular style of existentialist-phenomenological philosophizing however is less concerned with systematically analyzing or defining concepts than with reflectively exploring philosophical ideas, often by discussing concrete examples in ordinary language, in order to evoke an understanding response in the reader as he or she begins, as it were, "to see what he is getting at." Marcel's indirect approach to philosophical ideas such as "mystery" and "availability" does not lend itself readily to crisp summarizing in academic journals primarily oriented to the problems of healthcare practice.

The attempts of later twentieth-century healthcare practitioners and researchers to analyze and define hope and related concepts, however, quite clearly are primarily concerned with the practicalities of healthcare. An early and frequently cited paper by the American nursing author *R. F. McGee* (1984), for example, summarized and drew together a variety of mainly psychological theories about hope, hopefulness, and hopelessness to propose a model in which patients were placed on a continuum from those who were "unrealistically hopeful" to those who were "unjustifiably hopeless," with intermediate cases of the "chronically fearful," the "fragile copers," and, ideally, the "realistic copers." The model assumed that most patients tended to fall into one or other of these "coping strategy" categories, but also that movement was possible and that if nurses were able to reliably assess and rank patients in these terms, appropriate interventions could be designed to move them toward realistic hopefulness. With similar ends in view, nurse researchers around this time began to construct psychometric scales based on various elements and dimensions of hope identified in mainly the psychological literature and designed to measure degrees of hope and hopelessness in patients (Miller and Powers 1988).

Instruments to measure hope and strategies to inspire or sustain it were to multiply in the nursing and healthcare literature over the last decades of the twentieth and first of the twenty-first centuries. There was considerable disagreement however about how exactly the relevant elements and dimensions of hope should be conceptualized. Part of the difficulty appears to be that writers and researchers approached the subject from different disciplinary traditions, so that hope could be conceptualized, for example, psychologically as an emotion or again more philosophically in

terms of an orientation toward meaning in life and in suffering. The difficulty was compounded, particularly in nursing, by research methodologies which sought to identify and refine concepts drawn out from the raw material of questionnaires or interviews with patients (at a single or different stages of one or more illness or condition) and then to develop these into theories about what the experience of hope, despair, and hopelessness meant to patients more generally. These methodologies produced many richly detailed accounts of how patients responded to being asked about these experiences, but at the same time tended to develop ever more complicated theories and models which in a number of respects often were inconsistent with one another.

A variety of interesting and, from the point of view of health professionals, potentially practical insights nevertheless have emerged from research of this nature, notably in relation to different dimensions of the dynamics of hope, despair, and hopelessness. The Finnish nursing researcher J. *Kylmä* and colleagues (Kylmä and Vehviläinen-Julkunen 1997; Kylmä et al. 2001; Kylmä 2005), for example, distinguish an "upward" from a "downward" "subprocess" in patients' experience of despair, the former implying that the energy of hope has not been entirely exhausted in despair: they suggest that clinical guidelines for identifying these subprocesses could assist health professionals in finding appropriate ways of encouraging (albeit not creating) hope in patients. British researchers *M. Corbett* and colleagues (Corbett et al. 2007) confirm the relevance of these findings and also of the distinction between generalized and particularized hope, in their research into patients' experience of living with low back pain. Kylmä and colleagues also discern "subprocesses" (abandoning hope and being without it) in hopelessness, and they identify hopelessness rather than despair as the "polar opposite" of hope, claiming that their research is the first to make these distinctions – the first presumably in the healthcare literature: in philosophy, Waterworth (above and apparently independently) makes similar distinctions regarding hope, despair, and hopelessness.

Further aspects of the dynamics of hope in patients across a broad range of conditions are described by Canadian nursing researchers *J. M. Morse* and *J. Penrod* (1999). They see hope as the final stage of a process through which, optimally, a patient becomes able to accept their condition in ways that make the present bearable and to "reformulate" their self-understanding. The earlier stages, which are not (unlike the deliberations of hope) "strategies" but "reflexive responses," can progress from "enduring" (initially being unable to take in what has happened) through "uncertainty" (beginning to recognize the implications) to "suffering" (despairingly "acknowledging" rather than fully accepting), but patients may also become "stuck" in any one of these stages and the suggested clinical relevance of these findings is to assist in identifying "strategies for promoting patient comfort" appropriate to each stage.

Despite these and many other efforts by healthcare researchers to define hope, to identify its components and dynamics, and to construct scales, models, and theories which might ultimately be of clinical relevance, agreement on what hope is remains elusive. Not surprisingly, reports on research into hope frequently conclude with the recommendation that more research is required to clarify the concept. Reviewing and

reflecting on this failure, some healthcare researchers now argue that attempts to understand the essence of hope as if it were a single identifiable phenomenon are fundamentally mistaken and that different research approaches are required in order to understand hope in ways that are practically relevant to patient care and clinical practice.

Among the most illuminating of these approaches is that of the Australian research psychologist and medical oncologist *J. A. Eliott* and *I. N. Olver* (2002, 2007, 2009), who analyze the ways in which cancer patients speak about hope, spontaneously (in the course of interviews about their views on resuscitation decisions, first with outpatients and then with patients aware they would soon die), and also when dying patients were asked directly to talk about hope. When speaking spontaneously, the patients spoke in different ways of hope, when using hope as a noun and when using hope as a verb.

As a noun (often as in "no hope" of cure or remission), hope was spoken of predominantly as something either present or absent, objectively determined by the doctor, with consequences which the patient could not alter. Also as a noun however, hope was sometimes spoken of as something subjectively determined by the patient, which might grow or diminish, depending on the patient's interpretation (which might differ from the doctor's interpretation) of the care and treatment they were receiving and also depending on their hopes for medical progress generally or even for medical "miracles." This "subjective" hope, Eliott and Olver argue, may well be a helpful "coping resource" for a patient: to decry it as "false" or "unrealistic" hope is to confuse it unhelpfully with "objective" (but never infallible) medically defined hope.

Spoken of by patients as a verb, hope was not only again subjective but also essentially personal ("I hope"), active (willing rather than just wishing), and having a beneficial if indeterminable effect not dissimilar to that of a placebo. This was the case even with cancer patients who were aware that they had only a short time to live, but whose positive use of hope as a verb reflected their desire to live in the present as fully and meaningfully as they were able, and to remain engaged, valuing and valued by others, especially those closest to them.

When directly asked about their understanding of hope, dying patients spoke of it as an essential aspect of human life and resource for living ("there's always hope"), but again also in terms of the consequences of having greater or less hope. Hope was related not to death but to making life as good and meaningful as possible, again with an emphasis on valuing and being valued by others, especially family or religious community, both in the time remaining and as far as possible in the memories one left behind. There was even a sense in which to go on hoping might be "morally required," since not to do so could be felt as failing to value not only the patient's own life but also their relationships with those who cared for them. (There are echoes here of hope as an Aristotelian virtue or a Kantian imperative.)

Considering the practical implications of their findings for patient care, Eliott and Olver argue that instead of trying to calibrate patients' levels of hope and devise interventions to adjust them to some optimal therapeutic standard, clinicians should

first of all listen, in order to learn what kind of hope a particular patient is expressing or inquiring about. A dying patient's hope to live (meaningfully, now, with others, and in memory long) need not be inconsistent with their acceptance that there is "no hope" (of a medical cure or remission). This is because – as a nursing researcher commenting on their work points out (Dorcy 2010) –"language is not a neutral, transparent description of the objective world" but "a culturally contextual explication of the speaker's intentions."

That last observation, in conclusion, may provide a clue to why hope has been analyzed by philosophers and by health professionals and researchers in the many diverse and sometimes inconsistent ways reviewed above. Human perception is inevitably selective and contextual: our understanding of the world and of one another is always from some perspective or other, formed by the different cultures we inherit and inhabit. It is not surprising therefore that philosophers, concerned with truth and precision, and healthcare professionals and researchers, concerned with health and healing, come to different (albeit sometimes overlapping) conclusions about hope. Nor is it surprising that patients (who could be anyone) may conceive of hope in ways that can seem to the philosopher confused or to the professional unrealistic. Yet neither the philosopher nor the professional can deny that for any account either of truth or of health to be sufficient, it cannot replace without remainder the first person perspective by the third person perspective – what I think or feel or hope by what another says about what I think or feel or hope.

This is not to adopt Humpty Dumpty's "a word means just what I choose it to mean": "I hope" cannot mean "I despair." "I hope," however, can mean just that. The poet-philosopher Coleridge mentioned at the beginning wrote in a late poem, probably mourning a lost love, "Hope without an object cannot live." For once the sage was wrong, and perhaps he demonstrated that when, knowing that he was on his deathbed, he remarked that he felt he "even could be witty." To hope, not necessarily for anything in particular, can be to intend to be resolute, albeit without any particular resolutions. It can, of course, be many other things as well. What do patients mean by hope? If in doubt, ask the patient.

Definitions of Key Terms

Virtue	positive moral disposition integrated with the whole person
Propositional attitude	attitude toward a statement about a future possibility
Hopelessness	total loss of hope
Despair	oscillation between hope and hopelessness
Direct or particularized hope	hope that or for something identifiable
Indirect or generalized hope	hope expressive of values or meaning

Summary Points

- Hope is not simply an emotion nor simply a propositional attitude, and it may be a virtue.
- Indirect or generalized hope(fulness) exists as well as direct or particularized hope "that."
- Hope may still be alive in despair but not in hopelessness.
- Patients as well as doctors may accept "no hope" as medically determined.
- Patients also may see hope as subjectively or morally determinable by the patient.
- The clinical relevance of the patient's hope is discovered by individualized care.

References

Benzein E, Norberg A, Savemann B (2001) The meaning of the lived experience of hope in patients with cancer in palliative home care. Palliat Med 15:117–126

Canguilhem G (1991) The normal and the pathological. Zone Books, New York, p 196f

Corbett M, Foster NE, Ong BN (2007) Living with low back pain – stories of hope and despair. Soc Sci Med 65:1584–1594

Day JP (1969) Hope. Am Philos Q 6(2):89–102

Day JP (1998) More about hope and fear. Ethical Theory Moral Pract 1(1):121–123

Dorcy KS (2010) Hegemony, hermeneutics, and the heuristic of hope. Adv Nurs Sci 33(1):78–90

Eliott J, Olver I (2002) The discursive properties of "Hope": a qualitative analysis of cancer patients' speech. Qual Health Res 12(2):173–193

Eliott J, Olver I (2007) Hope and hoping in the talk of dying cancer patients. Soc Sci Med 64:138–149

Eliott J, Olver I (2009) Hope, life, and death: a qualitative analysis of dying cancer patients' talk about hope. Death Stud 33:609–638

Kylmä J (2005) Despair and hopelessness in the context of HIV – a metasynthesis on qualitative research findings. J Clin Nurs 14:813–821

Kylmä J, Vehviläinen-Julkunen K (1997) Hope in nursing research: a meta-analysis of the ontological and epistemological foundations of research on hope. J Adv Nurs 25:364–371

Kylmä J, Vehviläinen-Julkunen K, Lähdevirta J (2001) Hope, despair and hopelessness in living with HIV/AIDS: a grounded theory study. J Adv Nurs 33(6):764–775

Lohne V (2008) The battle between hoping and suffering. Adv Nurs Sci 31(3):237–248

Macquarrie J (1978) Christian hope. Mowbray, Oxford

McGee RF (1984) Hope: a factor influencing crisis resolution. Adv Nurs Sci 6:34–44

Miller JF, Powers MJ (1988) Development of an instrument to measure hope. Nurs Res 37(1):6–10

Mittleman A (2009) Hope in a democratic age. Oxford University Press, Oxford

Morse JM, Penrod J (1999) Linking concepts of enduring, uncertainty, suffering and hope. Image J Nurs Sch 31(2):145–150

Stempsey WF (2014) Hope for health and health care. Med Health Care Philos. doi:10.1007/s11019-014-9572-y

Waterworth JM (2004) A philosophical analysis of hope. Palgrave Macmillan, Houndmills

Woodring C (ed) (1990) The collected works of Samuel Taylor Coleridge, vol 14, Table talk I. Routledge, London, p 331

Dignity of the Patient

29

Andrew Edgar and Lennart Nordenfelt

Contents

A. Edgar (✉)
Cardiff University, Cardiff, UK
e-mail: edgar@cardiff.ac.uk

L. Nordenfelt
Ersta Sköndal University College, Stockholm, Sweden
e-mail: lennart.nordenfelt@liu.se

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_26

Abstract

This chapter will explore the concept of "dignity," with particular reference to its use in the health-care setting. There is a substantial philosophical literature on dignity, both in discussion of its relevance or otherwise to bioethics, and perhaps more fundamentally, as to the precise meaning of the term. The chapter proceeds by introducing the way in which dignity language is used by patients and patient advocates. It is, crucially, a term that non-philosophers find effective in articulating their moral demands and protests. The chapter will then review the use of the term in international and national policy documents, before reviewing a number of core positions in the philosophical debate. Ruth Macklin's rejection of the concept of "dignity," as at best reducible to the more fundamental concept of autonomy, provides a stepping-off point. Defenders of dignity may be seen to take a number of different tactics. A "metaphysical" conception, as found most influentially by Kant, argues forcefully that dignity is a mark of the moral status of humans regardless of any empirically perceptible capacities or qualities they may possess. Other philosophers (such as Nussbaum) either seek to link dignity to empirically identifiable capabilities or to differentiate and articulate diverse uses of the term, thereby untangling moral arguments and identifying core or dominant senses. Running through these debates are a series of problems that affect the dignity of the patient. Most importantly, there lies the problem of respecting patients who either are unwilling or unable to assert their own dignity claims. These may be patients whose subjective sense of dignity has been so eroded that they no longer recognize themselves to be worthy of dignified treatment; others may have lost the capacity for autonomous action altogether. The worth of the language of dignity, and of different accounts of dignity, may be seen to be tested by their applicability to such cases.

Introduction

The precise meaning of "dignity" is highly contested, as is its relevance to health-care ethics. Its use may at times amount to little more than a rhetorical flourish. However, it is a term to which not merely philosophers but also patients and their advocates appeal and appeal frequently. It expresses something fundamental about both human aspirations and human vulnerability. To claim that all humans have a right to be respected in terms of their dignity, regardless of their particular qualities, capacities, and achievement, may seem to be a fundamental assertion of moral values. This chapter will therefore review uses of dignity language by patients and their advocates, before briefly reviewing the history of the concept in policy documents. Philosophical discussion is wide-ranging and complex. No review can hope to cover, exhaustively, the range of material available. The discussion of moral theory that follows therefore seeks to highlight a number of representative arguments, both in defense of dignity and against it, with particular reference to the use of those arguments in defending the dignity of the patient.

Dignity, Policy Documents, and Ordinary Language

Introduction

The term "dignity" is much used in discussions of medical ethics and the ethical treatment of patients, be these discussions that are held within academia, within the medical profession, or by journalists, politicians, and policy makers, and of course as expressed in "lay" debates, not least in protests over the poor quality of patient care (see Lanigan (2008)).

A survey of members of the public and health-care professionals conducted by the UK Department of Health, on the theme of "dignity in care," is illustrative of the complex and diverse uses to which the word is put (see webarchive.nationalarchives.gov.uk/20140103222317/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4139552). Under the term "dignity" respondents included the failure of staff to treat patients as individuals or to address them appropriately (and with too much familiarity); the failure to respect privacy (particularly in the use of lavatories and bathrooms but also in the existence of mixed sex wards); the lack of assistance with eating; the lack of advocacy, including the problem of being unable to complain effectively; the problems the patient faced in maintaining an appropriate appearance; being deprived of a sense of purpose while in hospital; and finally the frustrations that language barriers create in communication. Respondents also complained that the term itself was poorly defined. The survey, while by no means exhaustive of uses of "dignity," might suggest that the term is a catchall for all types of complaint. However, more subtly, the survey highlights the importance of one's status both as a unique individual and as an autonomous agent, one's dependence upon others, and the vulnerability of one's self-esteem, thus one's sense of dignity, in the face of neglect or abuse. The control of privacy, the presentation of oneself to others, and effective communication, including complaining and if necessary having an advocate, may all be seen as preconditions of dignified agency. As such, the survey begins to highlight core issues within the philosophy of dignity, not least the centrality of notions of autonomy, privacy, and communication (and have one's communication respected).

The UK Department of Health's Dignity in Care initiative can be seen in part as an attempt to operationalize "dignity," so that demands for greater dignity might have a real influence on standards of care. In this, the department highlights the fact that the term is frequently poorly defined and used somewhat rhetorically. A brief overview of some social and health-care policy documents may illustrate both this weakness and developing efforts to define and operationalize "dignity" with some precision.

International Declarations and National Policy

While the Mexican constitution of 1917 would seem to be the first policy document actually to mention the concept of human dignity, the United Nations' *Universal*

Declaration of Human Rights of Man (1948), beginning with the assertions that: "All human beings are born free, equal in dignity and human rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood," remains the most famous of the early documents. The document is important, not least, in presenting dignity as the grounding of the universal and equal moral status of all human beings. The German Basic Law from 1949 is particularly explicit about fundamental human values, and the concept of "dignity" occupies a central place in it. In an inaugural paragraph, the constitution says: "Human dignity is inviolable. To respect and protect it is the duty of all state authority" (see www.iuscomp.org/gla/statutes/GG.htm). In both the German Basic Law and the *Universal Declaration*, the appeal to dignity may be seen to express a moral outrage at Nazism and a quiet assertion of "never again." Something similar is reflected in World Medical Association's *Declaration of Geneva* (1948) and later *Declaration of Helsinki* (1964). The former is a pledge expressing the doctor's duties toward their patients, which includes the duty to "practice my profession with conscience and dignity" (www.wma.net/en/30publications/10policies/g1/

index.html). The Helsinki Declaration addresses specifically the ethics of medical research (not least in the light of the atrocities of Nazi research) and includes the clause:

> It is the duty of physicians who are involved in medical research to protect the life, health, dignity, integrity, right to self-determination, privacy, and confidentiality of personal information of research subjects. The responsibility for the protection of research subjects must always rest with the physician or other health care professionals and never with the research subjects, even though they have given consent. (www.wma.net/en/30publications/10policies/b3/)

In later documents that refer to medicine and medical research, the Council of Europe and UNESCO have both created declarations and conventions that appeal to dignity. *The Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine* from 1997 is the main contribution from the Council of Europe, asserting that: "Parties to this Convention shall protect the dignity and identity of all human beings and guarantee everyone, without discrimination, respect for their integrity and other rights and fundamental freedoms with regard to the application of biology and medicine" (http://conventions.coe.int/Treaty/en/Treaties/Html/164.htm). UNESCO's *Universal Declaration on the Human Genome and Human Rights*, also from 1997, is significant in that it is the first text to treat the question of human dignity as having a bearing on problems regarding scientific progress. The declaration recognizes in particular that respect for human dignity must take precedence over scientific research on the human genome and its applications. The text says in its article 2: "Everyone has a right to respect for their dignity and for their rights regardless of their genetic characteristics" and "Dignity makes it imperative not to reduce individuals to their genetic characteristics and to respect their uniqueness and diversity" (http://www.unesco.org/new/en/social-and-human-sciences/themes/bioethics/human-genome-and-human-rights/).

Recently, national governments have made increasing reference to "dignity" in health-care policy. Thus, the Swedish *Health and Medical Services Act* (of 1997) asserts, in § 2, that: "Care shall be given with respect for the equal value of all human beings and for the dignity of the individual." In 2000, the UK Labour government's *New NHS Plan* similarly states that "The NHS will treat patients as individuals, with respect for their dignity…. They have the right to be treated with dignity and respect." In 2010, the UK's incoming Conservative-Liberal coalition government wanted "a sustainable adult social care system that gives people support and freedom to lead the life they choose, with dignity" (§ 1.17 [www.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf]).

The appeal to dignity in policy documents is laudable. However, it is not always immediately clear what work the concept of "dignity" is doing precisely or how well it has been theorized. Becker, for example, argues that there is little evidence that the drafts of the *Universal Declaration* engaged in any discussion of the meaning of the term (Becker 2001). The suspicion that "dignity" plays a largely rhetorical, rather than genuinely analytic or argumentative, role may be supported if the term is removed from any of the passages cited above. Their meaning is typically barely changed. This suggests that the meaning of terms such as "dignity" and "respect for dignity" is actually encapsulated in attendant phrases, such as a recognition of equal rights and the need to respect autonomy and consent, the need to avoid discrimination or objectification, or the need to recognize individual uniqueness and diversity and freedom of choice. This is illustrated by the *United Nations Principles for the Older Person* (1991, §§ 17 & 18 [http://www.un.org/documents/ga/res/46/a46r091.htm]) that uses "dignity" as a heading, under which it is asserted that: "Older persons should be able to live in dignity and security and be free of exploitation and physical or mental abuse. Older persons should be treated fairly regardless of age, gender, racial or ethnic background, disability and other status, and be valued independently of their economic contribution." It is not clear what the word "dignity" adds to this proposal. Thus, while the "dignity" of the human being typically is asserted as the foundation of his or her moral status, the exact nature of this foundation goes without explanation, leaving the practical import of the claim vague.

Some recent documents from governments and NGOs have begun to give more substance to the concept, through empirical research and conceptual reflection. This more rigorous approach serves to explicate, in some detail, the practice that a respect for dignity will entail. Perhaps, the work of the Department of Health's Dignity in Care Initiative that was encountered above has been the most significant, at least within the context of the UK health-care systems (www.dignityincare.org.uk). Here a (still somewhat nebulous) dictionary definition of "dignity" – a "state or quality or manner worthy of esteem and respect, and (by extension) self-respect" – is developed, with reference to good practice in care, as: "The kind of care, in any setting, which supports and promotes, and does not undermine, a person's self-respect regardless of any difference" (http://www.scie.org.uk/publications/guides/guide15/selectedresearch/whatdignitymeans.asp). This definition, informed by the survey material, summarized above, which opens an awareness of the concrete actions that undermine a sense of self-respect, grounds a "ten-point Dignity Challenge" for the medical practitioner:

Have a zero tolerance of all forms of abuse.
Support people with the same respect you would want for yourself or a member of your family.
Treat each person as an individual by offering a personalized service.
Enable people to maintain the maximum possible level of independence, choice, and control.
Listen and support people to express their needs and wants.
Respect people's right to privacy
Ensure people feel able to complain without fear of retribution.
Engage with family members and carers as care partners.
Assist people to maintain confidence and a positive self-esteem.
Act to alleviate people's loneliness and isolation (www.dignityincare.org.uk/About/The_10_Point_Dignity_Challenge/).

The UK Care Quality Commission's report on *Dignity and Nutrition* (2011, p. 4) similarly offers empirical examples of failure to respect dignity. It notes that: "People were not spoken to; people were left without call bells, ignored for hours on end, or not given assistance to do the basics of life – to eat, drink, or go to the toilet." Again, a core link is made between dignity and autonomy, but crucially within the context of a recognition of the potential vulnerability of the older person and thus the need to support them in achieving autonomy (or indeed in compensating for the autonomy that they have lost).

Empirical evidence and principles of dignity, such as that provided by the Care Quality Commission and the Dignity in Care Initiative, suggest that the language of dignity is appealed to in protest against certain morally offensive experiences and that these experiences have some degree of family resemblance, and it is this that the term "dignity" encapsulates. Dignity responds to the nuances of context and thus places different demands upon carers who have different categories of patient and who work in different care contexts. But further, such research and analysis allows the treatment of patients with dignity to become something that can be documented. In effect, dignified care is conceptualized in terms of a series of behaviors as well as the degree of awareness of patient need that is expected of the practitioner. The dignity of the patient is secured through the virtuous behavior of the practitioner. The failure to treat with dignity, and thus absence of the appropriate virtue, is readily identifiable, allowing sanctions and training as necessary.

In summary, policy documents have striven to give substance to the concept of "dignity," moving away from a somewhat rhetorical gesture that flags up the universal moral status of the human being. These recent documents, drawing on empirical research and reviews of the philosophical literature, explicate what respect for dignity might mean for nursing and other medical practice. Such efforts are driven by a recognition of concrete examples of bad practice and, most typically, as the Dignity in Care Initiative suggests, rest upon patients' subjectively perceived violation of their dignity as a loss of self-respect, self-worth, or self-esteem.

The Uses of "Dignity"

It may be suggested that the problems faced in formulating policy documents and declarations are indicative of the somewhat complex use of the term "dignity" in ordinary language. Much empirical work has been done of the use of the term dignity, particularly in reference to care of older people (Calnan 2005, Franklin et al. 2006). This includes research, by the Dignity and Older Europeans project (www.cardiff.ac.uk/socsi/dignity/europe/), exploring the different nuances of the term in its use across Europe, in different medical professions, and among lay people (Arino-Blasco et al. 2005; Bayer et al. 2005; Stratton and Tadd 2005; Tadd and Calman 2009). This research indicates, in line with the research of the Care Quality Commission and the Dignity in Care Initiative, that the word does not have a single and unambiguous use. Its use is not merely contextual, but dependent upon cultural norms and ideals of good human behavior and deportment. It may further be suggested (see Edgar 2003, p. 116) that the precise meaning of the word in any given context typically depends upon what its opposite is supposed to be. This is reflected in Shotton and Seedhouse's emphasis on the way in which the use of "dignity" is responsive to tension between a person's competences and the particular contexts in which they find themselves and thus "when we are in situations where we feel foolish, incompetent, inadequate or unusually vulnerable" (Shotton and Seedhouse 1998, pp. 246–247).

The predominantly negative, or perhaps more properly reactive, meaning of "dignity" can be seen in the following examples of violations of dignity:

> organ sales from living "donors"; seeking patent rights over human genes; making animal–human chimeras; obliging someone to live in abject poverty; pornography; torture; sex selection by preimplantation genetic diagnosis; death in irremediable physical or psychological suffering; abandonment to senility in a nursing home. . .. Socialists used to speak of the "dignity of labour". Pico della Mirandola wrote of all activities that extend human powers or show them off to aesthetic or practical advantage as contributions to the "dignity of man." (Ashcroft 2005 p. 679)

In these cases "dignity" is being opposed, variously, to the violation of bodily and species integrity, the lack of basic resources or opportunity to provide for oneself, the objectification of the human subject, the violation of human rights or autonomy, the prolonging of unnecessary suffering; and perhaps, in the case of the dead, the violation of the sensibility of survivors. In this summary list, it is only in the comments of the Renaissance scholar Mirandola (1985) that a positive definition of dignity is found, in terms of the potential of human development. Yet even here it may be suggested that Mirandola's meaning is most significantly reflected today in examples of abject poverty or unemployment (and to which might be added, denial of education and cultural opportunity), where it is that human potential that is being stifled.

It is this dependence of the meaning of dignity on its opposite that is reflected in policy documents, as well as everyday use of the term. Policy typically seeks to protect the vulnerable, and subjectively we feel no need to assert our dignity or

demand respect for our dignity unless that dignity is threatened. Dignity matters when we are embarrassed, humiliated, objectified, degraded, confused, and thus when we are most vulnerable (see Kaufmann et al. 2011). The appeal to concrete examples of bad practice is thus informative, not least for anchoring a particular use of "dignity." Yet, so too is the fundamental assertion of human dignity as the foundation or encapsulation of the moral value and status of the human being. Merely subjective accounts of dignity, which rely upon people consciously experiencing indignity, lead to a paradox. It is readily conceivable that someone may be so degraded that they internalize a sense of worthlessness and as such no longer protest their entitlement to be treated with dignity. It is through an assertion of the inalienability and objectivity of a right to be treated with dignity that the moral demand for the protection of such degraded and vulnerable people can be articulated.

In concluding this review, it may be acknowledged that policy documents can and do function effectively. They legislate against moral abuses and serve to reform behavior, including practices in health care. Moral theorists may nonetheless inquire into the logical coherence of the everyday use of "dignity," and seek to untangle the term's fluidity and ambiguity. If successful in this process of clarification, moral theory may redeem and ground the use of "dignity." If unsuccessful, moral theorists may reject the term as irredeemably incoherent.

Dignity and Moral Theory

Approaches to "Dignity" in Moral Theory

A brief history of early attempts to theorize "dignity" may serve to orient a discussion of moral theory. "Dignity" as a moral concept can be found in Cicero in the first-century BCE and in the Roman Stoics. While the dominant use of dignity [*dignitas*] in the Roman world referred to the social status that one achieved or that was bestowed, so that dignity was a mark of rank and hierarchical superiority, Cicero also treats dignity as the intrinsic quality that distinguishes humans from all other beings (1913). As such, dignity begins to ground the claim, which continues to be reflected in documents such as the *Universal Declaration*, that all human beings have moral status simply by the fact that they are human. Stoicism similarly looks to the concept of dignity to articulate the moral status that humans have, over and above the artificial distinctions of rank. For the Stoic, further, dignity lies in the capacity of humans to live well, thoughtfully and reflectively, in the face of suffering and adversity. It may be noted that, as such, the Stoic conception of dignity is that of a virtue. Dignity is expressed in the exercise an appropriate habit of behavior. The Renaissance scholar Pico della Mirandola argues in his *Oration on the Dignity of Man* (1984 [1481]) that the human being, in contrast to any other animal, is uniquely free and thus has a special dignity. Pico thinks that God has assigned human beings to be creatures of indeterminate nature. The human being is free to decide their own way of life.

Contrary to common belief, there is little basis for the present use of the term "human dignity" (as corresponding to the Hebrew *gedula* or the Greek *semnotes*) in the Christian scriptures nor in translating most of the influential medieval theologians such as Augustine. A theological interpretation of dignity is a product of the late nineteenth century. Dignity, as an expression of the universal and distinctive moral status of humanity, is interpreted in terms of the idea of human beings being made in the image of God. For that reason, the theologians claimed humans have intrinsic worth and value beyond all price. Pope Leo XIII officially introduced the idea of the dignity of man in his *Rerum Novarum* 1891. A much later, but very significant, papal pronouncement about dignity is that of John XXIII in his encyclical *Peace on Earth* from 1963 (see http://www.papalencyclicals.net/).

Kant's articulation of a theory of dignity is, in many respects, the most influential. For Kant "dignity" is explicitly expressive of the fundamental moral worth of human beings (Kant 2002) and as such encapsulates the notion of dignity that stands as the core of the *Universal Declaration* and other similar documents. While most objects and creatures are of worth only instrumentally, which is to say they have a "price," humans have intrinsic worth. The source of this worth lies in the capacity of humans to use reason and thus to be autonomous. Kant asserts that autonomy is "the ground of the dignity of the human and of every rational nature" (2002, p. 54). Kant's argument does not then simply equate dignity with the exercise of autonomy or freedom, and the subtlety of Kant's argument may be noted on two grounds. Firstly, autonomy is understood by Kant as a "noumenal" quality. That is to say that Kant is not concerned with the actual capacity to exercise autonomy possessed by empirical human beings. Autonomy as a noumenal property is rather the condition of possibility of autonomous action that cannot itself be empirically experienced. It is attributed to human beings regardless of their perceptible capacity to exercise that autonomy. Thus, a patient with advanced dementia or in a coma might be thought to retain their dignity, in Kant's sense, regardless of the fact that they are incapable of acting autonomously. Attributing dignity to all humanity is thus a call to recognize the moral status of one's fellow human being, regardless of their empirically identifiable qualities or capacities, and indeed a call upon the individual to respect their own dignity. Secondly, Kant's understanding of the relevant sense of "autonomy" may be further refined: it is the noumenal capacity for the autonomous use of reason, that is, the condition of possibility of dignity, and more precisely the use of reason to recognize and obey the moral law. Thus, dignity is entwined with humanity's unique and distinctive status as a moral being.

To move beyond historical accounts of dignity to contemporary debates, it may be suggested, following Ashcroft (2005), that moral theories may be categorized into a number of different groups, according to their approach toward "dignity." For Ashcroft, the first group rejects "dignity" as at worst incoherent and at best reducible to some other moral concepts, such as autonomy or integrity (Cochrane 2010; Harris 1998; Macklin 2003). The second group retains "dignity" as a useful concept, but seeks to articulate the concept more precisely, rejecting the notion that it is reducible to a single more basic concept, and thus arguing that different meanings of dignity need to be articulated, to at once avoid confusion and to articulate the range of appeal

and application of the concept (Beyleveld and Brownsword 2001). A third group seeks to ground the understanding of dignity in the exercise of certain human capabilities, functions, or processes of social interaction (Nussbaum 2008). The final group defends dignity, most strongly, as an inviolable metaphysical property and as the foundation of unique moral status of the human being as such, regardless of any particular qualities the individual human may possess or lack (Kant 2002, Kass 2002). Beyond this, a number of thinkers offer multi-part typologies of "dignity" that attempt to capture and map its different uses and meanings, albeit typically arguing for the interrelationship between meanings and the priority of an ultimate and grounding meaning (Nordenfelt 2004; Mann 1998).

The Redundancy of "Dignity"

The key representative of Ashcroft's first group is Ruth Macklin (2003), who boldly asserts the uselessness of the concept of "dignity." Her argument, in part, follows the observations already made about the use of dignity in early declarations and policy documents. The word is doing little other than making a rhetorical flourish. Going beyond this, the vacuity of the concept may be seen in the fact that it is appealed to by proponents at opposite sides in bioethical debate (Schulman 2008). Dignity can be used to defend abortion, euthanasia, and human enhancement or to condemn them. The nature of a dignity argument cannot then be reconstructed simply from an appeal to the word itself. Everything depends upon the way in which the word is interpreted. Macklin therefore argues that if the dignity claim does have substance, that substance is derived from other more rigorously formulated terms and most significantly from claims about autonomy. Consequently, respect for dignity is resolvable into the demand to sustain the conditions that enable the autonomy of beings capable of rational thought and action. As such, it may be suggested, Macklin's argument very much reflects what is sometimes referred to as "old bioethics." Here the focus is on the relationship between physician and patient, and the issue of autonomy and informed consent is placed as its center. This has been seen, for example, in the Declaration of Helsinki, noted above. Macklin's final point is to question the origin of the concept, suggesting that its continuing use owes much to Roman Catholicism, and by implication that "dignity" might too readily be tied to a single religious doctrine, and that this is undesirable in a multi-faith or secular age.

Macklin's criticism of dignity has some purchase with respect to the brief history given above. Cicero, the Stoics, and Pico ground dignity in the fact that humans are endowed with freedom and reason, and thus it may be readily suggested that the appeal to dignity is simply a celebration of human autonomy. Similarly, Pope Leo XXIII's *Rerum Novarum* ultimately links dignity to such natural rights such as private ownership of property, life, labor, and marriage, all of which may be interpreted as preconditions for an autonomous existence (see Leget 2013, p. 946). On Macklin's account, Kantian dignity is similarly reducible to autonomy, albeit to noumenal moral autonomy, rather than to the empirically perceptible exercise of autonomy. While Kant's account of the noumenal may not easily be defended, the

importance of the difference between the noumenal and the empirical perceptible nonetheless highlights potential problems in Macklin's approach.

Two problems may be considered. Firstly, if Macklin is correct and dignity means nothing more than the moral significance of exercise of autonomy, then it follows that dignity is irrelevant to those who lack autonomy. Dignity and dignified care would, for example, be irrelevant for patients with chronic disorders of consciousness (Varelius 2009), and there would be no sense in demanding respect for the dignity of the dead (Lantz 2009). If one is prepared to reject "dignity" from the moral vocabulary, then this is, of course, not a problem at all. Presumably, other sources of moral value for the non-autonomous could be found. However, respecting the moral intuitions that the term "dignity" expresses in lay uses, it may yet seem appropriate to use the language of dignity to defend the moral status of precisely those incapable of exercising autonomy. Dignified care for the comatose, or for those with advanced neural degeneration, seems to be an appropriately articulated demand and indeed a core use of the term "dignity." Respecting dignity cannot then simply be the safeguarding and enabling of the empirically perceptible exercise of autonomy, for it continues to apply even when that empirical capacity has been irrevocably lost.

Secondly, there is the now notorious case of dwarf tossing (Beyleveld and Brownsword 2001, pp. 25–27). In 1995, a French court ruled that the practice of dwarf tossing, whereby human dwarfs allowed themselves to be thrown as an entertainment, violated the dignity of the dwarf. While this might seem intuitively correct, one of the dwarfs in question, Manuel Wackenheim, protested the decision on the grounds that it violated his autonomy. He argued that he freely chose to be involved in the activity and further making it illegal would deprive him of a source of income. The court's decision followed a broadly Kantian line, in implying that just as one's dignity can be violated by others, so also one can violate one's own dignity, and thus one should be prevented from doing so. Here the appeal to dignity does not reduce to autonomy, but seems to run contrary to it. This problem will be addressed below, with respect to the work of Beyleveld and Brownsword. Prior to this, a recent version of the metaphysical theory of dignity will be rehearsed.

The Metaphysical Defense of Dignity

As noted above, defenders of dignity can be seen to take a number of strategies. Most fundamentally, as with Kant, dignity may be asserted as an ontological property that separates humans from all over creatures and that, as such, determines the moral status of the human being as a human being. This approach, implicit to the *Universal Declaration*, would, if successfully defended, overcome the problem of the dignity of marginal humans, such as those lacking autonomy, and perhaps more importantly, shift the bioethics debate away from a more or less exclusive focus on autonomy and informed consent. Such a focus is inadequate for dealing with the problems of a new bioethics, developed in the face of new medical technologies, such as reproductive and genetic technology, medical support at the end of life, and organ transplantation.

If a theory of dignity can encapsulate what is the moral core of humanity, it may then articulate the limits to which technology can be allowed to impinge upon and change that core – and thus the limits of posthuman dignity (Bostrom 2005). Equally, if dignity embraces all humans, simply on the basis that they are human and regardless of the qualities they possess, then marginal humans would be demonstrated to continue to be worthy of respect.

Leon R. Kass offers a metaphysical account of dignity that, like Kant's, does not rely upon either the particular empirical capacities of the individual human being or yet upon any biological facts that may be taken as defining the human species (Kass 2008). Kass is, nonetheless, critical of Kant. For Kass, the very austerity of Kant's vision fails to recognize the importance that the embodied and vulnerable nature of humanity has for understanding dignity, and as such Kant has little to contribute to the new bioethics. For Kass, an understanding of human dignity must recognize a tension between the "dependency" of human beings and their capacity for "transcendence." For Kass, dignity is manifest on a mundane level in humanity's struggle with its weaknesses and biological needs and thus in "endurance and equanimity, generosity and kindness, courage and self-command" (p. 220). But unlike any other animals, Kass argues, humans are always capable of aspiring to something higher. In a move that would alarm Macklin, Kass illustrates his points through an appeal to Judaic-Christian scripture and the creation stories in Genesis. He argues that the idea of seeing humans in the likeness of God offers a genuinely philosophical insight into human nature and its dignity and not a parochially theological one. In the Judaic-Christian creation myth, humans share with God the capacity to "speak, plan, create, contemplate, and judge" (p. 226). Yet, in this, humans are only *like* God, not being themselves divine. Humans remain animals and are thus vulnerable at the physical and embodied level. This lowly, dependent aspect of humanity is what drives human aspiration toward the transcendent, aspiring to new ways of living, be this through moral and artistic or scientific and technological advance. It is in this aspiration that human dignity is manifest.

A core problem encountered by a metaphysical account of dignity, and thus any account that argues that, simply by being part of the human species, one has a special moral status, is that it seems vulnerable to the charge of speciesism (Rachels 1990, p. 181). That is to say, it appears to privilege humans simply for being humans. As soon as specific human qualities are identified as the basis for dignity, then by extrapolation, if any other creature possesses those qualities, then it too has an equal claim to dignified treatment. An adult chimpanzee may thus have as much if not more autonomy than a young human child, and yet the metaphysical account seemingly attributes greater dignity to the human. Kass is aware of this problem and is dismissive of theories that link dignity to personhood, precisely because personhood is readily attributable to nonhuman animals. However, his own account does not obviously avoid the problem. At best, Kass can argue that marginal humans, those who do not manifest the core quality of aspiration, may demand respect precisely insofar as competent and aspirational humans manifest human dignity in their care for their vulnerable companions. The dependency of humans upon each other, and thus their social nature, becomes crucial.

Dignity and Capabilities

Ashcroft notes that a significant defense of dignity lies in establishing the link between dignity and the exercise of certain human capabilities, functions, or social competences. This approach also seeks to confront and avoid the implicit speciesism of metaphysical approaches. As developed by Martha Nussbaum, a capability approach to dignity may best be understood as an account of how an appropriate conception of "dignity" grounds a just political order. As such, it does not necessarily address the issue interpersonal patient care, but it will say something about the rights to health care and, in the concerns of the new bioethics, the regulation of medical technologies (including drugs and organ transplantation).

In Nussbaum's account of human dignity, the metaphysical essence or soul of the human being, which serves as the ground of dignity for Kass, Kant, and earlier traditions, is replaced by the human's empirically identifiable capacities for activity and striving (Nussbaum 2008, p. 249). Capabilities are the "undeveloped powers of the person that [are] the basic conditions for living a life worthy of human dignity" (p. 252). A series of such capabilities are attributed to humans (pp. 261–262), including the capacities to live; to enjoy bodily health and bodily integrity; to use the senses, imagination, and thought; to experience emotions and emotional attachments; to live with others in human society; to live to other species; to play; and to control one's environment. This list, Nussbaum claims, has validity independently of any particular world view or comprehensive doctrine (such as that of Kass's Judeo-Christian tradition).

Nussbaum is at pains to stress that the presence of no one capability is necessary for dignity, albeit that it is assumed that all humans have these capabilities and that there is a political imperative to develop them. Here, her argument at once resists Macklin's reductionism, and more importantly the danger of denying dignity, and thus moral status, to those humans who may lack say reason or autonomy (2008, p. 252). Nevertheless, the capacity to use practical reason and thus "to form a conception of the good and to engage in critical reflection about the planning of one's life" (p. 262) is important. This hints at something akin to Kass's transcendence, but it also entails that threats to dignity lie in actions and environments that inhibit the development and realization of a capability. Thus, to be denied the health care that one needs to live, or the health education that one needs in order to live well, is to have one's dignity undermined.

The capabilities approach to dignity has two important implications. Firstly, Nussbaum is willing to extend dignity to nonhuman animals who exhibit the appropriate capabilities. This, as she notes, has important implications for the use of animals in medical research (2008, p 255). However, a consequence of the centrality of striving to this account of dignity entails that "we would not accord equal human dignity to a person in a persistent vegetative state, or an anencephalic child, since it would appear that there is no striving there, no reaching out for functioning" (p. 252). While Nussbaum is critical of simple equations between the capacities of an adult chimpanzee and a human baby (for ultimately the human baby will develop and will do so within a human world), there is seemingly an implication

that dignity is not inalienable (despite Nussbaum's claims to the contrary). For the marginal case, such as the patient with a chronic disorder of consciousness and who has thus irrevocably lost the capacity to form a conception of the good, dignity is seemingly irrelevant.

Aesthetic Dignity

The problem of marginal patients, and thus those who seemingly lack a core grounding to dignity, may be addressed by theories that argue for the irreducibility of "dignity" to a single ground. In such a defense, different uses of the term "dignity" may be identified, often expressed in binary oppositions such as intrinsic-extrinsic, objective-subjective, public-private, inherent-achieved, and descriptive-prescriptive (see Jacobson 2007, p. 293). The importance of the language of dignity may then, in part, lie in the very fluidity of the term and the subtle way in which it can shift to embrace difficult cases and contexts. The recognition of tensions between uses is not then taken as an indication of the redundancy of the term, but potentially as a recognition that the appeal to dignity may still be valid even if one of the relevant moral properties is absent.

Pullman defends the use of dignity with reference to those who lack autonomy by drawing a distinction between moral and aesthetic dignity (2002). Pullman begins with the problem of suffering, and the possibility that the experience of pain may be so intense that the victim no longer wants to go on living. Pain or disability is so grievous that it has stripped the sufferer of autonomy, and thus dignity renders life pointless. He suggests that a moral of dignity that focuses exclusively on autonomy may lead to such a conclusion. Life is no longer worth living, and euthanasia is thereby supported in the name of dignity. Pullman holds such cases reveal that an exclusive emphasis on dignity in autonomy is insufficient. Such an approach confuses the Kantian noumenal account of dignity with an empirical account. As noted above, the assertion of noumenal dignity represents a call to respect the moral value of all humans, regardless of their actual capacities. It is the reduction of dignity to the empirically identifiable capacity to exercise autonomy that potentially allows for the removal of moral concern from those without autonomy.

If autonomy is the only worthwhile value, Pullman argues, then one cannot tell a meaningful story about a person without autonomy. However, by drawing on the material, cultural, and emotional resources of people around them, even those without autonomy can continue to tell "beautiful" stories of their lives, and thus continue, with dignity, in the face of pain and disability. Thus, for Pullman, the exclusive focus on autonomy does not merely neglect the fact that a dignified life is a meaningful one but also that such exclusivity treats the human being as an isolated individual, stripped from the society and culture within which they live, and upon which they depend more fundamentally than upon their autonomy, for their dignity. Aesthetic, as opposed to Kantian moral, dignity rests in the individual's ability to tell a meaningful or beautiful story about their lives. As such, he may be seen to appeal

back to the Stoic tradition of virtue as the ability to live well in the face of suffering. Moral dignity, in the Kantian sense, retains its relevance but does so in the moral obligation that those with autonomy have to those without autonomy. This is an obligation to aid, not necessarily in restoring autonomy, but in making the non-autonomous life beautiful and thus aesthetically dignified.

The aesthetic interpretation of dignity is explored by others. While Pullman focuses on narrative, others turn more overtly to the issue of appearance. Thus, Pols makes a distinction between dignity as "humanitas" and as "dignitas." While the former is substantially equivalent to Pullman's basic dignity and thus the recognition of the universal moral status of humans, "dignitas" is an "engagement with aesthetic values" (Pols 2013, p. 953). As such, it highlights the importance of how the individual appears to others and how they internalize that judgment. Dignified care thus ensures that patients look good. Kolnai (1976) similarly sees the concept of "dignity" (and here in distinction from the more qualified "human dignity") as embracing esthetic as well as moral qualities. He analyzes the particular and concrete features that strike us as eminently dignified in three classes: first, there are the qualities of composure, calmness, and restraint; second, there are the qualities of distinctness, delimitation, and distance; third, there is self-contained serenity. Here, dignity may be taken to embrace the behavior through which the person presents themselves to others, as well as, in terms of morality, the virtuous action.

Empowerment and Restraint

A problem that may be seen to be implicit to approaches such as Pullman's is that they acknowledge that different definitions of dignity may lead to different moral conclusions. If the concept of dignity is not to be rejected, such ambiguities require careful clarification, indicating how "dignity" may be doing different, and indeed contradictory, work in its application by different thinkers, but with the further requirement that different meanings need to be structured or mapped, thereby indicating which meaning ought to be dominant in any given context.

Beyleveld and Brownsword (2001) may be seen to offer a defense of dignity that works by drawing a distinction between two conceptions. They distinguish between dignity of empowerment and dignity of restraint. Here, it is the case of Manuel Wackenheim and the dignity of dwarf throwing poses a core challenge. Dignity of empowerment broadly defends autonomy (or more precisely agency) and thus the right of Wackenheim to choose his occupation, while dignity of restraint defends communally agreed-upon models of dignified behavior. Recognition of the fundamentally different grounding of these two conceptions of "dignity" begins to explain why appeals to dignity by different theorists can lead to radically different normative conclusions, but also why such tensions will and should arise.

While Beyleveld and Brownsword's account of dignity draws more heavily upon empowerment than restraint, it recognizes that neither is adequate on its own. In the context of the emergence of a new bioethics, and thus a series of problems that arise

out of new technologies, including transplant technology, genetic screening, and the technologies that extend the final period of life, they argue that the old bioethics, grounded in autonomy and informed consent, and thus empowerment, oversimplify the problem. Certain free choices by agents may violate important communally held values, and this must be respected. It is precisely this communal understands of what it is to be dignified that dignity of restraint embraces.

Beyleveld and Brownsword draw their own account of dignity from the tradition of Kant and more immediately from the work of Alan Gewirth. Gewirth's political philosophy argues that, in recognizing ourselves as agents, we acquire an obligation, at pain of self-contradiction, of recognizing and respecting the agency of others. This "Principle of Generic Consistency" effectively replaces Kant's summon bonum and the exultation to treat others always as ends and never merely as means. This further implies that Beyleveld and Brownsword place agency, rather than autonomy, at the ground of their ethical theory. While this might seem then to justify a dignity of empowerment, Beyleveld and Brownsword more subtly argue that dignity is a virtue. The precise nature of the virtue is explicated, initially, by noting the importance that Beyleveld and Brownsword place upon the vulnerability of the agent. Perhaps in contrast to certain accounts of autonomy, agency is fragile. On the one hand, humans may fail to act morally, and on the other, they act in the face of existential anxiety over their own inevitable death. Thus, at one level the virtue of dignity lies in the capacity to maintain a certain quality of behavior in the face of the contingencies of everyday life. As such, their account of dignity reflects something of Stoicism and of Kolnai's analysis. Yet further, if humans are recognized to be fundamentally social beings, as say Pullman argues, then, in part, this quality of behavior – what it is to be dignified – will be given by society. What dignity means in terms of restraint may vary from culture to culture. While this may be seen as a problem by some (Englehardt 2007), it does allow a step away from the overtly Western presuppositions inherent in the dignity of empowerment, allowing, for example, interpretations of dignity from within other philosophical traditions, such as Confucianism (Tao 2007). In the light of this analysis, a legal and moral judgment presupposes dignity as the virtue of balancing the demands of empowerment, and thus the agency of the individual, and restraint, which is to say the values of the society.

The problem of Manuel Wackenheim, and of more relevance to bioethics and the patient, the question of whether organs should be sold for transplant, the point at which life should be ended, and even the posthuman potential of genetic engineering are resolved through a diligent exercise of that virtue, albeit one that, again in contrast to the old bioethics, must recognize that social values shift over time and vary from culture to culture and thus that the very substance of dignified behavior, defended by the dignity of restraint, will change, not least in the light of virtuous debate.

Beyleveld and Brownsword are potentially left with the problem of the dignity of those lacking autonomy, or agency, not least in so far as those without agency seem unable to enter into a reciprocal relationship with the agent. Their solution is derived

from recognizing the social nature of humanity and its expression in the dignity of restraint. The virtue of dignity embraces and is expressed in respect for those lacking agency and by continuing to include them within the community. This is further underpinned by a precautionary principle. Given that one can never be sure that agency is lacking, say in patients with a chronic disorder of conscious, a precautionary principle, holding that if in doubt act as if agency is still present, entails that little will be lost if one is wrong, but much potentially gained if agency really is present.

Varieties of Dignity

Pullman, Pols, Kolnai, and Beyleveld and Brownsword have been taken as representatives of those many philosophers who seek to defend the use of dignity through a recognition of the complexity of uses of the term. The untheorized appeals to dignity found in policy documents can thereby be untangled, their arguments reconstructed, and their validity assessed. Lennart Nordenfelt similarly analyzes four distinct, but core, meanings of "dignity." These he maps according to their interrelationships and mutual dependency (2004: and see 2014).

Nordenfelt begins from the tension between attributed or extrinsic conceptions of dignity and the idea of dignity being intrinsic. This is, in effect, to begin from Cicero and the Stoic accounts of dignity. The moral conception of dignity, as something intrinsic to all humans, is pitted against *dignitas*, as socially bestowed or achieved status. Nordenfelt's point is that this contrast highlights a distinction between conceptions of dignity that may be a matter of degree (so that one can have more or less dignity) and dignity that is an absolute. Four types of dignity are proposed: the dignity of merit, the dignity of moral or existential stature, the dignity of identity, and the universal human dignity (*Menschenwürde*). These may be explained in turn.

Dignity of merit encompasses perhaps the most obvious examples of dignity that is bestowed by others. A person who has a rank or holds an office that entails a set of rights has a special dignity. This is probably the oldest sense of the Latin *dignitas*, which was used for referring to excellence and distinction, properties typically pertaining to senators and other people of high rank in the Roman republic and the later empire. This is a sense that is still flourishing in the Romance languages. The Spanish *dignidad* can refer to a person of a high rank, in particular in the clerical hierarchy, such as an archbishop. Similarly, this is the root of the English "dignitary." Such dignity entails that the office, rather than specifically the person holding the office, has a (perhaps legally enshrined) right to respect, although dignity of merit may also refer to dignity that comes through the individual performance (e.g., by a politician, soldier, or artist). In either case, dignity may be a matter of degree. The office of bishop has more dignity than that of the priest, and the war hero more than the common soldier.

Such a conception of dignity may appear to have little relevance to health care, beyond perhaps reference to hierarchies within the medical professions. Yet, given

that people typically derive a core sense of their dignity and self-respect from the social roles that they occupy, be these within the paid economy or through study, maintaining a household or informal caring, and given that illness typically undermines one's capacity to continue in such roles, the very fact of being ill may be understood as a challenge to a subjectively experienced sense of dignity. This may be particularly significant with chronic illness, where the opportunity to return to a previous social role may be permanently deferred. Health care may thus have to find substitutes for the patient's previous source of dignity of merit and to manage the transition into the new patient role.

Dignity as moral stature appeals to a sense of dignified conduct. A person can act with dignity, where dignity rests upon the deportment with which one acts and makes demands upon others. More significantly, one's sense of dignity and self-respect may be intimately entwined with one's ability to realize and act according to one's own moral values and principles. Here, crucially, there are no formal rights associated with the claim to dignity, as there are for dignity of merit, but rather a recognition of one's status by the relevant community to which one belongs and from which one seeks affirmation. Within health care, dignity of moral stature perhaps relates less to the behavior of patients (for it may be argued that patients have a right to care regardless of their moral stature). Rather, the dignity of carers may be threatened by working conditions that inhibit their ability to live according to their moral principles (e.g., if compromises over care have to be made due to shortages of resources).

In contrast, the dignity of identity is a type of dignity that may be particularly crucial in the analysis of many situations and problems in health care. This is the dignity that people attach to themselves (not necessarily consciously) as integrated and autonomous persons, persons with a history and a future, with all their relationships to other human beings. As such, the dignity of identity emphasizes the social nature of human beings, found, for example, in Pullman's arguments. Most people have a basic respect for their own identity, although it need not be at all remarkable from a moral or other point of view. But this self-respect can easily be shattered, for instance, by nature itself, in illness and the disability of illness and old age but also by the cruel or humiliating acts of other people. It is often the dignity of identity that one refers to when one claims that someone has lost his or her dignity. If the culture within which the patient lives denies the patient the resources through which they can construct a coherent and positive (or as Pullman would argue, "beautiful") narrative about themselves and their lives, then their dignity is threatened. Culturally ingrained prejudices, such as ageism, and discrimination against disabilities or mental illness undermine dignity, precisely in that they inhibit the person's positive articulation of themselves and thus sense of self-respect, as an older person, a person with disability or mental health problems.

The dignity of identity need not just be seen as a feeling or sense of worthiness, although it may express itself as a feeling or sense. The factors that ground the dignity of identity are the subject's integrity and autonomy, including his or her social relations. These factors are typically associated with a sense of integrity and autonomy.

When a person's integrity and autonomy are tampered with, this is typically associated with a feeling of humiliation or loss of self-respect on his or her part. Self-respect is thus an important concept in connection with the dignity of identity.

The three types of dignity introduced so far are quite different but have two important features in common. Firstly, people can have these types of dignity to various extents. Second, and consequently, all the three dignities can come and go. A fourth type of dignity, universal dignity or *Menschenwürde*, is thus proposed by Nordenfelt as a necessary complement. If dignity is merely subjectively experienced or bestowed by a relevant community, then the lack of subjectively felt self-respect or the withdraw of respect by others renders the patient morally vulnerable, as has already been argued above. Those incapable of asserting their dignity, including patients with chronic disorders of consciousness, or those who are the victims of prejudice, such as the mentally ill, may seem to have their moral status undermined. Nordenfelt therefore proposes a universal and inalienable dignity that all humans have to the same, absolute, degree. Such dignity, which has been seen above in the metaphysical accounts of Kant and Kass, cannot be taken from the human being. Such a claim to dignity, which is precisely the sense enshrines, for example, in the Universal Declaration, is a moral bulwark against discrimination and prejudice.

While Nordenfelt's model of the meanings of dignity serves effectively to disentangle different uses and crucially to highlight the problem of any reliance on a purely subjective or culturally relative and thus empirically experienced sense of dignity, it encounters the problem, noted in all metaphysical accounts, of failing to ground *Menschwürde* on any stronger grounds than the mere assertion of being human. Edgar (2003, 2004) has offered a partial defense in terms of something akin to a capabilities approach. He argues that a fully competent human will have developed certain competencies (the autonomous control of their bodies, the capacity to communicate, and skills in social interaction) that no other fully mature animal will have. The uniquely human potential to develop and lose such core competences grounds the claim to dignity. To lose any of these competencies, or to have any disregarded by other members of society (e.g., when a healthy carer or companion is spoken to and not the patient him- or herself), is to have one's dignity affronted. The demand of *Menschwürde* is thus to continue to respect and sustain the dignity of the patient, even in the face of the temporary or permanent loss of competence, for it is precisely in the absence of the core competence (even if that competence has never had the opportunity to develop) that we perceive the demand to be treated with dignity.

Conclusion

While some philosophers question the relevance of "dignity" as a morally concept, its continued use by both patients and their advocates, not least when challenging poor standards of care, is indicative of its continuing usefulness. It may be concluded that the very complexity and subtlety of the concept gives it a uniquely powerful

place in moral discourse about contemporary health care. As the survey data, review at the beginning of this chapter, indicates, dignity language is frequently used in protest over problems arising from a failure to understand the needs of patients or failures arising from underfunding of care. Further, as new challenges arise, not merely from new technologies, but more fundamentally from the increasing prevalence of chronic illness and increasing numbers of older patients and patients with dementia or chronic disorders of consciousness and thus an increase in patients whose personal sense of dignity is potentially undermined or who can no longer autonomously assert their dignity, it is precisely dignity, and not autonomy or informed consent, that must be central to any discussion of the ethics of health care.

Definition of Key Terms

Aesthetic dignity	A person's sense of dignity and self-esteem may lie in the way in which they present themselves to others (and thus in matters of physical appearance or dress) or in their capacity to construct positive and constructive narratives about themselves.
Dignity of constraint	A term proposed by Beyleveld and Brownsword to characterize the use of "dignity" to justify restricting the autonomy of individuals, in order to prevent them from committing action that damage their own or other people's dignity.
Dignity of empowerment	A term proposed by Beyleveld and Brownsword to characterize the use of "dignity" to enhance the autonomy and free choice of individuals.
Extrinsic dignity	Extrinsic dignities are dependent for their existence on some external fact, typically a human action. The most salient extrinsic dignities are such ranks and positions as are the result from nominations.
Human dignity	All people have this dignity to the same degree, i.e., people are equal with respect to this kind of dignity. It is significant that human dignity cannot be taken from the human being as long as he or she is alive.
Intrinsic dignity	The intrinsic dignity or the intrinsic value of a person defines the nature of the person and is identical with human dignity.
Metaphysical dignity	Also termed "Menschenwürde," this is a source of human dignity that is not reducible to any empirically observable properties or capacities of the human being or to the social ascribed moral worth of the individual.

Summary Points

- The term "dignity" is used widely by patients, carers, and medical professionals and has generated a substantial philosophical literature. The precise meaning of the term, and its application in applied ethics, is highly contested.
- Patients assert their dignity typically when that dignity is threatened. Dignity is understood in its absence, when a person feels undignified, humiliated, embarrassed, and vulnerable.
- Dignity may be understood as a subjective experience, linked to a sense of self-respect or self-esteem. A subjective sense of dignity is necessarily complemented by a sense of dignity as a universal and inalienable moral property of human beings, such that humans are worth of being treated with dignity, whether they subjectively recognize this or not.
- A number of philosophers have argued that "dignity," as a concept in bioethics, is redundant and may be reduced to other concepts, such as "autonomy."
- A numbers of strategies have been employed in order to defend the continuing relevance and irreducibility of "dignity" as a moral concept. Dignity may be understood as a metaphysical concept that does not refer to empirical manifest properties but to a moral essence of human beings. Conversely, dignity may be identified with the ability of human beings to exercise various capabilities or functions. Others map the different uses of the terms dignity, arguing that only by understanding this network of complementary uses can "dignity" be used with precision.

References

Arino-Blasco S, Tadd W, Boix-Ferrer JA (2005) Dignity and older people: the voice of professionals. Qual Ageing Policy, Pract Res 6(1):30–36

Ashcroft R (2005) Making sense of dignity. J Med Ethics 31:679–682

Bayer A, Tadd W, Krajicik S (2005) Dignity: The voice of older people. Qual Ageing Policy, Pract Res 6(1):22–29

Becker GK (2001) In search of humanity: Human dignity as a basic moral attitude. In: Hayry M, Takala T (eds) The future of value inquiry. Rodopi, Amsterdam, pp 53–65

Beyleveld D, Brownsword R (2001) Human dignity in bioethics and biolaw. Oxford University Press, Oxford

Bostrom N (2005) In defence of posthuman dignity. Bioethics 19(3):202–214

Calnan M (2005) Views on dignity in providing care for older people. Nurs Times 101(33):38–41

Cicero MT (1913) De officiis (trans: Miller W). Harvard University Press, Cambridge, MA

Cochrane A (2010) Undignified bioethics. Bioethics 24(5):234–241

Care Quality Commission (2011) Dignity and nutrition inspection programme: national overview. Care Quality Commission, London

Edgar A (2003) Velázquez and the representation of dignity. Med Health Care Philos 6:111–121

Edgar A (2004) A response to Nordenfelt's 'The varieties of dignity'. Health Care Anal 12 (2):83–89

Englehardt HT (2007) Long-term care: the family, post-modernity, and conflicting moral life-worlds. J Med Philos 32(5):519–536

Franklin L-L, Ternesdedt L, Nordenfelt L (2006) Views of dignity of elderly nursing home residents. Nurs Ethics 13(2):130–146
Harris J (1998) Cloning and human dignity. Camb Q Healthc Ethics 7(2):163–167
Jacobson N (2007) Dignity and health: a review. Soc Sci Med 64:292–302
Kant I (2002) [1785] Groundwork for the metaphysics of morals (trans: Wood AW). Yale University Press, New Haven
Kass LR (2002) Life, liberty and the defense of dignity. Encounter Books, San Francisco
Kass LR (2008) Defending human dignity. In: Lanigan BT (ed) Human dignity and bioethics. Nova Science, New York
Kaufmann P, Kuch H, Neuhäuser C, Webster E (eds) (2011) Humiliation, degradation, dehumanization: human dignity violated. Springer, New York
Kolnai A (1976) Dignity. Philosophy 51:251–271
Lanigan BT (ed) (2008) Human dignity and bioethics. Nova Science, New York
Lantz G (2009) Dignity and the dead. In: Nordenfelt L (ed) Dignity in care for older people. Blackwell, Oxford, pp 168–189
Leget C (2013) Analyzing dignity: a perspective from the ethics of care. Med Health Care Philos 16 (4):945–952
Macklin R (2003) Dignity is a useless concept. Br Med J 327:1419–1420
Mann J (1998) Dignity and health: the UDHR's revolutionary first article. Health Hum Right 3 (2):30–38
Nordenfelt L (2004) The varieties of dignity. Health Care Anal 12(2):69–81
Nordenfelt L (2014) Dignity and dementia: a conceptual exploration. In: Hydén LC, Lindemann H, Brockmeier J (eds) Beyond Loss: dementia, identity. Oxford University Press, Personhood
Nussbaum M (2008) Human dignity and political entitlements. In: Lanigan BT (ed) Human dignity and bioethics. Nova Science, New York
Pico della Mirandola G (1985 [1486]) The dignity of man. MacMillan, Basingstoke
Pols M (2013) Through the looking glass: good looks and dignity in care. Medicine, Health Care and Philosophy 16(4): 953–966
Pullman D (2002) Human dignity and the ethics and aesthetics of pain and suffering. Theor Med 23:75–94
Rachels J (1990) Created from animals: the moral implications of darwinism. Oxford University Press, New York
Schulman A (2008) Bioethics and the question of human dignity. In: Lanigan BT (ed) Human dignity and bioethics. Nova, New York, pp 3–13
Shotton L, Seedhouse D (1998) Practical dignity in caring. Nurs Ethics 5(3):246–255
Stratton D, Tadd W (2005) Dignity and older people: the voice of society. Qual Ageing Policy Pract Res 6(1):37–45
Tadd W, Calman M (2009) Caring for older people: why dignity matters – the European experience. In: Nordenfelt L (ed) Dignity in care for older people. Wiley-Blackwell, Oxford, pp 119–142
Tao J (2007) Dignity in long-term care for older persons: a Confucian perspective. J Med Philos 32 (5):465–481
United Nations (1948) Universal declaration of human rights. UN, Geneva
Varelius J (2009) Minimally conscious state and human dignity. Neuroethics 2(1):35–50

The Living Body and the Lived Body in the Clinical Encounter: How Does the Body Shape Ethical Practice

30

Dorothée Legrand

Contents

Abstract

How does the body shape the clinical encounter? In contemporary debates, evidence-based medicine is thought to favor exclusively the body as object: palpated by the clinician's hands, submitted to a laser, and captured in a medical image. Such approach to medicine has been challenged by alternative conceptions and practices, notably grounded on a phenomenological characterization of the body as distinctively subjective. Among such alternatives, narrative approaches to illness and medicine encourage, on the patient's side, a self-empowerment through the mastery of one's own illness story and, on the clinician's side, an empathic attitude allowing the interpretation of such narratives. In turn, such approaches are challenged by the consideration of the patient's and the clinician's relation to

D. Legrand (✉)
Archives Husserl, CNRS, Ecole normale supérieure, Paris Sciences et Lettres Research University, Paris, France
e-mail: dorothee.legrand@ens.fr

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_66

sufferance. Before the clinician can take the patient's sufferance as an object of scrutiny, at an intentional level, he may be experiencing it, at a pre-intentional level, through his own body, thanks to a bodily intersubjectivity in which one body resonates with the other. Moreover, also at a pre-intentional level, the patient's suffering is irreducible to any resonating affect in the clinician's body. As such, to suffer is to be passively subjected to self-entrapment. Only the encounter with another subject can open and thus alleviate such suffering, when the patient addresses it to the clinician as a call for recognition and when the clinician assumes the responsibility to give hospitality and to respond to such suffering. Such is the responsive stance held in the clinical encounter.

Introduction: Toward a Philosophy of the Clinical Encounter

Who does benefit from the interaction between philosophy and medicine? What could one expect from bringing together philosophy and medicine? Are we expecting an epistemic gain? Or rather a clinical one? Or both at once? Are we expecting that an epistemic gain would lead to a clinical one?

The philosophy *of* medicine may be notably concerned with *the clinicians' conceptual apparatus*, and their modes of operating with these concepts in their practice, to diagnose a pathology and determine the most appropriate treatment accordingly. The philosophy of medicine may here be conceived of as an epistemology prescribed from the perspective of an "outsider," i.e., from the perspective of a philosopher taking a step outside of illness, holding a stance that is foreign to both the patient and the clinician, applying philosophical concepts to a situation which he does not belong to, and relative to which he can keep the sanitized distance thought to be required for conceptual clarification.

The clinicians, however, cannot practice medicine without the patients. Thus, *the patients' experience* ought to be taken into account with precision and in its specificity, not only in the practice of medicine but also, and consequently, in the philosophy of medicine.

Moreover, in any case, medicine is first and foremost a *practice*, which cannot be conceived of without the encounter of a patient with a clinician. The philosophy of medicine, thus, must also be a philosophy of this practice, i.e., a *philosophy of the encounter* of a clinician with a patient. Here, philosophy *and* medicine can be articulated to each other as joint partners in the investigation of the conditions under which can occur the clinical encounter between clinicians and patients.

This is what the following investigation aims at contributing to. The reason for choosing such focus is that clinical practice starts neither with the clinician's mastery of some appropriate knowledge nor with the illness or pain the patient suffers from; rather any sort of clinical *practice*, be it technoscientific or tribal, starts and can only start from the encounter. To initiate our reflection on that matter, it must be underlined that the clinical encounter cannot be understood, and performed, without considering how it is an encounter of both the patient and the clinician with a

suffering body, a body that is both at once *lived* and *living* and, as such, vulnerable to pain and death. Simply stated, the question that will guide the forthcoming reflections is thus the following: *how does the body shape the clinical encounter*?

The Body as Subject and as Object

The investigation of the role the body may play in the clinical encounter benefits from conceptual distinctions offered by the phenomenological characterization of the structure of *intentional consciousness* and, in particular, the distinction between two inseparable poles of experience, a subjective pole and an objective pole (Legrand 2011, 2012). Schematically, one may simplistically represent intentional consciousness as an arrow, with its *subject* as its starting point and its *object* as its ending point. This simplistic schema suffices here to represent several important ideas:

- First, without its starting and ending points, there would simply be no way to draw the arrow, which schematizes that there is no intentional consciousness without both at once its subject and object: *subject and object of consciousness are inseparable from each other.*
- Moreover, again at the structural level, the subject at the starting point will never become the object at the ending point. This suggests that *the subject cannot be reduced to an object of experience*.

In the particular case of bodily consciousness, even the familiar experience of looking at one's image in a mirror involves experiencing oneself at once both as subject and as object. One experiences oneself as an *experiencing subject*, the one who is looking from "here" over "there" where one concurrently experiences oneself as an *experienced object*, as the one whose image is looked at.

In the situation of a clinical encounter, one may experience one's body as a *subject* giving to one's world a pervasively painful coloration, as a subject located "here," looking "over there," but unable to move "over there" because of, say, an excruciating back pain; and one may experience one's body as an *object* under one's own scrutiny, exteroceptively or interoceptively, an object palpated by the clinician's hands, an object submitted to a laser, and an object captured in a medical image.

While the whole history of Western medicine has been animated by oppositions between "the disease-as-scientifically-constructed" and the "illness-as-lived" (Zaner 1992, p. 111), in contemporary debates, evidence-based medicine is thought to favor exclusively *the body as object* and is challenged by alternative conceptions and practices, notably grounded on a phenomenological characterization of *the body as distinctively subjective*.

Among such alternatives, on the basis of a *narrative approach to medicine*, one may argue that, in the clinical encounter, the body "is not seeable as an object" (Charon 2008b, p. 290); the body is not only conceived of as an *object* of medical

manipulation but also as a *subject* telling a story which adequately trained clinicians can decipher to "increase their clinical effectiveness" (Charon 2008a, p. 26). The ill body resists self-understanding and must be "rescued from formlessness" thanks to an attention paid to its narratives (ibid.).

Additionally, in the framework of a *narrative approach to illness*, it has been argued that "the ill body is certainly not mute – it speaks eloquently in pains and symptoms – but it is inarticulate. We must speak for the body" (Frank 1995, p. 2). The patient's voice must articulate his inarticulate body; he must give his own voice to what his own illness means in his own life.

In the narrative approach to *medicine* and even more so in the narrative approach to *illness*, narration is thought to benefit the patient in that it promises him to keep or regain his sense of being himself, despite or thanks to what happens to him in illness. Indeed, "the performance of narrative [is] integral to the experience of identity" (Eakin 2004, p. 130). Narration would thus be akin to an "art of self-invention" (Eakin 1992, p. 71): "when we talk about ourselves, and even more when we fashion an I-character in an autobiography, we give a degree of permanence and narrative solidity – or body, we might say – to otherwise evanescent states of identity feeling" (Eakin 2004, p. 129). The therapeutic value of narration is thought to be linked to the "teller effect" that grants the narrator with the "psychological gratification" associated with one of the illusions "that fuels autobiography," namely, "the belief in the possibility of self-determination." In particular, narration is a performance by which the narrator avoids two pitfalls at once: the total loss of control that would allow illness to impose itself and the taking control of one's illness by another person, the clinician.

The Ethics of Self-Mastery

What emerges here is the idea that (some) narrative approaches draw on an *ethics of self-mastery*. Stated unambiguously, "the moral imperative of narrative ethics is perpetual self-reflection on the sort of person that one's story is shaping one into, entailing the requirement to change that self-story if the wrong self is being shaped" (Frank 1995, p. 158). This narrative approach to illness thus raises the act of telling one's story to the rank of a "moral imperative," and in so doing, it also promotes the act of constructing oneself, the self-determination of one's experience, by attributing meaning to one's suffering, taking control of the experience of illness, and domesticating the unknown. This practice would have therapeutic virtues, at least inasmuch as it is held to reduce the anxiety associated with the fact of being the object of forces that are uncontrollable, incomprehensible, undescribable, and unsayable: giving a meaning to the illness from which I am suffering gives me a power over it, at the very least the power of not succumbing to the unutterable meaninglessness into which suffering entraps me.

Interestingly, even though narrative-based medicine is designed to counter evidence-based medicine, they both share a reliance on an *ethics of self-mastery*. The latter relies on a "regulating ideal of the self" (Rose 1998, p. 2) which is

epitomized in technoscientific biomedicine but which is also at stake in the narrative approach. "Ethics [is] here understood as a way of understanding, fashioning, and managing ourselves in the everyday conduct of our lives" (Rose 2007, p. 257). This view puts an "increasing emphasis on the responsibility of individuals to manage their own affairs, to secure their own security [...]. Nowhere have these been more telling than in the field of health, where patients are increasingly urged to become active and responsible consumers of medical services and products" (ibid., p. 4). We live "in an age of [...] self maximization" (ibid., p. 8) where "health, understood as an imperative, for the self and for others, to maximize the vital forces and potentialities of the living body, has become a key element in contemporary ethical regimes" (ibid., p. 23).

However, a tension must be underlined here, between, on the one hand, such ethics of self-mastery and, on the other hand, the aforementioned phenomenological distinction between the body as subject and the body as object:

- On the one hand, an ethics of self-mastery focuses on the body as subject anchoring one's subjective experience to one's bodily standpoint, *at the expense of* the body as object experienced at a distance from oneself and imposing itself upon oneself (notably through illness).
- On the other hand, what emerges from within a phenomenological framework is neither a *discredit* of the body as object nor a *reduction* of the body as subject to the introspectively lived body. Rather, a phenomenological framework allows the consideration that the body is *multidimensional* and that it is as such that it participates to the clinical encounter, i.e., as a complex composition of multiple dimensions (subjective and objective) which remain irreducible to each other while being irremediably joint to each other.

Thus, the ethics of the sovereign self, which may be promoted both in humanistic and in technoscientific approaches to medicine and illness, can be reappraised and challenged from within the phenomenologically designed distinction between the body as subject and the body as object (Legrand 2013). Even if one focuses only on the body as lived by the patient, it may indeed be underlined that this "so-called 'lived body' involves experiences of one's body both as subject (*Leib*) and as object (*Körper*)" (Slatman 2014, p. 2). For that reason, "phenomenology in the field of health and medicine should abandon its unilateral criticism on the 'body as object ontology'" (ibid.) and rather admit that the body as object is an unavoidable dimension of the experience of the body, in general and in clinical situations too (Legrand 2015).

Clinical situations are challenging as they come with the fear of the clinician's power to "reinforce," "intensify" instead of "neutralizing [the] strangeness" of one's bodily experience (Slatman 2009, p. 122). But if clinical situations are disturbing, it is *not* because they reduce me to what I am *not*; it is rather because they show me what I am *too*, irremediably: a body as object. The standpoint of the clinician does impact my experience of my body. But it must be underlined that it does not do so by depriving me of my subjectively lived body but rather by revealing something of *me*,

my body as object, which always already, silently and impersonally, participates to my bodily experiences, unbeknownst to me.

As object, the patient's body is never reducible to a mere thing, neither from the standpoint of the patient nor from the standpoint of the clinician. In particular, as object, the patient's body is never reducible to dead matter and medical work would be utterly different if the clinical encounter unveiled the body as inert matter, as if one could attend to the dissection of one's own body when it has become a carcass. What is manipulated, measured, tested, controlled, and pictured by the clinician is not the *dead* body; rather, what is manifest in the clinical encounter is the *lived-living* body, the lived body that suffers and the living body that is not dead, *not yet*.

This latter consideration contrasts with the idea that technoscientific medicine would be based on an "ontology of death" (Leder 1992, p. 21) and would remain "a science or a practice of dead bodies" which "cannot show anything else than our thinghood" (Slatman 2009, p. 110). Contrastively, it must be recognized that the very practice of medicine imposes to take into account processes of healing and decay proper to a body which evolves toward death. It is to this body, *not-dead-yet*, to this body always susceptible to die, that both the patient and the clinician are subjected to (Zaner 1992, p. 105; Gadamer 1998, p. 100).

This dimension of my body is what I would rather "repress": the participation of my *living* body to my *lived* body (Merleau-Ponty 1962), the root of my subjectivity into my biological life, i.e., into my death. But despite this "organic repression," it would be a mistake to neglect how much my "*almost* impersonal" life and death are part of myself (ibid., p. 96).

What thus emerges here is the idea that, more than being a threat to bodily subjectivity, clinical situations may be a threat to the sovereign self, aiming at self-mastery. Operating such destitution of subjective sovereignty, the clinical encounter may participate to the realization that the body as object is distinct but tied to the body as subject (Legrand 2015). As such, it may question what we dare to bare about ourselves: it may question how much we are ready to lift the repression of life and death that unceasingly operates within our bodily subjectivity.

Empowerment

As was proposed just above, (some) narrative approaches to medicine and to illness are tied to an ethics of mastery of the *self*. Nonetheless, in such approaches, *others* are not obliterated. Indeed, if the patient must give his own voice to what his own illness means in his own life, this is nonetheless not meant to be a self-enclosed practice. Indeed, in doing so, the narrator is meant to create "alternative ways of being ill" (Frank 1995, p. 117), narrating "what is possible in impossible situations" (ibid., p. 133), notably in the aim of inspiring fellow-sufferers who may not have the possibility or ability to speak.

Even though experience in general, and even more so the experience of illness, "seems to be something that happens to us" (Sartwell 2000, p. 41) and is thus

"ateleological" in this sense (ibid., p. 16), narrative-based medicine inserts illness into a "teleological order" (ibid., p. 12) in that it seeks to *use* illness as a source of information, insight, understanding, and knowledge. In such teleological perspective, it is defended that, by telling their illness stories, the tellers "accept illness and seek to *use* it. Illness is the occasion of a journey that becomes a quest. [...] the quest is defined by the ill person's belief that something is to be gained through the experience" (Frank 1995, p. 115). Narratives are thus the occasion for realizing "a sense of purpose" (ibid., p. 117), and the purpose of illness is thought to be twofold. On the one hand, illness is "a calling, a vocation" (ibid., p. 166) for "a change of character through suffering" (ibid., p. 128); the tellers "realizing who they always have been, truly been, each becomes or prepares to become the re-created, moral version of that self" (ibid., p. 131). On the other hand, illness also imposes to the sufferer the "responsibility for testimony, and testimony implies risk: dying a messenger's death" (ibid., p. 166).

To elaborate on this latter point, Frank proposes a reading of Levinas' distinction between useless sufferance and a sufferance which acquires a meaning. Frank adapts this distinction to make it fit his own distinction between, on the one hand, "the unassumable, nameless suffering" (ibid., p. 178) and, on the other hand, the "just sufferance" which is the one narrated in "quest stories" where the sufferer is a witness of himself and a testimony for others. Sufferance ceases to be useless, Frank argues, when the sufferer "suffers for others" (ibid., p. 178); contrastively, sufferance "*becomes* useless" (ibid., p. 179, italics added) when it is left at what it is: meaningless. Frank considers how witnessing my own sufferance, paying attention to my own sufferance, calling upon others to require their attention upon my own sufferance, is what makes sufferance useful. Accordingly, it is by calling other's attention upon me that giving voice to my sufferance has the power to overpass its uselessness. What is useful is the sufferance I can narrate for others.

However, this view completely reverses the characterization of sufferance offered by Levinas in the text Franck refers to, i.e. *Useless Suffering* (Levinas 1998). Indeed, for Frank, it is when my sufferance is instrumentalized to reach the other that it ceases to be useless. On the contrary, Levinas insists on the "radical difference" between, on the one hand, others' suffering which is and is bound to remain "unforgivable" and which "solicits me and calls me" and, on the other hand, "my own experience of suffering, whose *constitutional or congenital uselessness* can take on a meaning, the only one of which suffering is capable, in becoming a suffering for the suffering (inexorable though it may be) of someone else" (ibid., p. 94, italics added). For Levinas, it is only "the suffering of suffering, the suffering for the *useless* suffering of the other, the [...] suffering in me for the unjustifiable suffering of the other" that may be said to be "just" (ibid., italics added). My sufferance is "just" when it mobilizes me into "a civilization called upon to feed human beings and to lighten their sufferings" (ibid.). The "inescapable obligation" to give "attention to the suffering of the other" (ibid.) is the one and only way my sufferance may be un-useless.

Two modes of characterizing sufferance can be distinguished here:

- The narrative approach involves a conception of the sufferer as (potentially) mastering his illness story, thereby using his sufferance as a means to become more (ac)knowledgeable and to give a testimony of one's own sufferance to others.
- This view, however, conflicts with the idea that suffering fundamentally confronts one with one's vulnerability, one's lack of control, the loss of one's sovereignty, and with one's responsibility to care for others' sufferance.

Outside of a teleological order and of an ethics of mastery of the sovereign self, how do these latter aspects of illness and sufferance participate to clinical practice? Before addressing this question by considering more in detail *the suffering body* and the manner in which it may shape the clinical encounter, let us stay within the narrative approach to further develop how it may impact the patient, not only in relation to the way he himself *speaks* about his illness story (to others) but now in relation to the way others may *listen* (or not) to the narrative of his illness.

Empathy

Illness stories are inevitably told within a context shaped and shared with others (Legrand 2013), a context which, currently, is growingly governed by what has been coined a "narrative imperialism" (Phelan 2005, p. 206), i.e., the idea that everyone lives and should live narrative lives and align with the narrative structure each and every aspect of all of one's experiences. Moreover, a major "conventional expectation of any narrative, held alike by listeners and storytellers, is for a past that leads into a present that sets in place a foreseeable future" (Frank 1995, p. 55). Given this requirement, inevitably, "the illness story is wrecked because its present is not what the past was supposed to lead up to, and the future is scarcely thinkable" (ibid.). Thus, in a narrative context, wrecked stories of illness and, even more, the chaotic words of helplessness uttered in sufferance "erect a wall around the teller" (Frank 1995, p. 102). Not only sufferance captures the person into "the claustrophobic terror of [. . .] muteness" (ibid., p. 109), but once one finds a voice to "speak *about* the chaos, from *outside* that chaos" (ibid., p. 190), no one is there to listen (ibid., p. 101) because what is uttered breaks the rules of narration. The lack of space the narrative approach leaves to unnarratable sufferance "only makes its horror worse" (ibid., p. 112).

It appears here that, when it becomes the *norm*, the narrative approach may become a burden: "identity narratives, delivered piecemeal every day, function as the signature for others of the individual's possession of a normal identity"; but "the verdict of those for whom we perform is virtually axiomatic: no satisfactory narrative, no self" (Eakin 2006, p. 182; see also Eakin 2001, p. 120). *Normally*, one shall not break the rules of narratives: "these rules are tacit because the daily performance

of identity story is instinctive and automatic" (ibid., p. 113), but failure to cope with these rules "may entail institutional confinement" (Eakin 2006, p. 182); "de-storied individuals" (Eakin 2001, p. 121) are rejected out of narrative normalcy by others. The call for narratives may be heard as a duty to make one's story fit into a narrative format which is audible enough to allow others to understand what is at stake: a narrative context "may be a burden on those ill and disabled people who do not fit or do not wish to fit into that script" (Garden 2010, p. 131).

On the basis of such considerations, what emerges is the idea that the call for illness narratives does not only give responsibility to the patients to tell their illness story, it also gives responsibility to others in general, and to clinicians in particular, to receive these narratives. More in particular, the narrative approach does not only advocate an *empowerment* of the patients which relies on an ethics of self-mastery, as it was proposed above; moreover, the call for illness narratives comes with a call for *empathy*, as shall be further detailed now.

The responsibility of others is clearly assumed in the narrative approach to *medicine*, developed alongside the narrative approach to *illness*. Narrative medicine has first been defined as "medicine practiced with narrative competence" (Charon 2001, p. 1897). That is, aside from approaches focusing on the narration of the patient himself, narrative medicine cultivates the act of listening and interpreting the medical doctor and emphasizes *his* narrative competence which involves "the ability to acknowledge, absorb, interpret, and act on the stories and plights of others" (ibid.), the skill to achieve a position that is "relaxed, absorbing, accepting, oceanic, filling" (Charon 2008a, p. 23).

Empathy is a key ingredient here. As Macnaughton underlines, "the 'practice of empathy' has become an icon of the growing medical humanities movement in the USA and the UK. US physicians have even gone so far as to adopt empathy as one of the accredited "skills" required by the American Council for Graduate Education" (2009, p. 190). She argues, however, that genuine empathy is neither possible nor advisable in the practice of medicine: what the clinician experiences "is so different from what the patient is feeling that it seems disrespectful to suggest that I somehow participate in his or her experience. [...] what we maintain is [...] the need to respond." A critical stance is also held by Garden who underlines that "the problem of empathy begins with the preoccupation with self that obscures the other. Empathy depends on the experiences and imagination of the person who is empathizing, and this dependency has the potential to obfuscate or exclude the patient's suffering and the meaning the patient makes of suffering" (2007, p. 555).

Empathy, conceived as such, is precisely what humanistic approaches to medicine and illness are meant to avoid. Thus, from within the phenomenological approach, and in the aim of reconsidering the relation that may be at stake between one and another in the clinical encounter, the very term "empathy," notoriously hard to define, may be avoided all together. With this terminological debate come important issues to be considered here, in the aim of better understanding the relation to the (ill) body which the clinical encounter entails.

Intercorporeality

David Kleinberg-Levin argues that "'sympathy' is a better word than 'empathy,' because the latter word, implying the logic of an inside and an outside, is inextricably entangled in a web of metaphysical assumptions – assumptions, in particular, about the subjectivity of the subject [. . .]" assumptions which would first posit the subject in isolation from others, and then impose to "articulate a magical escape from its solipsism"; by contrast, sympathy would rather be rooted in "intercorporeality" (Kleinberg-Levin 1999, p. 89). Intercorporeality, Kleinberg-Levin argues, on the basis of his reading of Merleau-Ponty, "transcends the subject-object structure" of intentional consciousness (ibid., p. 71) and allows us to conceive of intersubjectivity as built in the very structure of our body. As intercorporeal, our bodies are pre-intentionally structured intersubjectively, and in this sense the other always already "resonates" in me, no matter the particular stance I take on him via my intentional experiences of him. Conceiving of our "corporeal intersubjectivity" requires the "effacement of the omnipotent ego-logical subject" (ibid., p. 67), as well as it requires withdrawing from the perspective taken on the other as an object of intentional consciousness; intercorporeality rather "introduces substitutions and reversibilities that are radically decentering" (ibid., p. 69), as it is "a symmetrical relation, in fact, a relation of transpositions, reciprocities and reversibilities" between oneself and the other (Levin 2008, p. 23).

Following a comparable line of thought, Thomas Csordas differentiates two notions of intersubjectivity:

- On the one hand, a concept of intersubjectivity refers to "a relation between two isolated subjectivities or intentionalities" and "presupposes a Cartesian formulation of subjectivity, from which starting point intersubjectivity could only be construed as mutual representation of each subjectivity by the other" (1994, p. 285, note 13).
- On the other hand, intersubjectivity involves "the copresence of another myself," rooted in "an interweaving of familiar patterns of behavior" (ibid., p. 12).

Csordas rejects the former and espouses the latter conception of intersubjectivity. This provides him the relevant tool to qualify *healing processes* at stake in various sociocultural situations. If "we are not isolated subjectivities trapped within our bodies, but share an intersubjective milieu with others," then there must exist "a somatic mode of attention [which] means not only attention to and with one's own body, but [which] includes attention to the bodies of others" (Csordas 1993, p. 139). Under the heading "somatic modes of attention," diverse phenomena are gathered, among which:

- "One healer distinguished clearly that when the problem is internal, she typically 'sees' the organ, or cancer, appearing as a black mass, but when the problem is external, she typically 'hears' the word naming the illness or the body part, such as arms and legs" (ibid., p. 141).

- Csordas's "use of [his] own experience as an occasion for data collection," for example, when his "own crossed leg jumped as if it had been tapped by a doctor's hammer in a test of reflexes" during his observation of a patient (ibid., p. 145).

In such cases, there is at stake, Csordas argues, a somatic mode of attention which participates to the encounter between oneself and another person and thus participates to the healing practice. Such co-somatic, intercorporeal experiences operate an "intersubjective constitution of meaning" (Csordas 1994, p. 141) from which the healing practice cannot be severed. According to this view, the bodily state of another subject is experienced in resonance with one's own, immediately, "preobjectively" (Csordas 1994, p. 13), *pre-intentionally*, i.e., before the other is taken as the object of my perception, observation, or manipulation.

Likewise, Kleinberg-Levin emphasizes that corporeal intersubjectivity occurs "prior to every voluntarism" (Kleinberg-Levin 1999, p. 67). For Merleau-Ponty, indeed, our relationship to others is "deeper than any express perception or any judgment [...] We must return to the social with which we are in contact by the mere fact of existing, and which we carry about inseparably with us before any objectification" (Merleau-Ponty 1962, p. 362). Prior to any identification and appropriation of my experiences to myself, before any separation of myself as a subject singularly different from others, there is an "almost impersonal" layer of my existence (ibid.), where oneself and others find each other forming a community older than individuality (Kleinberg-Levin 1999, pp. 79–80).

Intercorporeal Ethics

Merleau-Ponty's notion of prepersonal intercorporeality, as read by Kleinberg-Levin, opens "the possibility of a reconciliation between the sensible and the ethical" (ibid., p. 68). In his view,

- There would be "a prelinguistic dimension or moment of embodied experience [...] prior to volition, prior to consciousness, and thus prior to (egological) memory, [where] the ethical responsibility to and for the other first takes hold" (Kleinberg-Levin 2008, p. 21).
- However, we would suffer from an oblivion of the "heteronomic body of pre-linguistically felt experience" (ibid.). Even though we are born in "intertwinings and reversibilities" with others, this does not mean that "ethical comportment is somehow predetermined to emerge automatically, directly and immediately" from this prepersonal layer of our existence; rather "evil is always a possibility because [intertwinings and reversibilities are] only one dimension of our embodiment, our bodily existence, and because, for the most part, we do not live and speak in contact with, and from out of, that dimension, but live and speak, instead, from out of the ego-logical structure constituted through the normal processes of socialisation" (Kleinberg-Levin 1999, p. 85).

- Nonetheless, redemption is possible, and it "depends on the remembrance, the recovery or restitution, of a forgotten intercorporeal attunement" (Kleinberg-Levin 2008, p. 39). Kleinberg-Levin denounces the "violence inherent in the ego's speech" (Kleinberg-Levin 1999, p. 86) and militates in favor of "self-development" and "self-realization" thanks to the "recuperation," the "recreation" of our intercorporeal prepersonal experience. Such return to prepersonal intercorporeality is "ethically imperative" (ibid.). What ought to be rediscovered is that "there is a latent [. . .] ethics schematized within perception, within the body of lived experience – within the [intercorporeal] attunement that constitutes my first mode of communication with others" (ibid., p. 85).

Thus, Kleinberg-Levin's philosophy is not only constative and descriptive but also performative and prescriptive. Indeed, he urges us to "realize the transformative potential in phenomenology: [. . .] because phenomenology is reflexive, is self-referential, its language is never merely descriptive, but is always functioning performatively, enacting and making true that which it is describing" (Kleinberg-Levin 2008, p. 42). Some responsibility thus falls to the philosopher and most specifically to the phenomenologist, who must admit that the "phenomenological language [is] designed to function *performatively*, making itself into a true description by altering the experience [it describes]. In this case, it would make itself true by actually *connecting* us to our lived experience of intercorporeality – to the dimensions of our experience from which we have become disconnected" (ibid., p. 40). It is thus our moral responsibility to recall the possibility that "a process of reconnection *could* be voluntarily undertaken" (ibid.).

One may wonder, however, what is at stake in such intentional reappropriation of pre-intentional intercorporeality. Is such reappropriation possible at all? And is it possible via a process of (phenomenological) reflection "voluntarily undertaken"? Would such process impact the very nature of intercorporeality, by introducing intentionality into what is first and foremost – and is meant to remain – a prepersonal engagement with another subject who is not – not yet – taken as an intentional object of empathic perception?

Rather than prescribing, as Kleinberg-Levin does, a return to the intersubjective body as subject, against its misappropriation by the individualistic body as object, Csordas assumes that two dimensions of the body are involved *together*: similarity and otherness.

On the one hand, as sketched above, healing would involve some *intersubjective similarity* insofar as it would be based on a somatic mode of attention, engaging the body through which you and I are similar to each other. Csordas agrees that "being a subject for oneself is entirely discontinuous with the lived experience of another person," but he underlines that, nonetheless, "perceiving another person is radically different from perceiving a thing because it is characterized by a "co-positing" of two subjects simultaneously [. . .] We are neither isolated cogitos that must bridge a gulf of solipsism nor participants in the same shared subjective substance. We are similar" (Csordas 2008, pp. 112–113). Such view points "our understanding toward direct recognition of the world and others based on similarity and analogy, and away

from an older hermeneutic akin to mindreading" (ibid., p. 115). As our bodies are similarly intersubjective, our encounter with others is intercorporeal.

But there is another dimension of our body at stake in illness and healing practices: its *otherness*. "This essential otherness originates in the limitations of our physical being that leave us with a sense of inescapable contingency, in the autonomic functioning of our bodies that insistently goes on without us, but which implicates us in anything that happens to our bodies, and in the possibility of seeing ourselves as objects from the perspective of another" (Csordas 1994, p. 158). Csordas highlights that this bodily otherness is interpreted, in some sociocultural contexts, as a "sacred 'otherness'" (ibid., p. 82): "The sacred is an existential encounter with Otherness that is a touchstone of our humanity. It is a touchstone because it defines us by what we are not — by what is beyond our limits, or what touches us precisely at our limits. [...] this sense of otherness itself is phenomenologically grounded in our embodiment" (ibid., p. 5). Religious healing would thus be based on a recognition of the otherness of *our own* body "here magnified to cosmological proportions" (ibid., p. 226) and attributed to the uncanny presence of divine or evil spirits.

In Csordas' conception of healing, thus, two processes operate in the (clinical) encounter of one subject with another:

- The *pre-objective similarity between oneself and another* (that is how the patient and the healer would communicate with each other, through their intercorporeality)
- The *radically objectified dissimilarity of oneself with oneself* (this is what would be attributed to alien forces)

This view presents the advantage of avoiding two pitfalls:

- It takes into account the body as *subject* and avoids reducing it to the introspectively lived body, to rather encompass its *intersubjective* dimension, at a *pre-intentional*, pre-objective level.
- It does not neglect the body as *object*, nor does it condemn it to a negative force of alienation, one should seek redemption from; rather, the body as object is precisely taken as a dimension of *ourselves* which confronts us with *our own otherness*.

One of the main questions that comes here, in the present context of an investigation of the clinical encounter *with another*, arises clearly in reaction to Csordas' description of his own motive: he aims at designing "a cultural phenomenology of healing that seeks the locus of therapeutic efficacy *in the self*" (ibid., p. 5, emphasis added). And indeed, in his view, the body of the other person is read through its similarity with *mine*, and what healers may refer to as alien forces is translated by the anthropologist into the otherness of *my own* body. But how does the body participate to the encounter with *another* subject who is neither similar to me nor alien to me, another subject who I can reduce neither to the body I identify myself with nor to the

body which I expulse from me as an alien force? By considering how the body shapes the clinical encounter with another subject, is it possible to design "a phenomenology of healing that seeks the locus of therapeutic efficacy" neither in the self nor in the other, but in the encounter of oneself with another?

Suffering

To further consider the body as it participates to the clinical encounter, it seems relevant to focus on the *suffering* body, the body of a patient who suffers in such a way that he is led to meet with a clinician.

What is suffering? Michel Henry proposes a radical answer to this question. Suffering resides in the very structure of subjective experience. No matter what the subject experiences, no matter whether the subject experiences something nice or nasty, whether he experiences himself falling ill or healing, he necessarily experiences *himself*, and such *self-affection* is fundamentally suffering: "suffering forms the tissue of existence" (1976, p. 659).

Crucially, suffering is "auto-affection," and as such it "does not and could not result from the affection [. . .] by a foreign being" (ibid., p. 659). "Pure pain does not refer to anything but itself, it is given over to itself, immersed into itself, overcame by itself, crushed by its own weight. Pure pain is [. . .] self-immanence [. . .] – a suffering without horizon, without hope, entirely occupied with itself" (Henry 2000, p. 84).

In this view, suffering "acquires a value and meaning" and even becomes "sacred" insofar as "it refers to him who suffers": "any sufferance feels itself and thus carries in it an "ego", the ego who suffers and without who no suffering is possible" (ibid., p. 30). Moreover, as immanent, sufferance ensures to the sufferer his "separation [. . .] from everything which [he] is not" (Henry 1976, p. 673). As such, suffering contains in and of itself a joy, i.e., the enjoyment of the "perfect adherence of identity" (ibid., p. 660), the "enjoyment of self" (ibid., p. 670), the joy the subject experiences by the very fact of being "forced back by his very suffering [. . .] to himself" (ibid., p. 672). Contrastively, any attempt to sever oneself from one's suffering would lead to despair by revealing "the impossibility for the ego of breaking the bond which attaches him to himself" (ibid., p. 677).

For Henry, if suffering is self-experience, it is also "original passivity," "impotence" (ibid., p. 659). "In suffering, feeling experiences itself in its absolute passivity with regard to self, in its impotence at changing itself, it experiences itself and has the experience of self as irremediably handed over to itself in order to be what it is, as loaded forever with the weight of its own Being" (ibid., p. 658). But "the passivity of sufferance must be radically distinguished from what we usually mean by this term" (Henry 2000, p. 86); the passivity of sufferance is not a passivity relative to something else than oneself, something which would be "foreign" or "anterior" relative to oneself and which would impose itself over oneself; sufferance is not something relative to which one would discover oneself passive (ibid., p. 86). The passivity of suffering is powerlessness, impotence, impossibility to escape *oneself*,

to distance oneself *from oneself* (ibid., p. 88). In Henry's view, therefore, *passivity involves no alterity*: the suffering that resides in self-affection is radically *passive* and radically *immanent*.

To better understand this conception of suffering, it must be underlined that, if suffering is thought to be immanent in such a way, it is because it is *pre-intentional*: for Henry, suffering is not the intentional experience of sufferance, as if it could be kept at an experiential distance from the subject which would take it as its object. Rather, suffering is the very structure of subjective experience, i.e., it is a self-affection which is necessary and necessarily prior to any intentional encounter with any object of experience.

Intentionality, Henry reminds us, is the property of consciousness to be consciousness of something else than itself. In the context of phenomenology, nothing is more banal than this assertion, but Henry decorticates what's hidden behind this seemingly innocent notion of intentionality. If consciousness is intentional, Henry insists, it means that it aims at what is *other* than itself, which means, in turn, that consciousness "turns away from itself in such a radical and violent manner that it is entirely oriented toward other than itself, toward the outside" (Henry 1995, p. 386). If that is the case, then consciousness itself does not and cannot appear to *itself* thanks to an act of intentionality. Thus, reducing consciousness to intentionality is a philosophical decision that contains an "extraordinary violence" insofar as consciousness thereby "sinks into the night" (ibid., p. 391), deprived of any consciousness of itself, reduced to transcendence, at the service of the manifestation of the world.

Henry rejects this view radically, and it is only within the context of his characterization of a non-intentional consciousness that one may understand his characterization of suffering as passive immanence. If "it is not intentionality that accomplishes its own revelation," then another mode of manifestation must be designed, a mode of manifestation which does not give any object, but the very experience of experience, a mode of manifestation which is not the manifestation of something else but self-manifestation (ibid., p. 391). Such self-affection is heterogeneous relative to intentionality (ibid., p. 392). While intentionality is structured by the distinction between a subject and an object, self-affection is not: it lies entirely on the side of the subject affected by itself. And in Henry's terms, this is suffering: a mode of being irremediably affected by oneself – to be radically distinguished from any local experience of some sufferance taken as an intentional object by a subject confronted to pain.

Now, in the present context of an investigation of the clinical encounter of a subject with another one, Henry's conception of suffering confronts us with an obvious question: how may suffering play any role in the clinical encounter with *another* subject, if suffering is *one's* feeling of *one's own* passive *self*-affection? How does immanent suffering participate to the clinical encounter? How is encountering another subject possible at all for a subject who is primarily self-affected? To consider these questions, it is relevant to now turn to another philosopher, Emmanuel Levinas, who is also concerned with "an *affectivity without intentionality* (as Michel Henry clearly noted in his *Essence of Manifestation*)" (2000, p. 17), but who never reduces such non-intentional affectivity to immanence.

As detailed just above, for Henry, suffering is the passivity of one's affection of oneself by oneself. This characterization of suffering, as fundamentally passive, resonates with Levinas' own conception. For Levinas, indeed, suffering is "vulnerability," "pure pathos" (1998, p. 107). It is through the very fact of being *subjected* to suffering that I live it as suffering. Passivity is the very mode by which suffering imposes itself onto my experience. "We do not only know suffering as a disagreeable sensation [. . .] The whole acuity of suffering lies in the impossibility of fleeing it, of being protected in oneself from oneself" (1979, p. 238). Extreme passivity, impotence, radical solitude, or suffering is being enchained to oneself without any possible retreat, without any possibility of assuming oneself, and without any possibility of taking care of oneself.

Such suffering is imposed by "the living human corporeality, as a possibility of pain, a sensibility which of itself is the susceptibility to being hurt, a self uncovered, exposed" (1991, p. 51). In particular, it is through the subject's exposure to another that sufferance may be imposed in a way that "breaks through the crust of its egoism and as it were displaces its center of gravity outside of itself" (1979, p. 239). Already here, it appears that Levinas differs markedly from Henry: the target against which Levinas develops his philosophy is not the transcendent world but the immanent subject, the subject "in possession of itself" in self-affection, or aiming at such possession in an ethics of self-mastery. For Levinas, such subject "is transfigured" by its subjection to suffering and this transfiguration consists in "existing for the Other" (1979, pp. 245–246).

Existing for the other is first and foremost being concerned by the other. More precisely, and more radically, being for the other, caring for the other, bearing "the wretchedness and bankruptcy of the other" (1991, p. 117) is not a choice, a vocation. I *must* carry you, I must, I cannot not carry you. The encounter with another subject, an encounter from which I cannot withdraw and in which I must bear the other's sufferance, is not the result of a conscious deliberation nor of a rational choice; rather, it is imposed to the subject who "finds himself committed" to the other he encounters (1991, p. 122).

Now it can be better understood how much this view differs from the idea – presented above on the basis of Frank's reading of Levinas – according to which *one's own* sufferance may become *useful* if one uses it to give to others the testimony of one's singular capability to cope with it. There is no such usefulness for Levinas, quite the opposite: suffering is irremediably *useless*, and if it may make sense, it is insofar as it bears the sufferance *of the other*, always unjustifiable. One never gets out of one's suffering by an empowering act of self-narration. Suffering, one's own and others', is beyond any capture by any act of intentional consciousness. It is in this sense, precisely, that suffering is more passive than passivity; in particular, it is more passive than the passivity of an intentional object taken as the target of an act of consciousness performed by a sovereign subject of experience.

It can be measured here how much Levinas' notion of suffering differs from Henry's. For both authors, suffering is passivity, but for Henry, as explained above, such passivity involves no alterity but rather the surrendering of oneself to oneself, whereas for Levinas, the passivity of suffering is paradigmatic of one's passivity relative to one's commitment to the other subject. Particularly relevant for our

investigation of the clinical encounter, it is because suffering is so radically passive that salvation may come from the outside and only from the outside of sufferance. *Exposure to the other* comes as a possibility of calling for salvation from the *self-enclosure* of sufferance. Contrary to Henry who argues that the self-enclosure of sufferance cannot and should not be broken up, but can and should rather be nurtured into the joy of being affected by oneself, in the immanence of one's self-affection, Levinas contrastively argues that the self-enclosure which defines suffering must be opened and can only be opened by another than the sufferer himself, since the latter is, per definition, self-entrapped.

Even though to suffer is to be passively subjected to sufferance, one's suffering may be expressed into a cry, a call, a demand for salvation. Here suffering may be *transformed* into an address to the other, and this is exactly what occurs in a clinical encounter (Ingerslev Ryberg and Legrand 2016). The tears, the grimace, the complaint, the narration of his pain by the patient *call* the clinician to *respond* to the others' suffering. But what is such response? In the clinical encounter, before asking for a remedy, a subject who suffers first demands for his suffering to be *recognized* and therefore to be listened to. Thus, the first act by which the clinician responds to the suffering of the patient is the recognition of this suffering and, by virtue of this, the recognition of the subject who is caught in it. In this view, responding is most primarily the act of someone who *listens* to the demand of someone else; responding to the other is recognizing that he is soliciting my listening to the *singular* demand that he addresses me.

Among other clinical approaches, the psychoanalyst holds a "responsive stance," as clearly expressed by Jacques Lacan as he underlines that "the psychoanalyst [...] comes upon the simple fact that language, prior to signifying something, signifies to someone. It is simply because the analyst is there listening that the man who speaks addresses him" (Lacan 2002, p. 66). Listening in this way is assuming that the speaker specifically addresses the listener. It becomes clear here that the silence of the psychoanalyst does not minimize his act of responding to the patient, but on the contrary maximizes it, if the analyst turns this suspension of his speech into an act of listening responsively. To listen is at once both to respond to the speech given *by* the patient and to give speech *to* the patient who responds to the listening silence. Such responsive listening is not a pure passivity; it is an act of mine that starts elsewhere, there where I am not, there where I shall never be, and there where the other is.

What emerges here is the idea that responsive listening has a metamorphic power. Indeed, as the patient is listened to responsively, he may be led to address his sufferance to the clinician: he may cry, moan, grimace, and also talk, describe, narrate, and communicate his sufferance to the other. The impact of such communication is not to offload the heavy burden of physical or psychical pain. Quite the contrary, the impact is to take *responsibility* for one's own sufferance, that is, to respond to it. And if that happens in the clinical encounter, it is because, by the very act of listening to me responsively, you position me in a space of responsiveness where I am in a position to respond to my sufferance as my own singular state, because it is as such that you respond to it and that I address it to you.

By contrast with what has been described above in the framework of narrative approaches to illness and medicine, here *to speak* does not involve any *sovereignty*

of the speaking subject because, here, to speak involves to listen: *to respond to one's sufferance*, by narrating it, is most notably *to respond to the other* who is listening responsively. It is because it is responded to by another that a complaint, a cry, a face marked with suffering becomes an "original call for aid, for curative help, help from the other me whose alterity, whose exteriority promises salvation" (Levinas 1998, p. 107). It is addressing the other that metamorphoses unassumable suffering into speech; in other terms, it is not self-mastery that may control pain; rather, it is encountering another subject that can metamorphose a destructive force into an address.

Moreover, and again by contrast with what has been described above, here *to listen* does not involve any *empathic* assimilation of the others' experiences. In particular, the other who speaks and who I listen to cannot be absorbed into a narrative; it cannot be captured as a thematic object of description or knowledge. No narrative, no matter how detailed it could be, will ever capture the other as he addresses me singularly, because his address is not determined by himself but by our encounter, as it is partly determined by my response, at minima by the fact that I listen to him or not.

With Levinas, a radical conclusion may be drawn from these considerations: the "divergence that inevitably opens between the Other as my theme and the Other as my interlocutor [. . .] announces the ethical inviolability of the Other" (Levinas 1979, p. 195). That which remains irreducible is the other who comes to meet me, to address me, imposing upon me to respond to him, if only by listening to the singularity of his vulnerability to suffering. No matter how much I know the details of his intimate life, no matter how much I apply my expertise in an attempt to lighten his distress, the other who addresses me "remains infinitely transcendent, infinitely foreign" (ibid., p. 194), and it is as such that he is inviolable. The encounter between one and another starts with the recognition of the inviolability of the other, the recognition of the other as a stranger who I encounter without assimilating it to me or assimilating myself to him, as a stranger who I address myself to, who I listen, who addresses me, who I respond to. And this is where ethics starts: where an encounter occurs between one and another.

Bodily Ethics

A call shall be heard where "a moan, a cry, a groan or a sigh slips through" (Levinas 1998, p. 93). In such a call, suffering is transformed: while it is "intrinsically senseless and condemned to itself with no way out, a beyond appears in the form of the interhuman" (ibid., pp. 93–94). For Levinas, it is precisely here that "the anthropological category of the medical" imposes itself as "primordial, irreducible and ethical" (ibid., p. 93).

This characterization of the medical as ethical deserves closer scrutiny. Following the manner in which this term is used by Levinas himself, an encounter *with another* is in and of itself ethical: ethics is "the relationship of man to man" (1979, p. 79); ethics is "the non-indifference to another" (1991, 48). As ethical, the encounter with the other "is not a modality of cognition" and it is "irreducible to the circulation of

information" (1991, p. 48). Ethics is "the welcoming of the other" (1979, p. 43), the other who arrives without being taken as an object of intentional consciousness, the other who arrives and imposes his otherness before the subject can experience any similarity or dissimilarity with him. As ethical, an encounter with the other starts from the other, from "the strangeness of the Other, his irreducibility to the I, to my thoughts and my possessions"; as ethical, an encounter is thus "precisely accomplished as a calling into question of my spontaneity" (ibid.). Such "calling into question of my spontaneity by the presence of the Other" (ibid.) is precisely what the term "ethics" names.

This notion of ethics is minimal because it arises "before the bipolarity of good and evil presented to choice" (1991, p. 122), and it is radical precisely for the same reason: because it arises before any moral judgment of what is good and what is bad. In the radical minimality of the term ethics, to encounter the other is in and of itself to be engaged in an ethical relationship which is "irreducible" (ibid., p. 135): there is no encounter which is not ethical, in this minimal and radical sense of the term "ethics." If an encounter were unethical, in the sense the term "ethics" is given here, it would not be an encounter of the other *as another*, i.e., it would not be an encounter, in the sense this latter term is given here.

However, even though the ethical dimension of the encounter is irreducible in such a way, it is crucial to underline that such ethics is not an abstraction: it exists only as performed within a concrete encounter with the other. This immediately raises the question of its implementation: How is the clinical encounter *practiced* ethically?

Practicing a clinical work inspired by a Levinasian ethics, Paul Komesaroff focuses on the "ethical interchange in the clinic" (2008, p. 14) and underlines "the 'microethical structure' of medicine" (ibid., p. 6). For him, "ethics is what happens in every interaction between every doctor and every patient" (ibid., p. 27). That is, he is interested in ethics as it occurs "at the level of individual experience, the local or 'microethical' level, where one person engages another face to face" (ibid., p. xv). In this view, "every clinical relationship consists of a continuous series of ethical events, each of infinitesimal dimension and often inconspicuous to the participants" (ibid., p. 5). For example, "how does one gain the trust of a person one has never met before, to such an extent that she will grant access to her most private experiences?" (ibid., p. 28).

Confronted to such unceasingly unprecedented issues, medicine is "a practice of ethics" (ibid., p. 5). Such practice is not a matter of "adding empathy and friendship to the clinical discourse, as these may well hinder critical reflection on the part of both doctor and patient" (Lingis 2008, p. x). Rather, the clinical encounter occurs between two partners who remain irreducible to each other. It is not and does not aim at being only a relationship of mutual understanding, "of support and affirmation, but also of subversion and confrontation" (ibid., p. xii); the encounter "is asymmetrical and non-reciprocal [because] the otherness of the other is irreducible and unfathomable" (Komesaroff 2001, p. 324). In particular, thanks to the impossibility to reduce the patient's experience to the clinician's expertise, (micro)ethical clinical practices involve respecting the singularity of each participants, i.e., their alterity relative to each other, as well as their codependency.

In this view, (micro)ethics is intrinsic to the encounter. In other terms, ethics cannot be determined by an external discourse, and it is "radically disengaged from any unitary notion of the good" (Komesaroff 2008, p. 17). In the spirit of ethical committees exterior to the practice into which the clinical encounter is immersed, "it is assumed that ethical problems are solvable and that the solutions can be discovered through the application of rational thought (ibid., p. xvii). Although there is some heterogeneity among theories in biomedical ethics, they share a commitment to the "possibility of universal moral principles" (ibid., p. 10). By contrast, to hold a microethical stance is to be aware of the "constant process by which ethical issues arise, are dealt with in the course of the interaction and subsequently pass away" (ibid., p. 6): it is from the encounter with the other that minimal ethics surges. Thus, if it can be characterized as the encounter of one subject with the suffering of another, then the clinical encounter *is* ethical: it imposes itself as ethical, insofar as the other's suffering remains irremediably irreducible to mine.

Now it appears that the body which participates to the clinical encounter and shapes it as ethical is neither the body of a sovereign subject nor the body as an inanimate object; rather it is the body as anchoring the position from where one subject may address another, and *as such* it is also, and by the same token, the body "of flesh and blood in matter" (Levinas 1991, p. 78). Indeed, for Levinas, the one who addresses himself to the other, the one who responds to the other, is not made of evanescent words but of matter. As such, "matter is the very locus of the for-the-other" (ibid., p. 77). Indeed, it is only insofar as the subject is "of flesh and blood, a man that is hungry and eats, entrails in a skin" that he is *thereby* "capable of giving the bread out of his mouth, or giving his skin" (ibid.). To be for the other is not to nurture "elevated feelings" but to tear away the "bread from the mouth that tastes it, to give it to the other" (ibid., p. 64); it consists in "nourishing, clothing, lodging, in maternal relations, in which matter shows itself for the first time in its materiality" (ibid., p. 77).

Levinas' view of the body stands in sharp contrast with Henry's. For the latter, "our body is originally neither a biological body nor a living body *nor a human body*; it belongs to an ontological region radically different which is the region of absolute subjectivity"; this body is a "*transcendental* body": "a body which is an 'I'" (1975, p. 8). Throughout his work, Henry differentiates radically the feeling body as subject which is non-intentionally self-affected, on the one hand, and, on the other hand, the felt body as object taken at the target of intentional consciousness. By contrast, Levinas considers the "whole" body, not only the subjective flesh of the lived body, but also the living body always vulnerable to death, and as he underlines explicitly himself: "in [such] corporeality are united the traits we have enumerated" to characterize the encounter between two subjects: one's body is "for the other, despite oneself" (Levinas 1991, p. 55). In such a way, what appears now is how the encounter is bodily and *as such* ethical. In other terms, in the view which has been unfolded here, the clinical encounter intrinsically involves a *bodily ethics*. Indeed, as characterized here, an ethical stance imposes to give hospitality to a realm of otherness which notably materializes itself in the body to be taken care of, the patient's body which remains irreducibly other relative to both the patient himself and the clinician.

Definitions of Key Terms

Intentionality	Consciousness is intentional inasmuch as it is consciousness of something else than itself. Intentional consciousness is structured by two inseparable poles of experience, a subjective pole and an objective pole. Schematically, one may simplistically represent intentional consciousness as an arrow, with its subject as its starting point and its object as its ending point.
The body as subject	As subject, the body anchors one's experiential perspective.
The body as object	As object, the body is taken as a target of a conscious experience.
Intercorporeality	As intercorporeal, our bodies are pre-intentionally structured intersubjectively, in the sense that the other always already resonates in me, before I perceive him as another and interpret his behavior.
Suffering	At a non-intentional level, the subject is suffering inasmuch as it is affected by itself without any possible retreat; suffering is being subjected to passivity, powerlessness, and impotence; suffering is being confronted to the impossibility to escape oneself, to distance oneself from oneself.
Bodily ethics	A bodily ethics is a minimal ethics which starts with the encounter of another as another, at the bodily level. Such a bodily ethics imposes to give hospitality to a realm of otherness which notably materializes itself in the body to be taken care of, the patient's body which remains irreducibly other relative to both the patient himself and the clinician.
Responsive stance	To hold a responsive stance is to assume that one is specifically addressed by another subject and to respond accordingly. Holding a responsive stance, the clinician assumes the responsibility of encountering the patient as a subject by who he is singularly addressed.

Summary Points

- Medicine is first and foremost a practice, which cannot be conceived of without the encounter of a patient with a clinician. The philosophy of medicine, thus, must also be a philosophy of this practice, i.e., a philosophy of the encounter of a clinician with a patient.
- The clinical encounter cannot be understood, and performed, without considering how it is an encounter of both the patient and the clinician with a suffering body. How does the body participate to the encounter with another subject?

- In contemporary debates, evidence-based medicine is thought to favor exclusively the body as object and is challenged by alternative conceptions and practices, notably grounded on a phenomenological characterization of the body as distinctively subjective.
- In narrative approaches to illness and to medicine, narration is thought to benefit the patient in that it promises him to keep or regain his sense of being himself, despite or thanks to what happens to him in illness. Narrative approaches draw on an ethics of self-mastery.
- The narrative competence of the clinician involves holding an empathic stance upon the patient's illness story.
- At a pre-intentional level, the encounter with another subject is not the result of a conscious deliberation nor of a rational choice. Rather, it is intercorporeal: our bodies are pre-intentionally structured intersubjectively, in the sense that the other always already resonates in me, no matter the particular stance I take on him via my intentional experiences of him. Moreover, at a pre-intentional level, the subject cannot withdraw from the encounter and must bear the other's sufferance that imposes itself.
- To suffer is to be passively subjected to sufferance without any possibility of protecting oneself from oneself. The self-enclosure which defines suffering can only be opened by another than the sufferer himself, since the latter is, per definition, self-entrapped.
- When expressed into a call for salvation, suffering may be transformed into an address to the other. A responsive stance allows the clinician to assume that the patient addresses his singular sufferance specifically to him and allows him to respond accordingly.
- As ethical, the encounter between one and another starts with the recognition of the inviolability of the otherness of the other.
- A bodily ethics imposes to give hospitality to a realm of otherness which notably materializes itself in the body to be taken care of, the patient's body which remains irreducibly other relative to both the patient himself and the clinician.

References

Charon R (2001) Narrative medicine. A model for empathy, reflection, profession and trust. JAMA 286:1897–1902

Charon R (2008a) Where does narrative medicine come from? Drives, diseases, attention, and the body. In: Rudnytsky PL, Charon R (eds) Psychoanalysis and narrative medicine. State University of New York Press, Albany, pp 23–36

Charon R (2008b) Material and metaphor: narrative treatment for the embodied self. In: Rudnytsky PL, Charon R (eds) Psychoanalysis and narrative medicine. State University of New York Press, Albany, pp 287–294

Csordas T (1993) Somatic modes of attention. Cult Anthropol 8:135–156

Csordas T (1994) The sacred self: a cultural phenomenology of charismatic healing. University of California Press, Berkeley

Csordas T (2008) Intersubjectivity and intercorporeality. Subjectivity 22:110–121

Eakin PJ (1992) Touching the world: reference in autobiography. Princeton University Press, Princeton
Eakin PJ (2001) Breaking rules: the consequences of self-narration. Biography 24(1):113–127
Eakin PJ (2004) What are we reading when we read autobiography? Narrative 12(2):121–132
Eakin PJ (2006) Narrative identity and narrative imperialism. Narrative 14(2):180–187
Frank A (1995) The wounded storyteller: body, illness, and ethics. University of Chicago Press, Chicago
Gadamer HG (1998) Philosophie de la santé. Grasset, Paris
Garden R (2007) The problem of empathy. New Literary History 38(3):551–567
Garden R (2010) Telling stories about illness and disability: the limits and lessons of narrative. Perspect Biol Med 53(1):121–135
Henry M (1975) Philosophy and phenomenology of the body (trans: Etzkorn G). Martinus Nijhoff, The Hague
Henry M (1976) The essence of manifestation (trans: Etzkorn G). Martinus Nijhoff, The Hague
Henry M (1995) Phénoménologie non-intentionnelle : une tâche de la phénoménologie à venir. In: Janicaud D (ed) L'intentionnalité en question. Vrin, Paris
Henry M (2000) Incarnation. Le Seuil, Paris
Ingerslev Ryberg L, Legrand D (2016) Clinical response to bodily symptoms in psychopathology. Philos Psychiatry Psychol 23:3
Kleinberg-Levin D (1999) A responsive voice: language without the modern subject. Chiasmi Int 1:65–103
Kleinberg-Levin D (2008) Before the voice of reason: echoes of responsibility in Merleau-Ponty's ecology and Levinas's ethics. State University of New York Press, New York
Komesaroff P (2001) The many faces of the clinic: a Levinasian view. In: Toombs K (ed) Handbook of phenomenology and medicine. Kluwer, Dordrecht, pp 317–330
Komesaroff PA (2008) Experiments in love and death: medicine, postmodernism, microethics and the body. Melbourne University Publishing, Melbourne
Lacan J (2002) Beyond the "reality principle". In: Ecrits (trans: Fink B). Norton, New York, pp 58–74
Leder D (1992) A tale of two bodies: the cartesian corpse and the lived body. In: Leder D (ed) The body in medical thought and practice. Philosophy and medicine, vol 43. Kluwer, Dordrecht, pp 17–35
Legrand D (2011) Phenomenological dimensions of bodily self-consciousness. In: Gallagher S (ed) Oxford handbook of the self. Oxford University Press, Oxford, pp 204–227
Legrand D (2012) Self-consciousness and world-consciousness. In: Zahavi D (ed) Oxford handbook of contemporary phenomenology. Oxford University Press, Oxford
Legrand D (2013) The patient's voice in psychoanalysis, narrative medicine and patient-based research: What is the clinical impact of patient-based research? Bull Graduate Sch Educ (Hiroshima University) part I 62:29–38
Legrand D (2015) The body and its image in the clinical encounter. In: Elpidorou A, Dahlstrom D, Hopp W (eds) The reappearance of the mind: essays on the intersections between phenomenology and the philosophy of mind. Routledge, London
Levinas E (1979) Totality and infinity (trans: Lingis A). Martinus Nijhoff Publishers/Duquesne University Press
Levinas E (1991) Otherwise than being or beyond the essence (trans: Lingis A). Kluwer, Dordrecht
Levinas E (1998) Useless suffering. In: Entre nous, On Thinking-of-the-Other (Trans: Smith MB, Harshav B). Columbia University Press, New York
Levinas E (2000) God, death and time (trans: Bergo B). Stanford University Press
Lingis A (2008) Foreword. In: Komesaroff PA (ed) Experiments in love and death: medicine, postmodernism, microethics and the body. Melbourne University Publishing, Melbourne, pp ix–xii
Macnaughton J (2009) The art of medicine. The dangerous practice of empathy. Lancet 373:1940–1941

Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge, London/New York
Phelan J (2005) Editor's Column: Who's Here? Narrative, 13, 3:205–210
Rose N (1998) Inventing our selves: psychology, power, and personhood. Cambridge University Press, Cambridge
Rose N (2007) The politics of life itself: biomedicine, power, and subjectivity in the twenty-first century. Princeton University Press, Princeton
Sartwell C (2000) End of story. State University of New York Press, New York
Slatman J (2009) Transparent bodies: revealing the myth of interiority. In: van de Vall R, Zwijnenberg R (eds) The body within. Art, medicine and visualization. Brill's studies on art, art history, and intellectual history, vol 3. Brill Academic, Leiden, pp 107–122
Slatman J (2014) Multiple dimensions of embodiment in medical practices. Med Healthc Philos. 17, 4:549–57
Zaner RM (1992) Parted bodies, departed souls: the body in ancient medicine and anatomy. In: Leder D (ed) The body in medical thought and practice. Philosophy and medicine, vol 43. Kluwer, Dordrecht, pp 101–122

Trust and Mistrust Between Patients and Doctors

31

John Saunders

Contents

Abstract

Concerns about trust are long-standing, although mistrust may be increasing. Daily life could not flourish without trust. It can be defined in a variety of ways, but health-care programs have not demonstrated proven interventions by which it can be increased. It is intrinsically and instrumentally valuable. Good health care requires trust in systems as well as individuals. Trustworthiness is a virtue, whereas not all are trusted even if trustworthy. Trust functions at different levels of knowledge. An ethos of mistrust leads to a contractual relationship, with an infinite regress as to where one places trust. There is a paradox between trust and rationality: this depends on how rationality is construed, but in some situations we may trust against the evidence. The uncertain outcomes in medicine mean that sometimes trust may lead to disappointment, but trust should not automatically be abandoned.

J. Saunders (✉)
College of Human and Health Sciences, Swansea University, Swansea, UK
e-mail: saundersjohn@doctors.org.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_20

Introduction

Anxieties that trust between patients and doctors is threatened are long-standing: even a cursory examination of the medical literature will reveal concerns for over 30 years. Just as the economy always needs fixing, so trust always seems to be threatened. Some reasons seem easy to find. In the UK, for example, a series of scandals in health care has involved both institutions and health carers – individual doctors and nurses. Institutional examples include:

- The Bristol heart scandal in which high deaths in children's cardiac surgery in the 1990s led to a major inquiry with sweeping criticisms of the unit.
- The organ retention scandal at the Royal Liverpool Children's Hospital (Alder Hey). Here the unauthorized retention and disposal of body parts from 850 infants led to the UK's Human Tissue Act in 2004.
- The poor care standards in the Mid-Staffordshire Trust at Stafford hospital. This led to a major inquiry by Robert Francis. Between 2005 and 2008, 400–1200 excess deaths for a hospital of this size were estimated.

Individual examples include the cases of:

- Harold Shipman, a general practitioner who murdered up to 250 patients. He was convicted in 2000.
- Rodney Ledward, an incompetent gynecologist, found guilty of professional misconduct in 1998. Fifty-eight women subsequently alleged sexual assault.
- Beverley Allitt, a nurse, who was convicted in 1993 of murdering four children, attempted murder of three, and seriously harmed another six.

Mass media have often played up these reports: the role of the popular *Daily Mail*, for example, in negative reporting on the UK's National Health Service (NHS) is a major example. Despite this, trust in doctors remains high. "Fortunately most people still seem to think highly of their own physicians, even though they have doubts about physicians in general" (Alper 1988). That was written in 1988. Thirty years later, post Alder Hey and Bristol, there was little deviation in support for doctors even at time of high adverse publicity. Respondents with experience of the NHS were more likely to state that they thought doctors did their job very well (Ford 2007). In the UK, a survey showed that 90 % of adults trusted their doctors to tell them the truth (Ipsos-Mori 2014), putting them above all other groups and far beyond journalists or politicians. On the other hand, perhaps patients may trust blindly when some skepticism is warranted. "Much care that is needed is never provided, and ineffective and inappropriate care is common" (Mechanic 2004). Even setting well-publicized scandals to one side, there have, of course, been many individual professionals in medicine who have violated trust: the more extreme examples feature in the disciplinary hearings of the UK's General Medical Council, available to all to read on the Internet.

This chapter will discuss the importance of trust for individuals and for society and its particular importance in health care. The understanding of what constitutes trust will be compared with trustworthiness and the latter considered from an Aristotelian perspective. Trust will be considered from a phenomenological perspective and the dangers of an ethos of mistrust. Trust relates to autonomy, but in the final section, its relationship to rationality leads to paradox.

Trust in Daily Life

Nobody lives without trust: trust in materials, trust in systems, trust in organizations, trust in other people, and trust in ourselves. Without trust, life could not function. There would be endless searching for evidence, then validating the evidence, then validating the validation in an infinite regress.

In his "Essays on the Laws of Nature" (Locke 1663), John Locke describes trust as "the bond of society." Self interest, says Locke, cannot be the basis of the laws of nature. People need each other and must therefore cooperate. Man is naturally a social animal. A more recent commentator states that trust "provides the glue that makes cooperation possible without costly and intrusive regulation" (Mechanic 2004). It is important in its own right to give relationships intrinsic value; it is also important for instrumental reasons. To achieve certain ends in medicine, care must be sought, information must be disclosed, and treatment must be followed.

Trust is what gives individuals sufficient confidence that a material, a system, an organization, or other people will do what is expected and will do it, moreover, without the need for checks, tests, assessments, proofs, experiments, or special assurances. The bridge is trusted to carry the weight of those that use it, airline security is trusted to identify the terrorist, the health screening service is trusted to identify the victim correctly, others are trusted to give the assistance that is expected, and individuals trust themselves that they will rise to a particular challenge. A declaration of trust is a declaration about a forthcoming action or possible action: even a religious believer who trusts in God will believe that God will *do* something, answer some prayer, give some aid, and sustain in life's tribulations. Trust orientates itself to action.

At the personal level, somebody who is trusted is deemed "trustworthy": deserving of trust by the person that knows him or her and perhaps deserving the trust of others too. This is a positive description: nobody would want to be described as untrustworthy. Trustworthiness seems to function as a virtue.

This functions at societal level too. As a society, governments are trusted to promote and defend the national interest by its citizens, to create just laws, to extirpate unfairness of a gross kind, or to root out corruption.

Whether another person or institution is trusted will depend on whether that other person is thought trustworthy (Hardin 2005) in the situation in which an individual finds him or herself. That introduces a cognitive element into what might otherwise be considered a moral virtue. On what are such judgments based?

O'Neill, who has contributed most notably to recent discussions of trust, introduces her Reith lectures with the following observations (O'Neill 2002a, vii):

> Trusting is not a matter of blind deference, but of placing – or refusing – trust with good judgement. So we need social and political institutions that allow us to judge where to place our trust. Yet some fashionable ways of trying to make institutions and professionals trustworthy undermine our abilities to place and refuse trust with discrimination.

She goes on to give examples: perverse incentives, rewards for contributions that cause bigger adverse consequences in a related area; micromanagement, losing sight of what is important by a focus on minor unimportant matters; false goals, aiming for what is irrelevant; and so on.

Trust is multidimensional although frequently treated as a single dimension in research studies. It includes factors such as fidelity, honesty, confidentiality, and competence. A variety of definitions have been proposed:

- Worthiness of being relied on, fidelity, a resting on the integrity, and friendship of another (Kirkpatrick 1983)
- The process by which barriers to cooperation and compliance are overcome (Dibben et al. 2000)
- A quality with a subjective component that requires an optimistic acceptance of vulnerability (Hall et al. 2001)
- The willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party (Mayer et al. 1995)
- The expectation of the public that those who serve them will perform their responsibilities in a technically proficient way (competence), that they will assume responsibility not inappropriately defer to others (control), and that they will make their patients' welfare their highest priority (agency) (Mechanic and Schlesinger 1996)

It is claimed that trust is a coherent psychological construct that can be reliably measured and distinguished from satisfaction. Satisfaction is an evaluation of previous experiences, whereas trust is primarily future oriented (Hall et al. 2002). Nevertheless, no methods have been demonstrated to improve patients' trust in doctors in a variety of studies (Rolfe et al. 2014).

Trust Between Patients and Doctors

So whether health care or personal and daily societal life are being discussed, trust is needed. Patients need to trust their doctors; doctors need to trust their patients. "Trust – a firm belief in the honesty, integrity, reliability, and justice of another

person or thing – is the critical foundation of an effective patient-physician relationship" (Hillman 1996). The UK's General Medical Council, in its guide to practice, states simply that patients need good doctors and good doctors are trustworthy (GMC 2013). Similarly the Royal College of Physicians of London in its report on professionalism asserts that "securing trust is the most important purpose of medical professionalism. Moreover "trust – and so professionalism – operates at two levels: in the doctor providing care (individual professionalism) and in the system where that care is given (institutional professionalism)"(RCP 2005). That trust is based on several sources: on patients' beliefs that doctors are technically competent, on interpersonal competence, and on indications that the doctor is their ally and, on occasions, their advocate. Good doctors demonstrate good communication skills, the ability to listen, and the evidence that they care. Patients do not expect intimacy, but they do seek respect and responsiveness.

It has been stated that 40 % of deaths in the USA may relate to behavior patterns – among which smoking, diet, and physical activity are prominent. Adherence to prescribed medication in chronic diseases may be as low as 50 %, large numbers of patients miss follow-up appointments, and many drop out altogether. Addressing behavior patterns requires partnerships between doctors and patients. Achieving this means gaining respect and trust (Parekh 2011). Medicine is primarily a profession and not a business. "The consultation room is not a bargain counter. . .Words and deeds must emphasise the difference between (doctors') work and that of the market place" (Ingelfinger 1972).

Doctors are also the gateway to trust in the health-care system, and many believe that they have a responsibility to cooperate with managers to build trust in systems (Mechanic 2004). Management changes or reorganization that is politically driven can reduce trust. All modern health-care systems are complex, and major interventions are likely to have many unintended consequences, some of which may impact on trust. For example, demanding the identification by doctors of certain groups of patients for the purposes of charging may have a significant impact on the doctor–patient relationship. The sharing of information required by statute in Multi-Agency Protection Arrangements could mean that confidential information could be released to others without patients realizing this (Jones 2007). This is not conducive to trust.

Trust is needed most when individuals are at their weakest and most vulnerable. It is when individuals are most dependent that trust is most easily given, because they are most needy: ill, psychologically disturbed, too old, too young, and too cutoff by language – a foreigner, literally or metaphorically, in a strange land. At the end of life, pursuing more and more interventions to less and less effect reaches a point when it is in the interests of neither patient nor doctor. "The greater the trust between physician and patient. . .the more willing patients will be to refrain from pursuing long odds to achieve bad ends" (Caplan 1996).

Society constantly changes and what has been termed the "moral fabric" of medicine has evolved significantly. This has been set out in six developments as follows by Shortell et al. (1998):

- Whereas doctors were formerly responsible only for individual patients, they now carry responsibilities for a wider population.
- Individual clinical responsibility has developed into team or group responsibility, thus necessitating a dialogue between doctors, other team members, and the patient.
- Reliance on mystique and prestige to underpin credibility and trust has changed to a need for performance data and documented evidence.
- Other public bodies now determine performance and accountability criteria along with the profession, instead of the profession determining this alone.
- Organizations exist to serve patient and societal interests instead of only doctors' interests.
- Doctors are accountable not only to patients and the profession but also to the health-care organization in which or for which they work.

Trustworthiness as a Virtue

Personal trustworthiness is one of a family of other-regarding characteristics or qualities that might include truthfulness, honesty, reliability, and consistency. They are features that affect what is done or decided: that is, they are *moral* qualities. Such qualities might be considered as *virtues*.

Virtues are developed by studying and copying the life well lived (Anscombe 1958). From this perspective, the relationship resembles that of an apprentice toward a master. Virtues are excellencies that contribute toward human flourishing. The "apprentice" sees the life that flourishes and emulates its exemplar with his or her practical demonstration of how to live. It could even be said that a person is followed rather than his creed. Virtue ethics requires a conception of the good of a human life, conceived as a unity, and demonstrates the creative, rational, social, and communicative qualities of human beings to an excellent degree. Aristotle's ethics is teleological, but not consequential. That is, it rests upon a consideration of the attainment of some good – the starting point of his Nicomachean Ethics (Aristotle 1984). Kant expounds a duty-based ethics, Aristotle a virtue-based one.

In virtue ethics, goodness is defined not as rightness but as a human excellence, and for this reason the practice of a virtue must aim at perfection. Its underlying question is not what should I do? Rather, it is how should I live? Virtue ethics emphasizes the character, here the trustworthiness, of the agent and accepts that the self is morally important. Virtue therefore applies to traits of character, to dispositions, and to character patterns that lead to behavioral consequences. Moral development therefore takes time as character forms and dispositions to do good develop. By contrast, deontology places its emphasis on action. Deontology sets out principles and rules in a quasi legal manner, with practice seen as obligation; virtue interprets practice as the expression of an underlying character sensitive to culture and community traditions.

The virtuous person will act in a certain way because he wants to act in this way, because acting like this will realize a virtuous end. In acting, character and

disposition are developed. The virtuous man will vary behavior to the context, in order to apply the particular virtue. If deontology could be caricatured as a rule-based cookbook, virtue ethics is more akin to a connoisseur sampling a fine food. This difference of approach necessitates an entirely different emphasis on the moral agent. The exemplar from whom the virtuous person will learn is likely to be someone with great experience of life and wisdom. While not everyone who has experience has the wisdom that comes from reflecting on it, without experience the ability to respond in different situations will be less. The person or institution that is trusted should have the flexibility to respond appropriately and in accordance with expectations in differing circumstances.

Nevertheless, trustworthiness does not sit easily with the Aristotelian doctrine of virtue as being a midpoint between extremes, for example, courage as a midpoint between rashness and cowardice and generosity as a midpoint between profligacy and miserliness. It is hard to conceive of criticizing somebody for being too trustworthy, just as it is difficult to criticize for being too just. By comparison, being too trusting represents gullibility, and trusting too little, being suspicious or even paranoid.

The Phenomenology of Trust

In his essay discussing the role of contract in trust, Pellegrino (1991) quotes two accounts that are suggested sociologically. According to the first, professionals have expert knowledge compared to their patients, they are independently certified as competent to practice that knowledge, and they have a fiduciary relationship that creates obligations both to the individual and also to society. A second view is that trust is a way of reducing complexity. Society is complicated and we must accommodate the freedoms of others. On this account, trust "is the generalized expectation that another will handle his freedom, his disturbing potential for diverse action in keeping with his personality, or rather in keeping with the personality which he has presented and made socially visible" (Pellegrino 1991, 71). On either account, there are gaps to be bridged: necessary contingent features in any interpersonal relationship. The professional cannot confidently know everything about the patient, and the definition of the patient's good may rest upon features unknown to the doctor. Fulfilling trust implies that at least some latitude is granted by the patient to the doctor. In turn the doctor will use such latitude wisely, neither assuming too much nor too little.

At a basic level, trust adheres to a particular social role. In an emergency, a patient probably has no knowledge of the doctor, any more than the aircraft passenger has any knowledge of the pilot. The patient trusts the doctor – or the pilot – because she/he knows that the individual has been trained and assessed in certain ways. Trust is largely about the system and not the individual. At a more complex level, patients will reveal to their doctor matters of social intimacy, behaviors that may be embarrassing (or even unlawful), the delicacies of family relationships, the secrets of their hearts, their vices, or foibles or subject themselves

to intimate bodily examinations from which any other person would be excluded. This creates an expectation and a patient must be prepared to trust that professional at a personal level not to disclose the intimacies divulged. In the end, that trust is unavoidable. If another opinion is sought, then there is the decision as to which opinion should be followed – or whether to seek a further opinion in an infinite regress to find the "truth."

The exercise of patient autonomy (understood as choice) might seek to limit the latitude granted by trust by resorting to contract, but even a contract requires trust that it will be followed. An advance refusal of medical treatment, or indeed an advance directive of any sort, is an attempt to limit medical discretion and to replace it with an explicit policy of management. But the doctor still has to be trusted to have the latitude to decide whether the situation fits that specified in the directive: what an ordinary or extraordinary measure might be and what a reasonable chance of success might be. The doctor's choices may have been limited by the directive, but somebody has to be trusted to ensure that the directive is followed. "Living wills cannot supplant trust because their execution depends on it" (Pellegrino 1991). A contract may reduce the doctor's latitude for action but can neither envision all circumstances nor do without trust that they will be implemented.

An Ethos of Mistrust

As already noted, there are situations where patients might have been well advised to withhold trust from their doctors. Incompetence, venality, dishonesty, and insensitivity are all part of the spectrum of human qualities and behaviors. Doctors are not separate from wider social trends, and there are no reasons to think that they are morally superior to the rest of the population. They are relatively wealthy, with life styles to match; the feminization of the profession in the UK has been predictably associated with more medical couples and household incomes to match; and the preference for nine to five working has increased. It is easy to understand if patients feel that doctors do have less interest in them personally. One response is to mistrust, to research the Internet, to use doctors as one resource among many, to view oneself as a consumer, and to be skeptical of "experts."

The danger of an erosion of trust is that ultimately the possibility of trusting relationships with professionals becomes impossible. "Can a sick person be healed – made "whole" again – when she is suspicious of the motives and methods of her healer? A sick person must be empowered to heal herself. Is this possible when the person empowering is suspected of fostering her own self-interest?"(Pellegrino 1991, 78).

Doctors' expertise is primarily in medicine and in its basic sciences. Among the many values that they hold, medical values are likely to be placed first. By "medical values," it is meant essentially scientific information of accepted reliability about what investigation or treatment is most likely to achieve the preferred outcome of the medical professional. These medical values however are not just brute facts

about the world. They are selected and presented from a particular viewpoint, according to the doctor's values. These may be very different from the patient's values. If it is thought that *some* standard of virtue is not inherent in medical practice, then doctors cannot be trusted to promote the patient's good. In this ethos, the only patient protection is more and more stringent contractual demands. As above, contracts do not abolish the need for trust but may shift responsibilities elsewhere. An external third party may be needed to police the contract and to ensure its application or applicability in a particular situation. If there is dispute, the patient will still have to trust one party – or seek yet another in a potentially infinite regress.

An ethos of mistrust is likely to turn a doctor–patient relationship into a primarily legalistic one. In this ethos, both parties will become more protective of their respective interests. The limits of obligation will be more closely adhered to, and the acts of supererogation that are commonplace in so many human relationships will disappear. The personal will be replaced by the impersonal, the desire to understand by the minimalism of contractual fulfillment, the responsibility of care by a shift of responsibility to the patient, and a shared enterprise to a solo achievement by the patient.

In a memorable editorial toward the end of his life, the editor of the New England Journal of Medicine wrote (Ingelfinger 1980):

> I do not want to be in the position of a shopper at the Casbah who negotiates and haggles with the physician about what is best. I want to believe that my physician is acting under a higher moral principle than a used car dealer. I'll go further than that. A physician who merely spreads an array of vendibles in front of his patient and then says "Go ahead, you choose, it's your life" is guilty of shirking his duty, if not of malpractice.

Fairly or not, the metaphor of the used car dealer is not one of an individual with a reputation for trustworthiness. It is the image of someone who might sell the customer damaged goods, goods that would not be wanted if the details were known, yet within the contract of sale. *Caveat emptor!* Buyer, beware.

The Patient's Good: Autonomy and Trust

The health-care professional is expected to promote the patient's good or, for a libertarian, the patient's interests. Yet it has already been acknowledged that the professional may only qua professional know the patient's medical good or medical best interests. The good of the patient may extend far beyond such narrow confines (Pellegrino 2008a). Individual good may conflict with social good, good in certain relationships may differ from that in others, and the perception of short-term good may differ from the more considered longer-term view. Doctors are trusted to promote the patient's good but often have a limited view on what factors might make up that good. Prescribing a statin drug to lower cholesterol is one of the commoner long-term recommendations that doctors make to patients, but few

explain the exact benefits for the particular patient, and many probably do not even know them exactly. Even in the limited field of medical good, doctors may in practice be trusted beyond a reasonable limit, when they should not. Such medical good may well influence the estimate of other "goods" that the patient may make. This is particularly difficult if the patient has never been mentally competent and therefore always unable to express any preferences or values. The doctor's judgment will then be guided by accounts by others of, for example, the things that the patient appears to enjoy and what might be balanced against the aim to restore this capacity – the recent experiential interests. Even then, it is possible that the patient's interests may differ from those of family members, and in general, the interests that can be known – the medical good – may be the best guide. Nevertheless, if lifesaving treatment is advocated by the family, perhaps for reasons of love for a lifelong severely deformed or disabled adult, it may be judged that the costs – in every sense – may be excessive and burdensome for the patient. Doctors are trusted to take these decisions and to communicate them sensitively.

There is a large literature on the conflict between the good and the desired: between beneficence as paternalism and autonomy. Doctors are trusted some latitude to judge this. Informing an anxious patient of the possibility of a very unlikely life-threatening diagnosis may meet the conditions of a contract to give full information, but if the probability is low, a limited paternalism may be welcomed by the patient, if it were possible to know this in advance. The point surely is that the tacit presumption of full information should be interpreted in the concrete situation, where the presumption may be wrong. Trust gives the doctor the latitude to make that judgment: a limited paternalism and, some might argue, not really paternalism at all. We trust the doctor to inform us what we need to know to exercise choice or to understand why certain treatments cannot help us, but not to give a detailed list of everything inappropriate or irrelevant unless the patient requests more information. The more distress or even harm that might result should correlate with the strength of the doctor's conviction that this information is really wanted. As far as possible, the patient should set the agenda. But since information is inevitably infinite, somebody must judge what constitutes "full" information. Doctors are trusted to make that compromise because they are among those who "have the most experience of the subtle and paradoxical ways that human beings may react to illness and to fear; and who have had the greatest opportunity, from first-hand experience of when to speak out and when to keep silent" (Brewin 1985). But doctors "who *ignore* (my italics) the patient's notion of the good, violate the good of the patient as a self-determining rational being" (Pellegrino 2008b, 74).

The difficulty of medical practice in many Western societies is compounded by the pluralism of individual and social values. Multiculturalism has swept away some of the old assumptions. Nevertheless, the primacy remains the welfare of the patient: it is toward the healing of the patient that medicine aims, with an understanding of healing that extends beyond that of cure, or even alleviation, to the "good death." "Every art and every inquiry, and similarly every action and choice, is thought to aim at some good; and for this reason the good has rightly been

declared to be that at which all things aim. . .the end of the medical art is health" (Aristotle 1984, 1094a1). Palliative care adjusts healing acts to the good that can be achieved in the face of death. Care is never futile.

The greatest enemy of trust is deception. Deception is not a passive process, a lazy disrespect or error, and an ignorant assumption. It is an active process to lead to a belief that is not true, often a covert way of gaining an advantage or avoiding a responsibility. The rejection of deception is a fundamental human obligation that stems from the Kantian concept of a principled autonomy. The latter provides the ethical basis for trustworthy action, which in turn provides the evidence for trust itself.

Kantian autonomy deserves a fuller explication. It stems from Kant's focus on a test by which principles of action *could* be chosen by all, that is, fit to be universal laws. This is in contrast to concepts of individual autonomy, where the emphasis is on free choice, a form of individualism. Kant's autonomy is *principled* and set out as self-legislation. Fundamental ethical principles should presuppose what is required to be a principle for all (O'Neill 2002b). "Nobody who is committed to principled autonomy can make deception of others basic to his or her life and action because deception cannot serve as a principle for all." The effect of widespread deception would be catastrophic for trust. Some of the implications that follow are set out by O'Neill: refraining from lying, from false promising, from promise breaking, from misrepresentation, from passing off, and from plagiarism and, positively, for truthful communication, avoidance of exaggeration, simplicity, explicitness, and honesty in dealing with others. These are the qualities that contribute to trustworthiness. These qualities will not apply without exception. Civility demands some hypocrisies: gratitude may be expressed to the aunt who has given a ghastly necktie at Christmas; occasions for silence or discretion are recognized.

Paternalism is far from dead in British medical practice. Patients are often assured about a "little" problem and a "slight" discomfort and given overoptimistic descriptions of unpleasant procedures, such as colonoscopy (a flexible instrument passed up the bowel) or the claustrophobia of MRI scanning. It is all very well intended – or perhaps just medical ignorance. And there is still a prevailing belief in "consenting the patient," where that means getting a form signed. Euphemisms or anodyne descriptions are the soft end of deception, and the major deceits engaged in by the medical murderer are, of course, extremely rare. The response to a number of these major crises has probably been a contributory factor in maintaining the continuing high level of trust in health-care professionals in the UK.

Improving trustworthiness is a continuous process as medicine develops new procedures and opportunities for healing. Ethical principles require education and understanding but also confidence in judgment in their practical application to specific cases. A care plan for the terminally ill may have to meet many clinical, legal, and personal requirements, all of which must be balanced in sound judgment – not forgetting financial constraints. What ought to be done is always subject to what can be done. Over the last quarter of a century, guidelines, new laws, regulatory

bodies, and institutions have been created to improve patient safety and (in the case of data) privacy. Professional bodies provide interpretative guidance on many of these, as well as encouraging audit, appraisal, and openness. As O'Neill points out, these are not uniformly successful and some may have the reverse effect. League tables, ranked tables of particular measured standards of performance, are aimed to incentivize but may fail to measure what is important, and their results are widely misunderstood. Statistically, some institutions will do badly in one particular year for no better reason than the variation that can be predicted in any human institution. Fifty-percent will, of course, always be below average! The demands for accountability grow and the response itself may indicate that there is something wrong. Cassandra's misfortune, writes O'Neill, was that her prophecies were trustworthy, but still she was not trusted. Is more information or less required for trust?

The Paradox of Trust: Trust and Rationality

This seems to present a paradox. Character and qualities may be recognized by some form of connoisseurship, based on previous actions. Just as scientific knowledge is established by a process of induction, so too trustworthiness is established on a limited number of observations in a limited number of situations, and from these a conclusion is reached about the trustworthiness of the agent or institution. On this basis, with more experience, there will be a greater inclination to place trust in someone or some institution. Trust demands experience, and it would appear, the more experience, the better. Yet there is something unsatisfactory about the demand for more evidence. "If you really trusted me, you wouldn't want to see my documented credentials. You just don't trust me." There is something in this: moral sensibilities are affronted by the demand for more and more documentation or experience. A young child does not trust her parent because she has seen the parent defends or promotes her (the child's) interests in a new situation: she trusts – we might think, in O'Neill's terms – blindly. Nor is the persistently skeptical person appealing, who always seems to think the worst, ever suspicious of motives: the patient who records the consultation, not to ensure understanding by replaying the information but rather to check on the accuracy of everything said and to have the evidence to complain or prosecute if error has been made. Mostly such a patient who behaves badly, dislikes doctors, and takes some pleasure in the shortcomings of the doctor he is seeing might be judged to have an excessive concern with his own security and safety or to be someone who is overly cautious with others because he is bad himself (Baker 2008, 808).

On this basis, there does seem to be a return to some sort of Aristotelian balance in determining trust. There must surely be some grounds for placing trust, but this should be rather less than a demand for a veritable catalogue of evidence from different occasions. Is trust then more like faith? Some reasons are needed for placing faith in a person or indeed a religion, but there is not an expectation of what

might be construed as proof. Kant, in a different context, claimed to have disproved the existence of God to make way for faith. Affirming trust in someone or something is positive: "I trust him or her" is usually regarded as a compliment that most of us are pleased to hear.

Thus, a patient in an emotionally fraught and anxious situation may have difficulty trusting the doctor who appears too rational or difficulty doing so if appearing not rational enough. Reason may subvert trust. At the societal level (Fukuyama 1995), law, contract, and economic rationality are necessary but not sufficient basis for a society that wants to foster trust. There is also a need for individual reciprocity, moral obligation, duty toward community, and trust. The practice of such qualities is based on habit as much or more than calculation. Baker suggests that ordinary conceptions of rationality are not adequate to account for the phenomenon of trust (Baker 2008, 807). When something is believed, it is because the aim is truth. As she expresses it, it is rational to accept beliefs likely to be true. This implies that the reasons for our beliefs are those supporting the truth of what is believed. Yet, as observed above, the paradox seems to be that in some cases someone is trusted against the evidence. She proposes three varieties of trust. Firstly, in much daily life, things are just taken on trust. It is assumed that what has been told is true. Doctors, for example, do this most of the time listening to patients' stories or symptoms. There may be more skepticism in some situations than others, but all information cannot be checked. It must be trusted. But note that such trust is vulnerable to counter-evidence. Secondly, someone may be trusted because they just look honest or appear "decent." This is hardly an evidential standard that would ordinarily be accepted: rather the impressions are an evidence substitute. For example, people may trust professionals because they are smooth talking, courteous, and well presented. My mother trusted her orthopedic surgeon because he wore a white coat and was always courteous. Again, such trust is susceptible to counter-evidence. But Baker's third group presents us with the most interesting challenges. Here it may be found rational to hold beliefs *despite* counter-evidence and even find this praiseworthy. This she calls "special trust" or "friendship trust." If someone is trusted (and let us call them a friend) in this way, they will be supported in the face of accusations backed by evidence. It will be believed that there must be some explanation that exonerates the "friend." This is seen too in political life, where individuals have trust in the leader and believe that if only they could get through the lower-level officials, who present the evidence, to the top, the explanation would appear and the issue be resolved. There is no limit at which such trust may be broken in advance. One may know a "friend" is innocent because one knows one's "friend." It might be argued that this is not irrational because a person's character, motivations, and capabilities may indeed be known. Nevertheless, such knowledge does not conform to the usual ideas of rationality.

Of course, scientists may not reject a theory because of adverse evidence: rather they may seek more evidence to support the conjectures they have made. In any case, conjectures in the Popperian sense are not beliefs: they are proposals that the scientist seeks to falsify. One difference in personal trusting is that beliefs are

adjusted at a slower rate. In Baker's terms, trust *outruns* the evidence, the plain fact that should support trust. Such trust may not extend to others. An individual's trust may be limited if others perceive themselves at risk: an individual does not lock the silver up but may lock up someone else's, because he cannot be responsible for the other person's goods if that person doubts the facts.

It would trivialize trust to suggest it is merely a watered down version of full belief, acting "as if." A "friend" wants belief, not play acting. The claim to rationality could be abandoned altogether, but while a nonrational attitude may be helpful, it would eradicate trust if it has no rational basis.

Nor does it seem acceptable to accept inconsistency: "true belief is only one of a variety of needs." Rationality is not required in certain situations. This hardly seems satisfactory for an integrated life. A more fruitful approach may be to reconsider rationality.

Many moral judgments are based on what others may be expected to do. Trust is then indispensable for moral principles to take hold. It is essential for respect, which, from the perspective of Kant's moral philosophy, is essential to morality. Morality moreover demands trust in order for children to learn morality. Trust by children is not an "as if." People incapable of trusting could not engender trust in their children. Baker suggests that we reach a version of Pascal's wager: "despite our inability to establish the likelihood of our friend's decency, the belief itself is of enormous value." This is an end-directed rationality, but if the friend is honest with us, it is compatible with a truth-directed rationality too. Trust should therefore be looked at as an implicit "commitment, a state of conviction which is also an inclination of will," a hope "to be able to give a plausible account... of the way in which trust is subject to our control" and "insist that the rationality of trust is genuine rationality."

Trust risks disappointment, because life is not always predictable and even the most honorable people have their lapses. Uncertainty in medicine creates outcomes that may be not be predicted and still less anticipated. Patients and doctors still need to trust each other, for without trust, the best outcome cannot be secured. Better to be disappointed occasionally than not to trust.

Summary Points

- Trust is essential: without it, society could not function.
- Trust and trustworthiness are different concepts and either could exist without the other.
- Trust may be most needed when patients are most vulnerable and less able to make good judgments.
- Trustworthiness is a virtue.
- The mistrustful may enter an infinite regress of one person certifying the trustworthiness of another.
- Differing values between doctor and patient may lead to a contractual model of medicine and less trust.

- Deception is the greatest enemy of trust.
- Sometimes trust is irrational: it is given against the evidence.
- Occasional disappointment is better than never trusting.

References

Alper PR (1988) Trust at risk. West J Med 149:489–490
Anscombe GEM (1958) Modern moral philosophy. In: Crisp R, Slote M (eds) Virtue ethics. Oxford University Press, Oxford, pp 26–44, 1997
Aristotle (1984) Nicomachean ethics. In: Barnes J (ed) Complete works. Princeton University Press, Princeton 1094a1
Baker J (2008) Trust and rationality. In: Sosa E, Kim J, Fantl J, McGrath M (eds) Epistemology: an anthology. Blackwell, Oxford, p 808
Brewin T (1985) Truth, trust and paternalism. Lancet 8453:490–492
Caplan AL (1996) Odds and ends; trust and the debate over medical futility. Ann Intern Med 125:668–669
Ford J (2007) Public trust in doctors undented. BMJ 335:465
Fukuyama F (1995) Trust: the social virtue. The Free Press, New York, ch 2
General Medical Council (2013) Good medical practice. General Medical Council, London, p4, para 1
Hall MA, Dugan E, Zheng B, Mishra AK (2001) Trust in physicians and medical institutions: what is it, can it be measured, and does it matter? Milbank Q 79:613–639
Hall MA, Camacho F, Dugan E, Balkrishnan R (2002) Trust in the medical profession: conceptual and measurement issues. Health Serv Res 37:1419–1439
Hardin R (2005) Trust. In: Honderich T (ed) Oxford companion to philosophy. Oxford University Press, Oxford, p 926
Hillman AL (1996) Mediators of patient trust. JAMA 280:1703–1704
Dibben MR, Morris SE, Lean ME (2000) Situational trust and co-operative partnerships between physicians and their patients: a theoretical explanation transferable from business practice. QJM 93:55–61
Doctors in Society (2005) Medical professionalism in a changing world. Royal College of Physicians of London, p 2.8
Ingelfinger FJ (1972) Why trust the professionals? N Engl J Med 286:263–264
Ingelfinger FJ (1980) Arrogance. N Engl J Med 304:1507–1511
Ipsos Mori (2014) Trust in professions. https://www.ipsos-mori.com/researchpublications/researcharchive/15/Trust-in-Professions.aspx. Accessed 01 Aug 2015
Jones CN (2007) How much longer will patients trust us? BMJ 335:1226
Kirkpatrick E (ed) (1983) Chambers 20th century dictionary. Chambers, Edinburgh
Locke J (1663) Essays in the laws of nature. In: W van Lampen (ed) Clarendon Press, Oxford 1954. Quoted in: Hollis M. Trust within reason. Cambridge University Press 1998, Cambridge
Mayer RD, Davis JH, Schoorman FD (1995) An integrative model of organisation trust. Acad Manag Rev 20:709–733
Mechanic D (2004) In my chosen doctor I trust. BMJ 329:1418–1419
Mechanic D, Schlesinger M (1996) The impact of managed care on patients' trust in medical care and their physicians. JAMA 275:1693–1697
O'Neill O (2002a) A question of trust. Cambridge University Press, Cambridge
O'Neill O (2002b) Autonomy and trust in bioethics. Cambridge University Press, Cambridge. This section follows O'Neill's account, especially, pp 84–99
Parekh AK (2011) Winning their trust. N Engl J Med 364:e51
Pellegrino E (1991) Trust and distrust in professional ethics. In: Pellegrino ED, Veatch RM, Langan JP (eds) Ethics, trust and the professions. Georgetown University Press, Washington DC, pp 69–92

Pellegrino ED (2008a) Moral choice, the good of the patient, and the patient's good. In: Pellegrino ED, Engelhardt HT, Jotterand F (eds) The philosophy of medicine reborn. University of Notre Dame Press, Notre Dame, pp 163–186

Pellegrino ED (2008b) The internal morality of clinical medicine. In: Pellegrino ED, Engelhardt HT, Jotterand F (eds) The philosophy of medicine reborn. University of Notre Dame Press, Notre Dame, p 74

Rolfe A, Cash-gibson I, Car J, Sheikh A, McKinstry B (2014) Interventions for improving patients' trust in doctors and groups of doctors. Cochrane Database Syst Rev (3). Art No.: CD004134. doi:1002/14651858. CD004134.pub3

Shortell SM, Waters TM, Clarke KWB, Budetti PP (1998) Physicians as double agents. Maintaining trust in an era of multiple accountabilities. JAMA 280:1102–1108

Spirituality in Health Care

32

John Paley

Contents

Abstract

The discourse of spirituality-in-health-care (SIHC) exhibits a number of features that are not always present in the academic literature. One founding principle is that extravagant metaphysical claims can be made without having to be defended and in the expectation that they will not be challenged. These claims include the fundamental premise of SIHC discourse: everybody is spiritual, or has a spiritual dimension (the universality premise). While metaphysical claims are typically made in the absence of evidence or argument, a series of familiar epistemological tropes is used to secure an inviolate space in which challenges from a naturalistic perspective can be rendered otiose. At this point, SIHC discourse splits into two on mainly demographic grounds: an inflationary version in the USA and a deflationary version in the UK. Two distinctions between

J. Paley (✉)
Sheffield Hallam University, Sheffield, UK
e-mail: john.paley@btinternet.com

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_27

'spirituality' and 'religion' are adopted: inner/outer and broad/narrow, respectively. One interesting consequence of this demographic split is that the evidence for positive health outcomes as a result of religion/spirituality may apply only to the USA (and other religious countries), given the extent to which American culture is saturated with religion. Authors who adopt the broad/narrow distinction extend the denotation of 'spirituality' – what is to be deemed as an instance of either 'spiritual need' and 'spiritual care' – as much as possible in order that the universality premise will seem more plausible in a relatively secular society. This amounts to a classification project, an exercise in persuasive definition, in which the relevant 'deemings' require no defence, and in which a semantic bridge is constructed between the inflationary and deflationary poles. As a consequence, a discursive space is created and maintained for religious sensibilities in health care. The classification project is, for that reason, a broadly theological one.

Introduction: Two Literatures

Looked at from a certain angle, the literature dealing with spirituality and health divides into two. One is predominantly American, the other predominantly British. The two literatures share a fundamental premise, and have a common niggling worry; but they make different assumptions, refer to different contexts, and have different agendas. Contributing authors employ the same language, often submit to the same journals, and assume that they are engaged in the same debate. Much of the time, however, they are talking past each other, and there is not one debate but two, sometimes muddied and muddled by the misapprehension that everybody is talking about the same thing.

This chapter will try to disentangle the two literatures, and will consider several matters arising. It will ask a number of questions about the metaphysics, epistemology, and linguistics of spirituality-in-health-care, and it will make particular reference to the structure of discourse. Reference will be made to the evidence suggesting that spiritual practices and/or religious beliefs are associated with positive health outcomes, but only in the context of the philosophical topics which are the main focus of the discussion.

A useful start would be to consider the shared fundamental premise and the common niggling worry.

The Shared Fundamental Premise: Spirituality as Universal

The shared fundamental premise is simply stated. Everybody is spiritual, or has a "spiritual dimension." There are, of course, a number of variations on the theme. "Human persons are intrinsically spiritual" (Sulmasy 2002: 25). "Personhood is. . . an indivisible unity of body, mind, and spirit" (Hudson 2012: 05). "We are all

Spiritual beings trying to be Human" (Scottish Inter Faith Council 2008: 9). "We all have a spirituality and spiritual needs" (McSherry 2006: 59). "Spirituality is an essential element of humanity" (Puchalski et al. 2014: 10). "Spirituality is self-evident" (Cobb 2001: 11). "All people ... have an innate spirituality" (Gordon et al. 2011: 2).

In the discourse of spirituality-in-health-care (SIHC), this premise – the *universality of the spiritual* – is always stated or assumed. (One way in which it is assumed, illustrated by two of the quotations above, is through the apparently inclusive use of "we," "our," and "us," implying universality without making it explicit.) The universality premise is, as Cobb et al. (2012) suggest, the crucial axiom on which SIHC discourse is based. However, even when it is stated openly, it is always presented without evidence or argument. This gives rise to some important questions.

Consider the statement, "all people are spiritual" (Rumbold 2012: 181). What kind of claim is this? Is it an *empirical generalization*, akin to "all ravens are black"? If so, what evidence is offered in support of it? Or is it, perhaps, a *logical truth*, comparable to "all prime numbers greater than 2 are odd"? If so, can it be proved? Or is it an implied *stipulative definition*: "I will use the word 'spiritual' in such a way that it applies to everybody." (Compare: "I will use the word 'person' in such a way that it can be applied to the human zygote.") There have apparently been no attempts to provide empirical evidence in favor of the universality claim, and it does not seem capable of logical proof. So perhaps the stipulative option is the best bet.

The problem with this option is that it is a trifle flimsy, since it is the exercise of a semantic *fiat*. If you stipulate that "spirituality" will mean "valuing close relationships," or something like that, then of course everybody *is* spiritual, give or take the odd psychopath. But this is transparently a doubtful procedure: start with the conclusion, and then work back to a premise that will support it. "All people are spiritual" has to come out true, so "spiritual" is defined in such a way as to *make* it true. This attempt at reverse engineering renders the desired conclusion trivial and uninteresting. In any case, SIHC authors do not act as if they were merely stipulating. They talk as if they were struggling to elucidate how things are. The various attempts to conceptualize and measure spirituality imply that "spirituality/religiosity is a complex construct involving cognitive, emotional and behavioural aspects" (Büssing 2012: 329); so if this complex construct makes essential reference to the divine, or transcendence, or one's connection to nature, or the sacred, then a stipulative definition will look as if the author in question has ducked all the important questions.

However, some authors do seem to be implicitly proposing a definition of "spiritual." It is as if they were announcing, "The expression 'spiritual care' will be used in such a way that it applies to *w, x, y, z*...."; or "The phrase 'spiritual need' will be used in such a way that the following count as instances: *a, b, c, d*....." This is to propose a *classification*. It is to suggest that certain items should be classified as examples of *spiritual care* or *spiritual need*. But classification is never an ideologically innocent act. It is always motivated, and it is always designed to achieve a

particular outcome (Bowker and Star 2000). So it is legitimate to enquire what ideological function the statement "all people are spiritual" has if it is taken as a classification proposal. This "ideological function" question will resurface later in the chapter.

A different type of justification is sometimes implicit in references to "ancient wisdom." "The Ancient Wisdom traditions ... understood that mind, body, and spirit were all one... that mind, body and spirit are all interconnected" (Scottish Inter Faith Council 2008: 27, 10). This suggests that one argument for "all people are spiritual" is an appeal to authority, in this case an ancient (though unspecified) authority. Admittedly, references to "Ancient Wisdom" are somewhat unexpected in an "Information Resource for Healthcare Staff" issued by NHS Scotland and are difficult to reconcile with the idea of evidence-based medicine. But it is hard to find anything else in the literature that counts as even a weak argument in favor of the universality principle.

Some writers are content to note that most people agree with them. "Spirituality is commonly... accepted as a universal human phenomenon" (Bruce et al. 2011: 44). "It is now almost axiomatic in healthcare circles that 'all people are spiritual'" (Cobb et al. 2012: 487). This is another weak argument, akin to the *Consensus Gentium* argument for the existence of God (Kelly 2011): most people think there is a God, so there probably is one. It is not an appeal-to-authority argument but a clearly-the-majority-must-be-right argument. However, the audience for SIHC discourse accepts the universality principle for the same reasons – or lack of them – that SIHC writers do. To that extent, authors can justifiably assume that their readers will not worry about the absence of evidence or argument. Together, they form a cadre of spirituality enthusiasts, all of whom regard the universality principle as self-evident.

But it is not self-evident. In a multistage survey of 7403 English adults, 46 % of those interviewed did not identify as either religious or spiritual in outlook, and claimed not to have a religious or spiritual understanding of their lives (King et al. 2013). It is highly unlikely that many of those 46 % would agree with the statement: "all people are spiritual." It is far more likely that some of them would find the statement patronizing because it asserts *ex cathedra* that they *are* spiritual, in spite of their denials.

The universality principle is the fundamental premise of SIHC discourse, but the literature is bereft of arguments or evidence in its favor. The fact that it is now axiomatic in health care circles might be an example of a familiar apophthegm: if something is repeated often enough, people will eventually come to believe it.

The Common Niggling Worry: Spirituality and Religion

There is clearly some sort of connection between spirituality and religion if only because, historically, the term "spiritual" belongs to the discourse of specific religions (and, in particular, Christianity). Until recently, major reference works on religion (Eliade 1987; Bowker 1997) did not discuss spirituality in a generic

sense, divorced from specific religious traditions; and according to one historian/theologian (Principe 1983), "spirituality" only began to detach itself from its associations with Christian mysticism, piety, and the contemplative life during the 1950s.

The first book to use the contemporary language of spirituality is James (1968), a text permeated with references to "a search for meaning and significance," "wholeness," "the totality of human experience," and "spiritual life *as* life." Following the publication of this volume, it became possible to argue that spirituality is no longer "associated exclusively with any one Christian tradition, nor even necessarily with Christianity as a whole" (Sheldrake 1991: 50). The book marks the point at which spirituality is reinvented as a generic concept and, according to some authors, opens for business as a "kind of radical individualism that tends to elevate the self to a cosmic principle" (Bellah et al. 1985: 236).

But it also marks the point at which the relation between spirituality and religion becomes ambiguous. "With the emergence of spirituality, a tension appears to have risen between the constructs of religiousness and spirituality" (Zinnbauer and Pargament 2005: 24).

On the one hand, spirituality has achieved independence – a sort of relative autonomy – from religion. It is "understood to be a broader concept which includes, but is not confined to," religion (Gordon et al. 2011: 57) and "is not necessarily anchored in religion" (Stanworth 2004: 1). For some writers "broader" is something of an understatement, since the range of things that have been classified as "spiritual" is impressively wide. It includes humanistic and Jungian psychology, complementary therapies; the Tao and Buddhism; art, poetry and music, the contemplation of nature; the search for meaning and purpose, values; ecological concerns, political ideals, giving physical care; relationships, work, domestic chores, sport; unity, connectedness, transcendence (Roof 1999; Carrette and King 2005; Flanagan and Jupp 2007). From a SIHC perspective, *spiritual needs* can include the need for touch, the need for relationships, the need to nurture, the need for belonging, the need to experience oneself as a separate being. the need for affirmation. Examples of *spiritual care* include listening to the patient, instilling hope, shared decision-making, reflective practice, experiential learning, valuing staff, informing the patient of local resources, and cleaning the toilet seat for the next person (Paley 2015).

On the other hand, the relative autonomy is only relative. In many respects, spirituality and religion are still yoked together.

Consider, for example, language. In American English, about 67 % of all references to spirituality occur in religious contexts; in British English, the figure is 50 % (see the Corpus of Contemporary American English and the British National Corpus: Davies 2015). In sociolinguistic terms, the divorce is a long way from being complete. Moreover, discussions of spirituality in a health care context often use language which it is difficult not to see as religious, and specifically Christian. Words and phrases such as "sacred," "soul," "the ultimate," "transcendence," "higher power," "redemption," and "something bigger than ourselves" betray their origins in Christian theology. Writers who use this type of

language do not always notice its theological watermark. They assume that it makes as much sense – the same *kind* of sense – to unbelievers as it does to them. Their attempt to be inclusive, by attributing spirituality to everybody, is compromised by the fact that their descriptions of it trade on religious concepts, even as they insist that spirituality and religion are nonidentical. Even a word like "suffering," which at first sight seems theologically neutral, carries a Christian undertow, given the significance of suffering in the Christian narrative and in theological understandings of illness (Ford and Muers 2005).

For many authors, "the terms religion and spirituality are used more or less interchangeably" (Seybold 2012: 347); and it is routine to see expressions such as "religious and spiritual beliefs" (Puchalski et al. 2009), "spirituality/religion" (Burke and Neimeyer 2012), or "R/S" (Park 2007), as if the two terms were not worth distinguishing. Psychologists "traditionally regarded religion as a 'broad-band' construct, not explicitly differentiated from spirituality" (Zinnbauer and Pargament 2005); and although they do now distinguish between "religiousness" and "spirituality," the proposed distinctions tend to be a matter of nuance. According to Hill et al. (2000), the same basic concept lies at the core of both constructs – the *sacred*. Pargament (1997), for example, defines spirituality as a "search for the sacred" and religion as a "search for significance in ways related to the sacred." For writers of this persuasion, it would seem that spirituality has inherited the quiddity of religion – its DNA, so to speak – even if a declaration of independence has been made on its behalf.

So the common niggling worry is where to draw the boundary. Is spirituality an inflationary concept? That is, does it make *essential* reference to transcendent realities, the sacred, higher powers, the infinite, and cosmic forces, even if there is no mention of God, creeds, and rituals? Or is it deflationary? Does it confine itself mainly to relationships, an appreciation of art, literature, and music, the contemplation of nature, listening to the patient, and reflective practice (even if it acknowledges that, for some people, a relationship with God might be just as important as a relationship with members of their family)? Is the supernatural an *intrinsic* feature of spirituality (inflationary) or is it just one option among many others (deflationary)?

The sociological fact lurking beneath the inflationary/deflationary distinction is this. The US literature is largely inflationary; the UK literature is largely (but not exclusively) deflationary.

Deflationary and Inflationary

By any standard, according to Putnam and Campbell (2010), "the United States (as a whole) is a religious nation… [and]… ranks far ahead of virtually all other developed nations" (7–8). Belief in God ran at a steady 94 % between 1947 and 2001 (Norris and Inglehart 2004), and 80 % are absolutely *sure* that there is a God. Weekly church attendance is between 25 % and 40 %; 65 % of the population are members of a church or synagogue, with 38 % of them being active members

(Presser and Chaves 2007). A mere 18 % of Americans admit they never pray; a third believe that scripture is the actual word of God (Putnam and Campbell 2010).

In the UK, even if we combine the figures for belief in a personal God and belief in a "higher power," the total is no more than 47 % (Bruce 2002). Church attendance is estimated at around 7.5 % (Heelas and Woodhead 2004) and church membership at 10 % (Brierley 2000). In contrast to the USA, 54 % of the UK population claim that they never pray; only 9 % believe that scripture is the actual word of God (Putnam and Campbell 2010).

These are striking differences. But even more important, from the perspective of this chapter, is the fact that, in the USA, spirituality and religion are barely differentiated. Sixty percent see themselves as both spiritual *and* religious (Marler and Hadaway 2002); and while "four percent of the least religious describe themselves as very spiritual, 80 % of the most religious do" (Putnam and Campbell 2010: 21). The same writers add: "Among rank-and-file Americans spirituality and religiosity go hand in hand."

Again, the contrast is marked: as observed earlier, 46 % of English adults are neither religious nor spiritual (King et al. 2013), compared to a mere 12 % of Americans.

Failure to acknowledge these differences can lead to misconceptions. For example, Dyson et al. (1997) report that when Americans were asked to define spiritual well-being, "the majority of responses were given in terms of religious faith." They describe this as a "confusion," an observation which appears to betray unfamiliarity with the extent to which, in the USA, spirituality and religion are associated. By the same token, many American authors generalize from US data, apparently forgetting that in some other countries spirituality and religion are not as closely intertwined. For example, following a review of the American evidence, Zinnbauer and Pargament (2005: 29) conclude, "From these studies it appears that most people view themselves as both religious and spiritual." The "most people" is unqualified.

The universality principle will seem more plausible in the USA, where 80 % of the population see themselves as spiritual, than it will in the UK, where at most 54 % do. Furthermore, there will be a greater tendency to tie spirituality and religion together in the USA, where 60 % see themselves as both spiritual and religious, than in the UK, where the corresponding figure is no more than 35 %. It is likely, then, that authors in the UK will find it harder to convince British readers that "all people are spiritual" (in view of the greater reluctance of UK citizens to identify with a spiritual outlook on life) and will have to pursue a distinctive strategy in order to do so. It is equally likely that an inflationary view of spirituality will be more characteristic of American authors than of British authors, in view of the stronger association between spirituality and religion in the USA. Equally, it is probable that British writers will employ a deflationary strategy in order to make the universality principle more plausible. A view of spirituality in which an appreciation of nature is classified as a spiritual experience (at least for some people) will seem more plausible in the secular UK than a view which involves essential reference to various supernatural entities, the sacred, transcendent realities, and cosmic forces.

The numbers merely summarize the extent to which religion/spirituality saturates American culture, in contrast to more secular countries. Zuckerman's (2008) comparison of religiosity in the USA and the near lack of religiosity in Denmark and Sweden is illuminating in this respect, since he is an American sociologist accustomed to the view that a society without religion would lead to anarchy. "I spent my sojourn in Denmark in a relative state of awe," he says, noting that in the USA religion is ever present in the media, sporting events seldom begin without prayer, 64 % of Americans agree that politicians who don't believe in God are unfit for public office, 88 % believe in heaven, and 75 % in hell. In the USA it is generally assumed that morality would disappear without religion, that belief in an afterlife sustains those facing death, and that God provides answers to existential questions. In the USA, a bank manager can publicly advise one of his customers that turning to God will clear her debts, and nobody finds it strange or unusual. As Zuckerman, echoing Putnam and Campbell (2010), observes, "This is one religious country" (Zuckerman 2008, *passim*, 168).

None of this applies to Denmark, Sweden, or other European secular states (Norris and Inglehart 2004; Inglehart et al. 2004). Only 3 % of Danes go to church every week, and a mere 8 % agree that godless politicians are unfit for office. Only 18 % believe in heaven, and only 10 % in hell. In Denmark it is atheists who have an easier time facing death, and most people happily accept "that there is ultimately *no* meaning to life" (Zuckerman 2008: 5, 25). The vast divide between the religious USA and relatively nonreligious Europe should not be underestimated. As one Dane, who is now resident in California, noted, "American society is – all politics and media discussions – is based on that everybody is very devoted Christians" (181). He thinks that politicians praying to God is "just scary"; as for a recent president's claim that God talked to him, "that could be considered being a little mentally ill" (178). It is evident that secular Europe and religious America find each other almost incomprehensible.

The SIHC literature echoes this mutual incomprehensibility. American writing is largely inflationary. At its most extreme, it is explicitly and (to the secular mind) extravagantly supernatural. "Spirituality refers to the domain of spirit(s): God or gods, souls, angels, jinni, demons – and only by metaphorical extension to other intangible and invisible things" (Hufford and Bucklin 2006: 29). In contrast, writing in the UK is largely deflationary, inclined toward the mundane rather than the supernatural, and with a greater tendency to focus on relationships, the arts, nature, and daily tasks. "It is often the mundane rituals such as going to work, doing the washing or walking the dog that bring meaning and purpose to everyday life" (McSherry 2006: 49). It is from this perspective that there are two different literatures: one inflationary and largely American, the other deflationary and largely British.

If two distinct conceptions of spirituality appear to be on offer, inflationary and deflationary, one might expect them to be associated with two opposed strategies when the task is to define the relation between spirituality and religion. This expectation appears to be fulfilled in the literature.

Two Distinctions

Many inflationary writers, as already noted, make no distinction between "spiritual" and "religious" and treat them as interchangeable terms. This is not surprising, given the close association between religion and spirituality in the USA. Other inflationary authors, however, gravitate toward what might be called an *inner/outer distinction*. Variations on this theme include personal/institutional and private/public, the aim being to distinguish between a personal, subjective orientation and a public, observable orientation. The former is typified by personal beliefs and feelings. The latter is characterized by creeds and rituals. According to Walker and Pitts (1998), for example, spirituality is seen as a "personal affirmation of the transcendent"; in contrast, religion is regarded as "the creedal and ritual expression of spirituality that is associated with institutional church organizations" (409). Although this distinction has been criticized (Zinnbauer and Pargament 2005), the inflationary literature still tends to depict religion and spirituality as alternative, but not mutually incompatible, expressions of some underlying impulse (Oppy 2012). This impulse might be a "search for the sacred" (Hill and Pargament 2003), "belief in a higher power" (Walker and Pitts 1998), or something else which can be conceptualized as the common denominator of religious and spiritual life.

The deflationary literature adopts a *broad/narrow distinction*. The idea is that "spirituality" encompasses a very wide range of possibilities: virtually anything that people find emotionally satisfying, fulfilling, awe inspiring – or anything in which they find "meaning." At its most elastic, as observed earlier, "the spiritual" embraces virtually every aspect of human experience, confining the nonspiritual and secular to a "concern with fitted kitchen units and grouting" (Bruce 2002: 200). "Religion," on the other hand, is just one of the ways in which people can achieve satisfaction, meaning, or fulfillment. Historically, it was more important than it is now, and it is still salient for a large number of people; but it has become no more than a single option on the spiritual smörgåsbord. "Religion provides particular ways of making meaning" (Swinton 2012: 101); but "spirituality" offers a huge array of alternatives: creativity, nature, relationships, values, culture, and domestic chores. It is in a direct line of descent from "the totality of human experience," and "spiritual life *as* life" (James 1968).

There are, as one might anticipate, inflationary criticisms of deflation, and deflationary criticisms of inflation. Here is one inflationary complaint: "behaviors or lifestyles are not spiritual simply because they serve an integrative function in life. To say 'I find my spirituality in gardening' or 'Music is my spirituality' might indeed suggest that the person finds a great satisfaction and subjective well-being through gardening or playing music. . . but unless such lifestyles are responses to a perception of the Sacred. . . then it is inappropriate to refer to gardening or music as 'spiritual'" (Hill et al. 2000: 65). The irony is that Hill et al. take gardening and music as a sort of *reductio ad absurdum* of the "broad" conception of spirituality. Inflationist: "Define 'spiritual' in that way, and even gardening is spiritual!" Deflationist: "Yes, exactly!"

On the other side of the fence, deflationary authors suggest that the "inner" conception is discriminatory because it leaves too many people out. "If spirituality is defined only synonymously with religion and belief in God, then several persons, namely the atheists, agnostics, humanists and hedonists would be excluded from the possibility of using spiritual coping mechanisms. Therefore, spirituality applies to both believers and non-believers" (Baldacchino and Draper 2001: 835). For demographic reasons, the inner/outer distinction and the "perception of the Sacred" requirement are not plausibly consistent with the universality principle in the UK. So British authors need the "broad" conception in order to ensure that the maximum number of people can be classified as "spiritual."

One author attempts to have it both ways. For patient care, spirituality should be defined as "broadly as possibly so that all patients have an opportunity to have their spiritual needs addressed" (Koenig 2008: 18). However, for research purposes, he advocates that we "return the definition of spirituality to its origins in religion. . . If there is no connection with either religion or the supernatural, then I would not call a belief, practice, or experience spiritual. I would call it humanistic" (16–17). This is a somewhat uncomfortable position, given that research on spirituality in health would presumably require constant oscillation between the two definitions.

Research on Health and Religion/Spirituality

There is now an enormous empirical literature on religion and health. A good proportion of it appears to demonstrate that, in one form or another, religion and/or spirituality have a positive influence on health and well-being (Koenig et al. 2012). However, much of the research in this area is controversial because the findings are inconsistent, and because there are methodological weaknesses (it is argued) in many of the studies which seem to suggest health benefits.

There are also metadebates about the criteria used to select studies for systematic review. For example, in a review by Powell et al. (2003), which draws broadly skeptical conclusions concerning the claim that religion/spirituality brings positive health outcomes, exclusion and inclusion criteria were based on Cochrane Library techniques. However, a review which is more sympathetic to the "positive outcomes" claim describes the Cochrane criteria as "very conservative" and "a priori" (Oman and Thoresen 2005: 442), despite acknowledging that they have been "influential in medical research." In this domain, it appears, one writer's "minimally acceptable methodological standards" (Powell et al. 2003) are another writer's overly "conservative" criteria.

Other notable reviews in recent years are Chida et al. (2009) and Masters and Hooker (2015). However, this is not the place to conduct a full examination of the evidence, or assess the extent to which religion and spirituality produce health benefits. It *is* the place to ask whether the evidence might be influenced by cultural factors.

The most salient point is that research has focused almost exclusively on Christian populations in the USA (Abu-Raiya and Pargament 2012). One

consequence of this is the tendency to run religion and spirituality together. For example, Chida et al. (2009: 81) adopt the approach suggested by Hill et al. (2000): "Religiosity and spirituality can be defined broadly as any feelings, thoughts, experiences, and behaviors that arise from a search for the 'sacred'... the former implying group or social practices and doctrines and the latter tending to refer to personal experiences and beliefs." It is no surprise, then, to find a corresponding terminology in the reviews already cited: "religiosity/spirituality," "religion or spirituality," "RS," and "R/S."

Two general observations can be made about the research literature. First, the most convincing evidence of a connection between religion or spirituality and health outcomes (in the USA) concerns attendance at church or other religious services. In particular, there is a consistent finding that weekly attendance is associated with a reduced risk of mortality (Powell et al. 2003). For other independent and dependent variables, the results tend to be far more varied. Though recent studies have confirmed the "attendance" finding, it is not clear why regularly attending *religious* gatherings – as opposed to social gatherings of other kinds – has a protective effect (Masters and Hooker 2015).

Second, the reduction in mortality occasioned by religion or spirituality applies almost exclusively to healthy populations. There appears to be little or no protective effect in illness populations, suggesting that religious/spiritual involvement may be more important for health maintenance than for ameliorating existing disease.

The implications of these two observations are interesting. First, attendance counts as "religion" rather than "spirituality," whether the inner/outer or the broad/narrow distinction is adopted. Yet the American literature standardly classifies it as religion/spirituality or R/S. For example, Masters and Hooker (2015), having discussed several studies suggesting a connection between attendance and mortality as a result of cardiovascular disease, observe that "involvement in healthy behaviors accounts for some but not all of the beneficial association between R/S and CVD mortality" (524–525). An extrapolation from church attendance to "R/S" is a regular feature of the American literature, even though studies of attendance do not necessarily imply anything about nonreligious spirituality in either the "inner" or "broad" senses.

Second, studies which imply that healthy populations benefit from the protective effect of attendance do nothing to suggest that "spiritual" interventions will have a beneficial effect on people who are already ill. Instead, they suggest that church attendance should be the concern of public health initiatives, and not individual encounters between patients and chaplains in, say, a hospital. Whether promoting church attendance would be regarded as a legitimate use of public health resources is another question.

As already noted, the evidence for other positive outcomes is controversial. Putting these reservations aside for a moment, though, suppose that there were solid evidence not only for the attendance effect but for other religious/spiritual beliefs or practices bringing health benefits (not just reduced mortality), at least in the USA. What conclusions might be drawn from this?

Whatever the explanation of such findings turned out to be, it is probable that it would be most salient in cultures saturated with religion, like the USA. In a species as highly social as ours, what other people do, think, and believe is of overwhelming importance, to the extent that societies engage in systematic "mindshaping" in order to ensure the greatest degree of cognitive homogeneity (Zawidzki 2013). It is, of course, well known that socially integrated people live longer and have increased resistance to disease. The range of proposed mechanisms includes social comparison, social influence, self-esteem, sense of control, social support, belonging, and stress-buffering (Thoits 2011). Moreover, it is established that the presence or absence of social contact modifies activity in neural and endocrine systems affecting disease pathophysiology (Eisenberger and Cole 2012). However, the threat to survival as represented by a disengagement from, or reduced commitment to, beliefs and practices of a kind that saturate the culture to which the person concerned belongs may be an additional risk factor, over and above a lack of social support or companionship.

Consequently, in a religious culture, relative detachment from religious beliefs, church attendance, and other forms of affiliation are likely to be more deleterious to an individual's health than they might be in a secular culture. In religious cultures, "coming out" as an atheist can precipitate discrimination, social ostracism, a wholesale breakdown in family relationships, and other situations leading to severe stress (Zimmerman et al. 2015). In contrast, regular attendance, beliefs in accordance with social norms, and participation in religious/spiritual practices are likely to protect against negative health outcomes, just as the evidence from the USA (*ex hypothesi*) suggests. It is far less likely that any of this applies to secular cultures.

There is now evidence that the association between religion/spirituality and positive health outcomes, typical of the USA, does not generalize to secular societies. Attempts to replicate American research have produced mixed results in Europe; several studies based on the World and European Values Survey (WVS and EVS), and the Survey of Health, Ageing and Retirement in Europe (SHARE), have shown that religious commitments do enhance physical health, mental health, and life satisfaction – but only in countries where such commitments are the norm (Stavrova et al. 2013; Lun and Bond 2013). For example, frequency of prayer is associated with *decreased* mental and physical health in nine European countries (Hank and Schaan 2008); according to a study of 86 countries and 280,437 respondents, regarding God as an important part of one's life predicted *worse* health in secular societies (Hayward and Elliott 2014). In a recent UK study, "people who had a spiritual understanding of life had *worse* mental health than those with an understanding that was neither religious nor spiritual" (King et al. 2013: 71).

In summary, the best current evidence suggests that *attendance* at religious services *reduces mortality* in *healthy* populations, in highly *religious societies*. What does this conclusion imply about spirituality of a more subjective or "transcendental" kind? Ian Hacking (1983: 146) says, "We shall count as real what we can use to intervene in the world to affect something else, or what the world can use to affect us." In other words, whatever has causal consequences is real. If this

criterion is adopted in the SIHC context, the evidence suggests that the *real* is confined to attendance at religious services and the social contact associated with it. If two further conditions are met – continued health in the individual and a society in which religious beliefs are normative – then regular attendance at such services will have an observable effect: it will reduce the risk of death. Subjective and "transcendental" spirituality, on the other hand, do not appear to fulfill Hacking's criterion. There is no evidence that they have any causal consequences for nonhealthy people in nonreligious societies.

However, inflationist writers will probably reject Hacking's criterion. Many of them are skeptical about the application of causal concepts to this field, and some are prepared to countenance "super-empirical" entities and mechanisms, or "subtle energies that are beyond current modern scientific understanding" (Oman and Thoresen 2005: 440). In the next section, some examples of these metaphysical claims will be considered.

Inflationary Metaphysics

One feature of inflationary discourse is the extent to which authors grant themselves a license to make extravagant metaphysical statements without evidence or argument. Very roughly, these fall into three categories: statements about the self, statements about the universe, and statements about the relation between the self and cosmic forces. Typically, these statements are made without any reference to work done in what would appear to be the relevant disciplines – for example, psychology, cognitive science, neuroscience, philosophy, anthropology, biology, cosmology, astrophysics, and quantum mechanics.

For example, Hudson refers to the "unity of the human person in the totality of their being… A person is an indivisible unity of body/mind/spirit (or soul)… The patient is an embodied soul or ensouled body" (Hudson 2012: 106). The "indivisible unity" idea is popular in SIHC writing, but it is not transparent. Although Hudson contrasts this unity with Cartesian dualism, it is not clear what sort of unity she has in mind.

One well-canvassed alternative in modern philosophy is physicalism (Kim 2005). The person is a unity in the sense that she is completely physical – that is, physical without remainder. Every aspect of her existence, mind as well as body, is physical or supervenes on the physical. Physicalist philosophers do not typically talk about "spirit"; but, if they did, they would no doubt argue that spirit is physical as well. The suspicion is that Hudson would not accept this view; some inflationist writers explicitly reject it. It is hard to be certain, however, because she does not discuss it. Nor does she say what other kind of unity she is referring to.

Perhaps Hudson believes that body, mind, and spirit are all different substances (so that physicalism is a nonstarter) but that they are nevertheless inseparable, and the person is indivisible in this sense. There are three distinct "components," which are somehow interwoven. Quite clearly, a lot of work would be required to make this position look plausible and to explain how it is supposed to work. In what sense

are the three components independent substances? In what way(s) do they differ? How are they joined together? Why can they not be separated? How do they interact? Hudson does not explore any of these questions. Nor does she comment on the awkward fact that, despite her rejection of Cartesian dualism, Descartes' position was that a person is a single thing, a "substantial union" of distinct but nevertheless inseparable substances.

Many other questions suggest themselves. For example, is Hudson's thesis compatible with cognitive science evidence suggesting massive modularity of mind (Carruthers 2006)? *Prima facie*, the idea of an indeterminately large number of relatively independent cognitive modules, each dedicated to a specific function and dissociable from other modules, is inconsistent with the idea of a unified, integrated self. Similarly, the "indivisible unity" thesis appears, at first sight, to be incompatible with the evidence that moral judgment is not unified but resolves into a number of different systems (Cushman et al. 2010). But is this first impression warranted? Again, there is no indication of how Hudson would deal with this question. She is content to talk about "indivisible unity" without explaining what sort of unity this is.

As a second example, consider Swinton's (2012) reference to "hypothesized mechanisms, such as the existence of healing bioenergy. The literature within the area of prayer studies is indicative of the possibility that there may be supra-empirical dimensions to religion and spirituality that are currently not understood, but which may have healing capacities" (102). Here, Swinton apparently overlooks the findings of the most methodologically robust prayer studies, STEP and MANTRA II, in which no health benefits of intercessory prayer were identified (Sloan 2006). Even authors sympathetic to the potential efficacy of prayer concede that "there is little evidence that prayer is an effective treatment intervention" (Jantos 2012: 361). So the "healing capacities" of prayer can still legitimately be doubted.

More significantly, it is not clear what "healing bioenergy" is. There is nothing wrong with the concept of hypothesized mechanisms; but "healing bioenergy" does not itself refer to a mechanism. It is more a vague gesture toward an undefined something-or-other. "Bioenergy" is certainly a legitimate term, but it refers to the generation of sustainable energy from biomass in order to reduce CO_2 emissions (Bauen et al. 2009). In this context, the expression "healing bioenergy" does not make much sense. Presumably, then, Swinton does not have this use of "bioenergy" in mind. Unfortunately, he offers no explanation of what he does have in mind, and makes no comment on how his intended usage of "bioenergy" differs from the orthodox use. Nor does he provide an account of supraempirical dimensions.

Like many other inflationary writers, Hudson and Swinton apparently assume that they have a license to make extravagant metaphysical claims without the need to explain or justify them. De facto, of course, they *do* have such a license, since other SIHC writers rarely (if ever) challenge these claims. This is a founding principle of SIHC discourse, rather like the universality principle: statements about the unity of the self, cosmic forces, supraempirical entities, healing bioenergy, the connection between the self and universe, and so on, can be made

(a) without explanation or defence, (b) in the expectation that they will not be challenged.

In principle, however, these claims remain vulnerable to skeptical questions from writers less inclined to metaphysical extravagance. For this reason, they are often accompanied by one or more epistemological tropes designed to act as a philosophical firewall.

Inflationism's Epistemological Tropes

The reference to subtle energies "beyond current modern scientific understanding" (Oman and Thoresen 2005: 440) is an example of a familiar epistemological trope whose function is to place the study of spirituality outside the purview of science. The point of this maneuver is to secure a discursive space in which extravagant metaphysical claims can be made – a space to which scientific methods do not, by definition, apply, so that the metaphysical claims are protected from empirical testing and interrogation. It delineates a boundary beyond which science is impotent by epistemological *fiat*. Oman & Thoresen's reference is a relatively weak version of the trope, as it suggests that the subtle energies are *currently* beyond scientific understanding. A stronger version would suggest that these energies, supraempirical entities, or the spiritual realm will *always* be beyond scientific understanding. In SIHC discourse, this epistemological maneuver is often unsupported by argument. Rather like the universality principle, it is regarded by many SIHC authors as self-evident.

Not always, however. It is sometimes argued that science has a number of characteristics that disqualify it from investigating spirituality. These characteristics include reductionism, linearity, the requirement of objectivity, the goal of arriving at universal laws, falsifiability, replicable experiments, empiricism, and so on – all of which (it is suggested) are incompatible with the type of knowledge that is associated with spirituality. What is noticeable about these suggestions is that they rarely cite work in philosophy of science. Just as metaphysical claims are made without reference to psychology, philosophy, biology, or physics, so epistemological claims about science are made without reference to the relevant scientific and philosophical literature.

To the extent that influences can be determined, they belong predominantly to positivist interpretations of science, now long out of date. The underlying logic of these arguments is: the *positivist account* of scientific method does not fit our understanding of spirituality; therefore, the *scientific method* does not fit our understanding of spirituality; therefore, spirituality must be beyond scientific enquiry. To a great extent, it is the misdescription of scientific endeavor – judged by the achievements of the more recent philosophy of science – that gets the epistemological trope up and running.

Swinton (2012) again provides an excellent example of this logic. He takes falsifiability to be a criterion of science; but he assumes that it is the requirement that something can be "unquestionably disproved." It would be difficult to justify

this assumption on the basis of the philosophy of science literature. Even Popper accepted that a failed prediction may be the result of getting the mathematics wrong, accidental disturbances, instruments not working properly, bungling, incorrect auxiliary hypotheses, unsuspected misconceptions about how the instruments work, and other possibilities (Mayo 1996). All of these have to be investigated and checked; but it is never possible to rule out the alternatives *completely*. What is true is that the evidence can mount up, and further investigations can fail to identify any of the above errors. Eventually, the probability that such errors exist – despite all the failed attempts to locate them – drops below a certain conventionally determined threshold (although it never falls to zero). At which point, it is agreed that, pending further data, the hypothesis can now be rejected.

This is a long way from disproving something "unquestionably." So when Swinton suggests that "there is a God" or "I love you" are not falsifiable, because they can never be "unquestionably disproved," he is under a misapprehension about scientific procedure. Even if these statements cannot be "unquestionably disproved," the evidence against them can mount up until it becomes unreasonable to continue believing them. In the case of "I love you," for example, the evidence might include adultery, deception, hostility, abuse, and the discovery that one's partner has been saying "I love you" to someone else. Investigations can be made to determine whether there are other explanations; but there comes a point with love, as with scientific investigation, when the hypothesis can justifiably be rejected. (Some writers, of course, would make exactly the same point about "there is a God.") At any rate, the idea that the scientist lives "in two knowledge worlds" because "I love you" is something she knows to be true *intuitively* rather than scientifically (Swinton 2012: 100) is clearly open to question.

A related epistemological trope refers to "ways of knowing." For example, Cobb et al. (2012: 489) talk of "the ways we know about spirituality," which include narrative, "hermeneutic ways of knowing," and "research methods and paradigms that are congruent with the spiritual realities," while Swinton says that spirituality is itself "a form of knowledge." In this respect, the SIHC literature echoes claims that there are different paths to knowledge – paths which adopt nonscientific methods of "knowing." There are, it is suggested, women's ways of knowing, nurses' ways of knowing, embodied ways of knowing, and so on. These "ways of knowing" often have what Sloan (2006: 11) calls a "reverence for the subjective as a source of truth"; and most do not appear to possess a means of detecting and correcting error, which is arguably a necessary condition for anything that counts as a "way of *knowing*" (Mayo 1996) as opposed to a "way of *coming to believe*."

Placing spiritual matters beyond the scope of science by fiat, listing characteristics of scientific enquiry that supposedly make it unsuitable for investigating spirituality, suggesting that there are other ways of knowing are all examples of epistemological tropes that regularly accompany metaphysical claims. Their function is to secure a discursive space in which those claims are inviolable. Since any challenges will involve requests for evidence, or a precise specification of what the claim amounts to, they will presuppose scientific methods that are not "congruent with spiritual realities." On that basis, inflationary authors feel able to dismiss them.

Deflationary Deeming

If inflationary discourse is marked by extravagant metaphysics and a firewall of epistemological tropes, deflationary discourse is marked by a "stretch dynamic" (Paley 2008) that increases the extension of the words "spiritual" and "spirituality" – that is, the range of things they apply to – as widely as possible. This is a requirement of the universality principle, since in the UK and Europe the claim that "all people are spiritual" will be unpersuasive unless "spiritual" is defined in such a way that it designates situations, activities, and psychological states which fall well short of a commitment to "the supernatural" or "the sacred." As a result, deflationary contributions to the SIHC literature typically propose, or presuppose, claims about what shall count as "spirituality," "spiritual need," or "spiritual care." As noted earlier, the underlying logic of these contributions is to create and sustain a particular *classification*. Deflationary discourse is an exercise in the politics of meaning.

What is absent from this literature, however, is any theory of "deeming," that is to say, an account of the criteria according to which a need can be *deemed* a "spiritual need," and an intervention can be *deemed* "spiritual care."

Consider, for example, the urgent need to evacuate one's bowels. Can this be deemed a "spiritual need"? Presumably not; and there is nothing in the literature to imply that anyone thinks it is. Now consider a severe relationship problem and the stress to which it gives rise. Can this be deemed a "spiritual need"? According to the deflationary SIHC literature, it can (Murray et al. 2004). How about a man in hospital worrying about gambling debts? Is that a "spiritual need"? According to McSherry (2006), it is. Further examples of "spiritual need," taken from McSherry's book, include a retired teacher missing the job, a woman suffering personality change as a result of dementia, a man with aphasia not wanting to have a gastronomy tube inserted, a man worried about his family after suffering a myocardial infarction, and a vegan refusing hospital meals. This implies a very broad definition of "spiritual need," in the absence of any argument about why this definition is either necessary or desirable.

In deflationary writing there is little discussion of warrants according to which a vegan refusing meals or a man worrying about his gambling debts can be deemed examples of "spiritual need." What discussion there is consists largely of references to "meaning and purpose," implying that whatever is considered meaningful by a particular person can, *ipso facto*, be deemed "spiritual." As noted earlier, inflationary writers regard this implication as absurd (Hill et al. 2000). For deflationary writers, though, the project is to identify "what is spiritual" and "what is meaningful" as roughly coextensive, leaving the boundary between them – to the extent that they acknowledge one – fuzzy. There is a project of classification here which is never explicitly discussed.

Classification is a core topic in anthropology, especially cognitive anthropology, and there is a growing literature on the construction of classification categories in various contexts – including the workplace – and the historically specific work that goes into creating and sustaining them (Bowker and Star 2000). Classification

proposals are necessarily motivated, and it is always possible to ask who benefits from a particular set of categories. "Each standard and each category valorizes some point of view and silences another… Classifications are powerful technologies… [They] should be recognised as the significant site of political and ethical work that they are" (319). But the classifications we use and the labor that goes into producing them are, for the most part, invisible. Indeed, classification work is often not recognized as such. Debates in which the critical terms are said to be "complex" or "elusive," but in which the most significant theory-laden concepts are nevertheless represented as self-evident, are frequently the marker of a classification project and the ideological investment that accompanies it.

In the deflationary SIHC literature, the primary strategy is one of persuasive definition, a philosophical move first described 70 years ago (Stevenson 1938; Macagno and Walton 2008). According to Stevenson, persuasive definition involves a recalibration of the "descriptive meaning" of a word, while retaining its "emotive meaning." The point would not be expressed in this way now. A distinction would be made between "denotation," what the word refers to, and "connotation," the network of cultural associations linked to the word (Murphy 2010). Recalibrating the denotation does not alter the connotation. In the case of spirituality, persuasive definition involves the recalibration of what the word "spiritual" applies to, its denotation, while retaining the cultural associations of the term, its connotation.

The cultural associations in question can be traced to religion, especially Christianity. "Spiritual" drags in its wake a matrix of inescapably Christian undertones. Wherever "spirituality" goes, it is accompanied by a cloud of religious meanings, as evidenced by the language it adopts: "sacred," "transcendent," "the ultimate," "higher power," and so on. It began as a religious concept (Principe 1983; Sheldrake 1991), and it remains entangled in theological ideas. On this view, those who wish to disseminate spirituality discourse are creating or maintaining a space for religious sensibilities in health care. As many of those who write about spiritual care in the UK are chaplains, ministers, or theologians (Swinton, Kelly, Cobb, Pattison, Mitchell, Gordon, Robinson, Brown, Nolan), this is perhaps not surprising. The theologically inflected language that characterizes their writing is probably inevitable.

Expanding the denotation of "spiritual need" and "spiritual care" makes room for religious sensibilities by attaching the religious connotations of "spirituality" to a great variety of health-related concepts and practices. This is routine in the religious USA (Demerath 2000), where the sacralization of institutions outside the church is well established; but in the more secular UK, it is a relatively new departure.

McCutcheon (2010) has observed, "the only reason scholars find religion everywhere in the world… is because those very scholars approach the world – in fact, make their world – by using this term defined broadly enough, so as always to find sufficient things that they can deem religion – suggesting to me that a theory of *deeming*… is far more required than a theory of religion" (1188–1189). Replace "religion" with "spirituality," and "the world" with "health care," and this statement

aptly summarizes the key point made in this section. The SIHC literature lacks a "theory of deeming" – a theory of what is *deemed* to be "spiritual" in health care, and why. This is not provided by deflationary writers, who deem *a, b, c, d* as examples of "spiritual need," or *w, x, y, z* as examples of "spiritual care," without explaining why the deeming is warranted.

Oscillation: The Semantics of Ambiguity

Although it is possible to differentiate two distinct SIHC literatures, the one inflationary and the other deflationary, there is some degree of overlap brought about by the tendency of UK authors to oscillate between them. The idea that spirituality is concerned with relationships, nature, listening to patients, culture, and mundane rituals plays better in the secular UK than talk of "the sacred," "higher levels of being," "the absolute," and the "interconnectedness of mind, body and spirit." But given the theological background of some UK authors, and the tension created by the fact that it is not immediately obvious why the generic adjective "spiritual" is required to designate these familiar pursuits, a certain ambiguity creeps into the UK literature, one which permits "spirituality" to be blandly reassuring on the one hand but teasingly indicative of something "beyond" on the other.

Many authors oscillate between these two poles. In one paragraph, they talk about the relationships, or the dignity of life. In the next, they emphasize the "allure of eternity," or something mysterious which lies "beyond mortal limits" (Cobb 2001: 27; Stanworth 2004: 108). One moment, the "transcendent" is deflationary, no more than a "reaching out" beyond the self to other people. The next, it is inflationary on a cosmic scale: an indivisible, universal consciousness outside space and time (Chung et al. 2007).

Sometimes, the ambiguity seems almost deliberate. For example, Nolan and Holloway (2014) say that "in the literature on spirituality in healthcare, 'God' is not the God of traditional (Judaeo-Christian) religion, but that to which an individual ascribes ultimate value – whatever that may be" (43). This is the *Looking Glass* approach to meaning: "'When *I* use a word,' Humpty Dumpty said, in a scornful tone, 'it means just what I choose it to mean – neither more nor less'" (Carroll 1871). This Humpty Dumpty semantics creates ambiguity quite unnecessarily. Why use a theologically impregnated term such as "God" to refer to someone's values, even their ultimate values? Who, apart from those with a prior commitment to the Judaeo-Christian tradition, would find it necessary or useful to use "God" in that particular way? Who, apart from the religious, thinks in terms of *ultimate* values? No doubt Nolan and Holloway believe that they are being inclusive, making the term "God" available to nonbelievers. But their way of looking at things is *only* available to Christians (and those affiliated to other religions). If a word belongs, as "God" obviously does, to a particular ideology, then a broadening of its lexical range – undertaken by people *inside* the ideology – will not seem persuasive to anyone *outside* it, even if the broadening is designed to "include" the outsiders.

The idea of "something greater than oneself" involves another ambiguity. "Spiritual development is the process of growing the intrinsic human capacity for self-transcendence in which the self is embedded in something greater than the self" (Roehlkepartain et al. 2006: 6). "Something greater than the self" is an equivocal expression. It could be taken as intimating supernatural dimensions, higher powers, or forces connecting us with the whole of creation. But it could also apply to "a civil community, a youth gang, or a multinational corporation" (Oman 2015: 40). However, as with "transcendence," the ambiguity of "something greater" does have a discursive function. It makes possible the reassurance of "reaching out to others," but at the same time it has intimations of a supernatural *beyond*. The ability to switch from one to the other and back again, *pro re nata*, is not without discursive value.

Perhaps the most familiar example of ambiguity is "meaning and purpose." In this case, the deflationary pole refers to what might be termed "immanent meaning." There are two variations on this theme. The first is emotional salience, as in "Music means a lot to me." The second is an inference marker or causal hypothesis. "More breastfeeding means less ovarian cancer," or "Catching the earlier train means that I will get to the concert on time." Similarly, people have purposes, both short-term and long-term. "My purpose in visiting the Guggenheim is to see Richard Serra's The Matter of Time." In none of this is there any hint of a perspective in which an entire life has meaning or purpose. Immanent meaning is concerned solely with the facets, fragments, and filaments of everyday experience *within* a life. There is no obvious reason, independently of a theological project, to classify these filaments and facets as "spiritual."

In contrast, what can be termed "transcendent meaning" implies a perspective from which the events of someone's life, or even that life considered as a whole, can be fitted into a larger narrative. There is still a causal sequence, but now the inference is teleological: "*X* happened in order to bring about *Y*," or "*X* happened, but *Y* was brought about in order to compensate for the effects of *X*." Implicit in these causal stories is the idea that some *über* agent is at work. What happens serves the agent's purpose, or he/she/it can order things so as to reward, punish, rectify, alleviate, console. This agent might be either a personal god, an impersonal cosmic force, or a fair and balanced universe. Indeed, belief in a just world (Lerner 1980) is arguably intrinsic to transcendent meaning. In this sense, the "search for meaning" is a search for a believable *über* narrative that not only *explains* why some tragic event happened but also *justifies* it, adopting a perspective radically "outside" the life of the individual concerned. From this perspective, each person has a purpose. It is not self-chosen but is ordained by the *über* agent, just as a character in a book has a narrative purpose ordained by the author.

Like other expressions such as "transcendence," "something greater than ourselves," or "ultimate value," "meaning and purpose in life" functions as a bridge between inflationary and deflationary discourse. All of them can be used to designate something familiar, tangible, and ordinary; but each of them also has a reference to the supernatural, "radical otherness" or "the mysterious," which is sometimes made explicit, but which often remains unnamed but

present. For the theological classification project, this is a significant discursive resource.

Summary: The Structure of Discourse

The discourse of spirituality-in-health-care exhibits a number of features that are not always present in the academic literature. One founding principle is that extravagant metaphysical claims can be made or presupposed without having to be defended, without reference to the relevant disciplines, and in the expectation that they will not be challenged. Many of these claims, like the universality principle, are taken as self-evident. At the same time, a series of familiar epistemological tropes is used to secure an inviolate space in which challenges from a naturalistic perspective can be rendered otiose. At this point, SIHC discourse splits into two on mainly demographic grounds: an inflationary version in the USA and a deflationary version in the UK. Two distinctions between "spirituality" and "religion" are adopted: inner/outer and broad/narrow, respectively. One interesting consequence of this demographic split is that the evidence for positive health outcomes as a result of religion/spirituality may apply only to the USA (and other religious countries), given the extent to which American culture is saturated with religion. Authors who adopt the broad/narrow distinction extend the denotation of "spiritual need" and "spiritual care" as much as possible in order that the universality principle will seem more plausible in a secular society. This amounts to a classification project, an exercise in persuasive definition, in which the relevant "deemings" require no defence and in which a semantic bridge is constructed between the inflationary and deflationary poles. As a consequence, a discursive space is created and maintained for religious sensibilities in health care. The classification project is, for that reason, a broadly theological one.

References

Abu-Raiya H, Pargament KI (2012) On the links between religion and health: what has empirical research taught us? In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 333–339

Baldacchino D, Draper P (2001) Spiritual coping strategies: a review of the nursing research literature. J Adv Nurs 34(6):833–841. doi:10.1046/j.1365-2648.2001.01814.x

Bauen A, Berndes G, Junginger M, Londo M, Vuille F (2009) Bioenergy: a sustainable and reliable energy source: a review of status and prospects. Implementing Agreement on Bioenergy. International Energy Agency, Paris

Bellah RN, Madsen R, Sullivan WM, Swidler A, Tipton SM (1985) Habits of the heart. University of California Press, Berkeley

Bowker J (ed) (1997) The Oxford dictionary of world religions. Oxford University Press, Oxford

Bowker GC, Star SL (2000) Sorting things out: classification and its consequences. The MIT Press, Cambridge, MA

Brierley P (2000) The tide is running out: what the English Church Attendance Survey reveals. Christian Research, London

Bruce S (2002) God is dead: secularization in the West. Blackwell, Oxford
Bruce A, Sheilds L, Molzahn A (2011) Language and the (im)possibilities of articulating spirituality. J Holist Nurs 29(1):44–52. doi:10.1177/0898010110381116
Burke LA, Neimeyer RA (2012) Meaning making. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 127–133
Büssing A (2012) Measures. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 323–331
Carrette J, King R (2005) Selling spirituality. Routledge, London
Carroll L (1871) Through the looking glass, and what Alice found there. Macmillan, London
Carruthers P (2006) The architecture of the mind. Oxford University Press, Oxford
Chida Y, Steptoe A, Powell LH (2009) Religiosity/spirituality and mortality. Psychother Psychosom 78(2):81–90. doi:10.1159/000190791
Chung LYF, Wong FKY, Chan MF (2007) Relationship of nurses' spirituality to their understanding and practice of spiritual care. J Adv Nurs 58(2):158–170
Cobb M (2001) The dying soul: spiritual care at the end of life. Open University Press, Buckingham
Cobb M, Puchalski C, Rumbold B (2012) The future of spirituality and healthcare. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 487–492
Cushman F, Young L, Greene JD (2010) Multi-system moral psychology. In: Doris J (ed) The moral psychology handbook. Oxford University Press, Oxford, pp 47–71
Davies M (2015) corpus.byu.edu. Brigham Young University. Available at http://corpus.byu.edu/. Accessed 16 June 2015
Demerath NJ (2000) The varieties of sacred experience: finding the sacred in a secular grove. J Sci Study Relig 39(1):1–11. doi:10.1111/0021-8294.00001
Dyson J, Cobb M, Forman D (1997) The meaning of spirituality: a literature review. J Adv Nurs 26:1183–1188. doi:10.1046/j.1365-2648.1997.00446.x
Eisenberger NI, Cole SW (2012) Social neuroscience and health: neurophysiological mechanisms linking social ties with physical health. Nat Neurosci 15(5):1–6. doi:10.1038/nn.308
Eliade M (ed) (1987) Encyclopaedia of religion. Collier-Macmillan, London
Flanagan K, Jupp PC (eds) (2007) A sociology of spirituality. Ashgate, Aldershot
Ford DF, Muers R (eds) (2005) The modern theologians: an introduction to Christian theology since 1918. Blackwell, Oxford
Gordon T, Kelly E, Mitchell D (2011) Spiritual care for healthcare professionals: reflecting on clinical practice. Radcliffe Publishing, London
Hacking I (1983) Representing and intervening: introductory topics in the philosophy of natural science. Cambridge University Press, Cambridge, UK
Hank K, Schaan B (2008) Cross-national variations in the correlation between frequency of prayer and health among older Europeans. Res Aging 30(1):36–54. doi:10.1177/0164027507307923
Hayward RD, Elliott M (2014) Cross-national analysis of the influence of cultural norms and government restrictions on the relationship between religion and well-being. Rev Relig Res 56(1):23–43. doi:10.1007/s13644-013-0135-0
Heelas P, Woodhead L (2004) The spiritual revolution: why religion is giving way to spirituality. Blackwell, Oxford
Hill PC, Pargament KI (2003) Advances in the conceptualization and measurement of religion and spirituality. Am Psychol 58(1):64–74
Hill PC, Pargament KI, Hood RW, McCullough ME, Swyers JP, Larson DB, Zinnbauer BJ (2000) Conceptualizing religion and spirituality: points of commonality, points of departure. J Theory Soc Behav 30(1):51–77
Hudson R (2012) Personhood. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 105–111

Hufford DJ, Bucklin MA (2006) The spirit of spiritual healing in the United States. In: Koss-Chioino J, Hefner P (eds) Spiritual transformation and healing: anthropological, theological, neuroscientific, and clinical perspectives. AltaMira Press, Lanham, pp 25–42
Inglehart R, Basanez M, Diez-Medrano J, Halman L, Luijkx R (2004) Human beliefs and values: a cross-cultural sourcebook based on the 1999–2002 value surveys. Siglo Veintiuon Editores, Buenos Aires
James E (ed) (1968) Spirituality for today. SCM Press, London
Jantos M (2012) Prayer and meditation. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 359–365
Kelly T (2011) Consensus Gentium: reflections on the "common consent" argument for the existence of God. In: Clark KJ, Vanarragon RJ (eds) Evidence and religious belief. Oxford University Press, Oxford, pp 135–156
Kim J (2005) Physicalism, or something near enough. Princeton University Press, Princeton
King M, Marston L, McManus S, Brugha T, Meltzer H, Bebbington P (2013) Religion, spirituality and mental health results from a national study of English households. Br J Psychiatry 202:68–73. doi:10.1192/bjp.bp.112.112003
Koenig HG (2008) Medicine, religion, and health: where science and spirituality meet. Templeton Foundation Press, West Conshohocken
Koenig HG, King DE, Carson VB (2012) Handbook of religion and health, 2nd edn. Oxford University Press, New York
Lerner MJ (1980) The belief in a just world: a fundamental delusion. Plenum, New York
Lun VM-C, Bond MH (2013) Examining the relation of religion and spirituality to subjective well-being across national cultures. Psychol Relig Spiritual 5(4):304–315. doi:10.1037/a0033641
Macagno F, Walton D (2008) The argumentative structure of persuasive definitions. Ethical Theory Moral Pract 11(5):525–549. doi:10.1007/s10677-008-9119-5
Marler PL, Hadaway CK (2002) "Being religious" or "being spiritual" in America: a zero-sum proposition? J Sci Study Relig 41(2):289–300
Masters KS, Hooker SA (2015) Religion, spirituality, and health. In: Paloutzian RF, Park CL (eds) Handbook of the psychology of religion and spirituality, 2nd edn. Guilford, New York, pp 519–539
Mayo DG (1996) Error and the growth of experimental knowledge. University of Chicago Press, Chicago
McCutcheon RT (2010) Will your cognitive anchor hold in the storms of culture? J Am Acad Relig 78(4):1182–1193. doi:10.1093/jaarel/lfq085
McSherry W (2006) Making sense of spirituality in nursing and health care practice: an interactive approach, 2nd edn. Jessica Kingsley, London
Murphy ML (2010) Lexical meaning. Cambridge University Press, Cambridge, UK
Murray SA, Kendall M, Boyd K, Worth A, Benton TF (2004) Exploring the spiritual needs of people dying of lung cancer or heart failure: a prospective qualitative interview study of patients and their carers. Palliat Med 18:39–45
Nolan S, Holloway M (2014) A-Z of spirituality. Palgrave Macmillan, Basingstoke
Norris P, Inglehart R (2004) Sacred and secular: religion and politics worldwide. Cambridge University Press, Cambridge, UK
Oman D (2015) Defining religion and spirituality. In: Paloutzian RF, Park CL (eds) Handbook of the psychology of religion and spirituality. Guilford, New York, pp 23–47
Oman D, Thoresen CE (2005) Do religion and spirituality influence health? In: Paloutzian RF, Park CL (eds) Handbook of the psychology of religion and spirituality. Guilford, New York, pp 435–459
Oppy G (2012) Philosophy. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 77–82
Paley J (2008) Spirituality and secularisation: nursing and the sociology of religion. J Clin Nurs 17:175–186

Paley J (2015) Why cognitive science of religion cannot rescue "spiritual care". Nurs Philos 16(4): 213–225
Pargament KI (1997) The psychology of religion and coping: theory, research, practice. Guilford, New York
Park CL (2007) Religiousness/spirituality and health: a meaning systems perspective. J Behav Med 30:319–328. doi:10.1007/s10865-007-9111-x
Powell LH, Shahabi L, Thoresen CE (2003) Religion and spirituality: linkages to physical health. Am Psychol 58(1):36–52
Presser S, Chaves M (2007) Is religious service attendance declining? J Sci Study Relig 46:417–423
Principe W (1983) Toward defining spirituality. Stud Relig 12:127–141
Puchalski C, Ferrell B, Virani R, Otis-Green S, Baird P, Bull J, Cochinov H, Handzo G, Nelson-Becker H, Prince-Paul M, Pugliese K, Sulmasy D (2009) Improving the quality of spiritual care as a dimension of palliative care: the report of the consensus conference. J Palliat Med 12 (10):885–904. doi:10.1089/jpm.2009.0142
Puchalski C, Blatt B, Kogan M, Butler A (2014) Spirituality and health: the development of a field. Acad Med 89(1):10–16. doi:10.1097/ACM.0000000000000083
Putnam RD, Campbell DE (2010) American grace: how religion divides and unites us. Simon & Schuster, New York
Roehlkepartain EC, Benson PL, King PE, Wagener LM (2006) Spiritual development in childhood and adolescence: moving to the scientific mainstream. In: Roehlkepartain EC, King PE, Wagener LM, Benson PL (eds) The handbook of spiritual development in childhood and adolescence. Sage, Thousand Oaks, pp 1–15
Roof WC (1999) Spiritual marketplace: baby boomers and the remaking of American religion. Princeton University Press, Princeton
Rumbold B (2012) Models of spiritual care. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 177–183
Scottish Inter Faith Council (2008) Religion and belief matter: an information resource for healthcare staff. Scottish Inter Faith Council, Glasgow
Seybold KS (2012) Cognitive sciences: a perspective on spirituality and religious experience. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 347–352
Sheldrake P (1991) Spirituality and history: questions of interpretation and method. SPCK, London
Sloan RP (2006) Blind faith: the unholy alliance of religion and medicine. St. Martin's Press, New York
Stanworth R (2004) Recognizing spiritual needs in people who are dying. Oxford University Press, Oxford
Stavrova O, Fetchenhauer D, Schlösser T (2013) Why are religious people happy? The effect of the social norm of religiosity across countries. Soc Sci Res 42(1):90–105. doi:10.1016/j.ssresearch.2012.07.002
Stevenson CL (1938) Persuasive definition. Mind 47:331–350. doi:10.1093/mind/XLVII.187.331
Sulmasy DP (2002) A biopsychosocial-spiritual model for the care of patients at end of life. Gerontologist 42:24–33. doi:10.1093/geront/42.suppl_3.24
Swinton J (2012) Healthcare spirituality: a question of knowledge. In: Cobb M, Puchalski C, Rumbold B (eds) Oxford textbook of spirituality in healthcare. Oxford University Press, Oxford, pp 99–104
Thoits P (2011) Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav 52(2):145–161. doi:10.1177/0022146510395592
Walker LF, Pitts RC (1998) Naturalistic concepts of moral maturity. Dev Psychol 34:393–412. doi:10.1037/0012-1649.34.3.403
Zawidzki TW (2013) Mindshaping: a new framework for understanding human social cognition. The MIT Press, Cambridge, MA

Zimmerman K, Smith J, Simonson K, Myers W (2015) Familia relationship outcomes of coming out as an atheist. Secularism Nonreligion. doi:10.5334/snr.aw

Zinnbauer BJ, Pargament KI (2005) Religiousness and spirituality. In: Paloutzian RF, Park CL (eds) Handbook of the psychology of religion and spirituality. Guilford Press, New York, pp 21–42

Zuckerman P (2008) Society without God: what the least religious nations can tell us about contentment. New York University Press, New York

Dying and the End of Life

33

James Stacey Taylor

Contents

Abstract

When it comes to dying and the end of life, two claims are generally accepted: that the difference between a human's being alive and being dead is clear and that death is a harm to the one who dies. But, despite appearances, both of these claims are controversial. There are in fact three competing accounts of what it is for a human to die: the whole-brain approach, the higher-brain approach, and the cardiopulmonary approach. Second, while it is generally accepted that death is a harm to the one who dies, this view is challenged by Epicurus and his followers, who argue that death cannot be a harm for the one who dies. This entry will outline the three primary accounts of what it is for a human to die and then consider the arguments for and against the Epicurean approach to the end of life. It will then consider the implications of the Epicurean position for some of the primary issues in philosophy of medicine that arise at the end of life. Finally, it

J. Stacey Taylor (✉)
Philosophy, Religion, and Classical Studies, The College of New Jersey, Ewing, NJ, USA
e-mail: jtaylor@tcnj.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_67

will consider the claim that the ending of a person's life removes the possibility that events that occur after it could harm her.

Introduction

When it comes to dying and the end of life, two claims are generally accepted: that the difference between a human's being alive and being dead is clear and that death is a harm to the one who dies. But, despite appearances, both of these claims are controversial. There are in fact three competing accounts of what it is for a human to die: the whole-brain approach, the higher-brain approach, and the cardiopulmonary approach. Second, while it is generally accepted that death is a harm to the one who dies, this view is challenged by Epicurus and his followers, who argue that death cannot be a harm for the one who dies. This entry will outline the three primary accounts of what it is for a human to die and then consider the arguments for and against the Epicurean approach to the end of life. It will then consider the implications of the Epicurean position for some of the primary bioethical issues that arise at the end of life.

The Definition of Death

The first issue that must be examined when addressing the question of what it is for a being to die is to determine what constitutes death. (This debate focuses on what constitutes human death, and so this issue will be the focus of this section, also.) Death has traditionally been defined in terms of cardiopulmonary function; a human was considered to be dead once her cardiopulmonary functioning had irreversibly ceased. (Persons whose cardiopulmonary functions could not continue except through artificial means would thus not on this criterion be considered dead, since this criterion does not require that cardiopulmonary function occur absent artificial sustenance.) To determine if a human was dead, one would check to determine if the heart and lungs were functioning, by, for example, checking for a pulse or holding a mirror up to the mouth to see if any moisture gathered to determine if respiration was occurring. Prior to the development of contemporary life-support mechanisms, which enable cardiopulmonary activity to occur without brainstem functioning, a functional cardiopulmonary system was indicative of a functioning brainstem. However, once it became possible for a human's cardiopulmonary functioning to be artificially sustained in the absence of a functioning brainstem, a question arose as to whether the cardiopulmonary functioning of a human's body or its brain function should take precedence in the determination of whether or not she was still alive.

In recent years, a near consensus in the United States has been reached that rather than continuing to accept the traditional cardiopulmonary standard for death, a brain-based standard should be adopted instead. The most widely accepted brain-based

standard for human death is the whole-brain standard. This standard is supported by the organismic view of death: the view that an organism's death occurs when it irreversibly ceases to function as an organism as a whole (Becker 1975). On this view, a being's life critically involves the integrated functioning of the being as a whole. While cardiopulmonary activity typically indicates the presence of life as it typically indicates that the being is functioning as an integrated whole, these phenomena can occur even if the being is not functioning as an integrated whole. The possibility of artificial life support, for example, makes it possible for a brain-dead human to exhibit these signs of life without functioning as an integral unit. A functioning brain, however, is necessary to ensure that a human being functions as an integrated organism, and so the presence of a living whole brain is required for a human to be alive.

The whole-brain approach to death has not gone unchallenged. The first concern is that the functioning of a whole brain is not necessary for integrative organismic functioning. Many functions, such as wound healing, assimilation of nutrients, and detoxification, can operate without the integrative aid of the brain (Shewmon 2001). This observation has been used to support the view that the brain merely enhances integrative functioning, rather than being necessary for it. The second concern arises from observation of patients suffering from locked-in syndrome. The brains of such patients appear to have no more integrative effects on their bodies than those of brain-dead individuals, and yet locked-in patients are undoubtedly alive. These two concerns have not, however, led to the abandonment of the whole-brain criterion for death, but, instead, to a move away from justifying it on the basis of the brain's role in integrating the organismic functioning of a being. In place of a focus on integrative functioning, this new justification focuses on the brain's necessity for a being's function qua the type of being that it is. Thus, on this account, a human dies when she is no longer able to receive and respond to stimuli from the external world or to act on the world to obtain what she needs from it (where this is an exclusive disjunction). Humans with locked-in syndrome, while unable to satisfy the second condition for being alive, satisfy the first, and so on this approach to justify the whole-brain definition of death would be alive (Bartlett and Youngner 1988).

It was noted that the consensus that a living whole brain is required for a human to be alive was reached in the United States. In the United Kingdom, however, a slightly different account of death has been reached: one that focuses not on the whole brain, but on the functioning of the brainstem – a criterion for death that is lower than in the United States and that requires fewer clinical tests to ascertain.

In contrast to the approach of the United Kingdom, some academics have argued that human death should be determined by the irreversible loss of the capacity for consciousness. On such a "higher-brain" approach to death, persons in persistent vegetative states would be considered dead, even though they posses sufficient brain activity to meet the criteria for life offered by both the brainstem and the whole-brain approaches to death.

While the brain-based approaches to death enjoy considerable support, the controversy surrounding the definition of death had led some medical ethicists to argue that patients should be given the option of deciding which criterion for death should be applied in their individual case (Veatch 2004). Such an approach could fit best with the recognition that cultural factors might play in a role in determining death; this approach would also fit with the current dominance of the principle of respect for autonomy in medical ethics.

Epicurus and the End of Life

The above outline focused on the debate between the proponents of the three main competing accounts of the conditions that must be met for a human to be considered to be dead. This debate is separate from the debate over whether or not death is a harm to the human who dies. Typically, it is believed that death – the end of life – is a harm to the human who dies. But this is not a universally shared view. Most famously, in his "Letter to Menoeceus," Epicurus held that:

> Accustom yourself to believing that death is nothing to us, for good and evil imply the capacity for sensation, and death is the privation of all sentience; therefore a correct understanding that death is nothing to us makes the mortality of life enjoyable, not by adding to life a limitless time, but by taking away the yearning after immortality. For life has no terrors for him who has thoroughly understood that there are no terrors for him in ceasing to live. Foolish, therefore, is the man who says that he fears death, not because it will pain when it comes, but because it pains in the prospect. Whatever causes no annoyance when it is present, causes only a groundless pain in the expectation. Death, therefore, the most awful of evils, is nothing to us, seeing that, when we are, death is not come, and, when death is come, we are not. It is nothing, then, either to the living or to the dead, for with the living it is not and the dead exist no longer. (Epicurus 2014)

The most salient feature of death, for Epicurus, is that it is the "privation of all sentience." Since this feature would be shared by the deaths of humans whether these are accounted to be deaths by cardiopulmonary criteria or by brain-based criteria, determining the precise definition of death is orthogonal to Epicurus' argument. For Epicurus, a person's death should be "nothing to her"; she should neither dread it nor welcome it. As Epicurus argued, a person who is alive has not yet been harmed by her death, for it has not yet occurred. (This is compatible with the claim that a person might be harmed by thoughts of her own death; in such a case, though, it would not be her death that harms her but her thoughts that have it as their intentional object.) And a person who is dead would not exist and so would not experience any sensation. Thus, for Epicurus, since a person is either alive or dead, and since when she is alive she has not been harmed by her own death, and when she is dead she cannot be harmed by her own death, a person can never be harmed by her own death. Hence, for Epicurus, a person's own death should have no value for her, either positive or negative. Rather than having meaning for her, then, it should be "nothing" to her.

The Epicurean Position and Bioethics

The Epicurean view that a person's death should be "nothing" to her has significant implications for bioethical issues that are associated with the end of life, such as abortion, euthanasia, suicide, and questions that surround the allocation of scarce medical resources. At first sight, it might appear that the Epicurean position that the end of life is nothing to the one who dies would render the problem of abortion a simple one. After all, if the death of a fetus is not a harm to it, then it seems that abortion cannot be a wrong to the fetus. Yet if it appears that the Epicurean position renders the problem of abortion a simple one, it also appears that it significantly complicates questions concerning the rationality of both voluntary active euthanasia and suicide. If the ending of a person's life is genuinely nothing to her, then it seems that she would have no reason to choose voluntary active euthanasia or suicide. After all, persons only choose to pursue courses of action that they believe to be beneficial for them in some way – and if the ending of a person's life is "nothing" to her, it cannot be beneficial to her. And, just as a person would appear on the Epicurean view to have no reason to choose voluntary active euthanasia or suicide if the ending of her life is nothing to her, then it would also seem that a person would have no reason to avoid her death, if it is genuinely nothing to her. This would appear to have important implications for the allocation of scarce medical resources. This is because persons do (on the Epicurean view) have a prima facie reason to avoid pain and suffering. Thus, if the Epicurean position that the end of a person's life is nothing to her is correct, and if it is true that persons thus have no prudential reason to avoid their own deaths, then it seems that given the fact that persons do have reason to avoid pain and suffering, medical resources should be distributed to alleviate pain and suffering and not necessarily for the prolongation of life.

These Epicurean positions strike many persons as being counterintuitive. Yet despite appearances the adherent of the Epicurean view of the value and meaning of death is not committed to them. While the Epicurean is committed to the view that ending the life of a fetus does not harm it, she is not thereby committed to the view that abortion is therefore not morally wrong. It is possible that an act could wrong an individual (such as the fetus) even if it did not harm her. An Epicurean is also not committed to holding that a person's desire for euthanasia or suicide is conceptually puzzling. First, an Epicurean could recognize that a person could be ignorant of the Epicurean view of the value of the ending of her life and hence (falsely, for the Epicurean) believe that it would benefit her to die. In such a case, a person might desire death for its own sake. Alternatively, an Epicurean could recognize that even an Epicurean could desire the end of her own life. She would not, however, desire it for its own sake – after all, it would be "nothing to her" and so would not have a positive valence – but she would desire it insofar as it would bring about the end of her pain. An Epicurean, then, could desire euthanasia or suicide, even if she could not desire the ending of her life for its own sake. Similarly, an Epicurean could accept that persons have reason to avoid the ending of their lives, even if their deaths would not harm them. It would not be irrational for a person to avoid the ending of her own life if her continued existence would be a necessary condition for her to

pursue projects that she valued. Such a person would not be avoiding death per se, but, instead, would merely be securing the conditions necessary (i.e., continued existence) for her to pursue that which she valued. Since this is so, an Epicurean could adhere to the view that the ending of a person's life is nothing to her and yet reject the view that scarce medical resources should not be used to save lives, but, instead, only used to alleviate the suffering of the sick and injured.

Objections to the Epicurean Position

Since the Epicurean is not committed to any of the counterintuitive bioethical implications that her position might at first sight appear to entail, the Epicurean view of the value of the ending of life cannot be rejected on the basis of its counterintuitive implications. It might, however, be rejected for theoretical reasons.

One of the common objections to the Epicurean view of the value of death is that it fails to recognize that a person might be harmed by the ending of her life through being deprived of the goods of life that she might otherwise have enjoyed. This objection has been famously pressed by Thomas Nagel (1970). Nagel offers an example of a person who has undergone a severe trauma that has reduced him to the status of a contented infant. Even though this person does not experience any adverse effects from his trauma – he is, after all, similar to a contented infant – Nagel holds that we would typically consider this event to be a great harm for him. We would do so, Nagel claims, since we would consider this person to have been deprived of the goods of life that he would have received had he not undergone this injury. Similarly, argues Nagel, we should consider the ending of a person's life to be a harm to her since it would similarly deprive her of the goods of life that she would have otherwise enjoyed.

Yet although this deprivation-based objection to the Epicurean view of the value of the ending of a person's life is initially plausible, the Epicurean has two responses to it. First, she might argue that there is an important disanalogy between the man when he is reduced to the status of a contented infant and a person whose life ends. In the former case, there is a clear subject to be the bearer of the harm – namely, the man who has been reduced to the status of a contented infant. In the latter case, however, there is no subject for the putative harm of the ending of the life to befall. Second, the Epicurean might contest the claim that the man in Nagel's case was harmed at all. She might argue that mental differences that exist between the man as he was prior to the trauma and the man who existed after the trauma are so significant that the man who was the victim of the trauma ceased to exist and was replaced by another being who happened to have the same body. While this case would be directly analogous to the case of death, the Epicurean would note that to hold that the man was clearly harmed by the ending of his life would simply beg the question against the Epicurean position (Warren 2004).

More persuasive challenges to the Epicurean view of death focus on its claim that a being is either dead or not dead. As the above outline of the debate over which criteria must be met for a being to be dead shows, it is not clear what it is for a being's

life to end. One might thus challenge the Epicurean view that a person is either dead or not dead on the grounds that personal existence does not admit of such binary ordering, but instead is a matter of degree. That is, that during the dying process, a person is neither completely alive, nor completely dead, but somewhere in between. Similarly, if one believes that time is continuous, then one might challenge the claim that a person ceases to exist at a particular point in time, as the Epicurean argument appears committed to. If one believes that time is continuous then one will believe that between any two points in time there will be an infinite number of other points in time. Since this is so, for any two points in time, A and B, where A is the last moment at which a person existed and B is the first moment of her nonexistence, there will be an infinite number of other points in time where the existence or otherwise of the person in question is indeterminate.

Harm After the End of Life

One might also object to the Epicurean position on the grounds that it is based on a mistaken account of well-being. For the Epicurean, a person's well-being is a function of her mental states: A person's well-being is enhanced by positive mental states and reduced by negative ones. Since dead humans have no mental states, it follows that death can be neither a harm nor a benefit to humans, since they do not have any mental states after their lives have ended. But this hedonic account of well-being is controversial, and many philosophers reject it in favor of an interest-based account of well-being, on which a person's well-being is a function of the thwarting or fulfillment of her interests. (Which of a person's interests count, and for how much, is a matter of considerable debate.) On this account, a person could be harmed or benefitted by an event that occurs after the end of her life if it either fulfills or thwarts one or more of her interests. On the interest-based account of well-being, then a person could still be subject to harms and benefits even after the end of her life.

This criticism of the Epicurean position could be understood as construing a person's death as a type of posthumous harm: If one accepts the Epicurean view that a person is either dead or not dead and holds that a person's interests in remaining alive are thwarted once she is dead, then the event that harms a person through thwarting her interest in staying alive (her death) will occur at the first moment of her nonexistence. (This is assuming that there is such a moment – an assumption that a proponent of the continuous view of time might take issue with.) The person would thus be harmed by an event that occurs after the end of her life: She would be subject to posthumous harm.

If this criticism of the Epicurean position is sound, it opens up a new avenue of inquiry in philosophy of medicine concerning the normative significance of the end of a person's life: Does the end of a person's life mean that her interests should no longer be considered as being morally relevant or not? The answer to this question is relevant to a wide range of issues in philosophy of medicine. These include questions concerning the ethics of posthumous organ procurement, questions concerning the

ethics of research on the dead, questions concerning the ethics of postmortem pregnancy, and questions concerning issues of posthumous medical confidentiality. If it is possible to harm persons after their lives have ended, then the interests of persons who are now dead should count in one's moral assessment of issues such as those just outlined, just as much as the interests of persons whose lives are still continuing. For example, if a person who had an interest in her body remaining intact after the end of her life would be harmed by the postmortem removal of her organs for transplant, this should count when one determines whether her organs should be removed or not. More generally, this question raises issues concerning the reason-giving force of a person's past interests while she is currently alive. Does the fact that in her past a person had an interest that a certain outcome transpire give her any reason today (i.e., when she lacks the interest in question) to bring about this outcome?

The possibility that persons might be harmed by events that occur after the end of their lives has been defended independently by George Pitcher and Joel Feinberg (Feinberg 1984; Pitcher 1984). On the Feinberg-Pitcher view, a person is harmed when her interests are thwarted. Since some of a person's interests extend beyond the scope of her life (such as, e.g., her interest in keeping her medical records confidential), thwarting these interests after her life has ended will harm her. Since corpses cannot be harmed, the subject to the posthumous harm will be the antemortem person – the person as she was when she was alive. This might appear to involve backward causation, such that an event that occurred after a person's death causes her to be harmed during her life. But this need not be the case, for the relationship between the putative harmful event and the harm incurred as a result might be a conceptual relationship, rather than a causal one. For example, if a woman's daughter has a child after the woman is dead, the birth of this child will make it true that she is a grandmother, even though it would not cause this change in her status.

One might object to the possibility that a person could be harmed by an event that occurs after the end of her life by objecting to the interest-based account of well-being that it is based on: This would be the first move of an Epicurean. But one could also object to it on the grounds that while some properties (such as being a grandmother) could be attributed to persons on the basis of events that occur after the ends of their lives, these are limited to properties that can be uncontroversially attributed on the basis of the occurrence of certain events. (For example, when a person's child has a child, that person is uncontroversially a grandparent.) But, the proponent of this objection continues, "harm" is not a property whose ascription is uncontroversial in this way – there are, after all, at least two different accounts of what criteria must be met for a person to be harmed. There is thus no reason to believe that the property of harm can be ascribed to persons on the basis of events that occur after their lives have ended. Hence, this objection concludes, there is no reason to believe that it is possible to harm someone after the end of her life, even if one accepts the interest-based account of well-being.

Conclusion

There are three primary competing accounts of what conditions must be met for the process of dying to have ended and a person to be considered dead: the whole-brain approach, the higher-brain approach, and the cardiopulmonary approach. There are also debates over the questions of whether the ending of a person's life is a harm to her or not and whether a person can be harmed (or benefitted) by events that occur after her life has ended. All of these theoretical questions have significant implications for philosophy of medicine ranging from questions concerning ethical issues associated with organ procurement to questions concerning posthumous medical confidentiality.

Definition of Key Terms

Interest-based account of harm	A person is harmed if her interests are set back, where her interests are distinct from her preferences and desires.
Hedonic account of harm	A person is harmed if she experiences adverse mental states.
Cardiopulmonary function account of death	A human is dead once her cardiopulmonary functioning has irreversibly ceased.
Organismic account of death	An organism's death occurs when it irreversibly ceases to function as an organism as a whole.
Epicurean view of death	A person cannot be harmed or benefitted by her own death.

Summary

- There are important distinctions between different accounts of death, e.g., the cardiopulmonary account and the organismic account:
- The Epicurean view controversially holds that death is not a harm to the person who dies.
- The Epicurean view of death has significant implications for the philosophy of medicine.
- The Epicurean view of death faces several objections, arising from, for example, particular metaphysical views concerning the nature of time.
- There is considerable debate over whether or not posthumous harm is possible.

References

Bartlett E, Youngner S (1988) Human death and the destruction of the neocortex. In: Zaner R (ed) Death: beyond whole-brain criteria. Kluwer, Dordrecht

Becker L (1975) Human being: the boundaries of the concept. Philos Public Aff 4:334–359

Epicurus, "Letter to Menoeceus," at http://www.epicurus.net/en/menoeceus.html. Accessed 31 Oct 2014

Feinberg J (1984) Harm to others. Oxford University Press, Oxford

Nagel T (1970) Death. Noûs 4:73–80

Pitcher G (1984) The misfortunes of the dead. Am Philos Q 21:183–188

Shewmon DA (2001) The brain and somatic integration: insights into the standard biological rationale for equating "brain death" with death. J Med Philos 26:457–478

Veatch R (2004) Abandon the dead donor rule or change the definition of death? Kennedy Inst Ethics J 14:261–276

Warren J (2004) Facing death: epicurus and his critics. Oxford University Press, Oxford

"Lives at Risk" Study: Philosophical and Ethical Implications of Using Narrative Inquiry in Health Services Research

34

Ashrafunnesa Khanom, Sarah Wright, Marcus Doel, Melanie Storey, Clare Clement, and Frances Rapport

Contents

Abstract

Narratives ascribe meaning to individual experience and life events through a process of storytelling. Storytelling provides a context for understanding illness and health by mirroring life back to the self while at the same time disseminating personal inner thoughts of the storyteller out to the wider world.

A. Khanom (✉) • S. Wright • M. Storey • C. Clement
College of Medicine, Institute of Life Sciences 2, Swansea University, Swansea, Wales, UK
e-mail: a.khanom@swansea.ac.uk; S.Wright@Swansea.ac.uk; S.M.Storey@Swansea.ac.uk; C.Clement@Swansea.ac.uk

M. Doel
College of Science, Department of Geography, Wallace Building, Swansea University, Swansea, Wales, UK
e-mail: M.A.Doel@Swansea.ac.uk

F. Rapport
Centre for Healthcare Resilience & Implementation Science, Macquarie University, NSW, Sydney, Australia
e-mail: frances.rapport@mq.edu.au

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_80

This chapter will examine the contribution, meaning, and value of narrative inquiry within a health services research context. In so doing, it considers one specific case study, the "Lives at Risk" study, discussing some of the philosophical and ethical implications of narrative inquiry as it relates to this case.

Introduction

This chapter examines the contribution, meaning, and value of narrative inquiry within a health services research context. In so doing, it considers one specific case study, the "Lives at Risk" study (described in detail later in the chapter), discussing some of the philosophical and ethical implications of narrative inquiry as it relates to the case.

What Is Narrative Inquiry?

People have been telling stories (also known here as narratives) about how they live and what that means to them as a natural human response to life. A narrative inquiry approach within a health services research context has emerged relatively recently, from within the broader field of qualitative inquiry in social and human sciences. In other fields, such as education and sociology, narrative inquiry is much more familiar. Narrative inquiry, as a research method, has been said to contribute to knowledge development as a means of information transfer:

> . . .lived and told stories and the talk about the stories are one of the ways that we fill our world with meaning. (Clandinin and Rosiek 2007, p. 35)

Narrative inquiry is considered to be a form of storytelling (Frank 1995, p. 19). In this sense, the process of *telling* offers insights for the researcher into the social phenomena being considered, while adding to our understanding through the storyteller's (or research participant's) approach to presentation. Narrative inquiry, most typically, uses qualitative data collected from personal stories, autobiographies, journals, letters, interviews, and photographs to develop greater knowledge and understanding of the meanings people attach to their lives (Connelly and Clandinin 2000) and their sense of identity and selfhood (Ricoeur 1992). The uncovering of subjective truths by engaging with a variety of texts may produce a richer and more holistic understanding of the lived experience (Ricoeur 1991; Wiggs 2011) as opposed to considering individuals as "social variables" that can in some way be measured (Maynes et al. 2012, p. 16), apparent within impersonal statistical analyses.

Narrative inquiry data can be collected over a brief or extended period of time and can contribute to the development of broader understanding of what it means to be human within a social setting, as well as narrower understanding of an individual's personal experience (Connelly and Clandinin 2000; Nettelbeck 2008).

Understanding the experience of the research participant through narrative presentation allows the health services researcher to gain an "insider view" (Wang and Geale 2015, p. 2). The researcher aims to develop a deeper understanding of health issues as experienced by participants (Green 2013) and through the participants' own, enhanced self-awareness (Skultans 2000), while at the same time acknowledging participants' own subjectivity during the process of data collection, interpretation, and presentation of stories (Wang and Geale 2015).

Narrative inquiry, to reiterate, is principally an approach that seeks to identify, represent, and share knowledge and the experiences of others as they may apply to a range of different social phenomena. In effect, the process of telling a story enables knowledge to be transferred from the storyteller to the listener (Frank 1995, p. 19) while allowing meaning to be bestowed in the form of a collective narrative of a particular life experience (Bruner 2003). Knowledge transfer relies heavily on the storyteller's private memory and the moment in time in which the story is told, as much as it does to and the meanings assigned to it. Narratives are often interpreted and reinterpreted by the teller and the audience in light of their own perceived values, identity, and positions of power, where they can exercise choice on aspects of narratives with which they wish to identify and those to which they feel less affinity. Research participants are often clear about which aspects of their narratives are likely to resonate with a particular type of audience and so adjust them accordingly. Cathy Kohler Riessman (2008, p. 3) summarizes this by saying:

> . . ..a speaker connects events in a sequence that is consequential for later action and for the meanings that the speaker wants listeners to take-away (sic) from their story. Events perceived by the speaker as important are selected, organised, connected and evaluated as meaningful for a particular audience.

Kohler Riessman also helpfully outlines the many uses of narratives, from constructing individual and group identities to initiating social action and giving voice to those who are least heard. However, Frank adds a note of caution, stating that the process of telling stories, although beneficial, could affect the research participant in some unforeseen and "dangerous" way (Frank 2010, p. 71), a point which the researcher should be aware of at all times when engaging with participants. Thus, narratives are not "static entities," and they should be understood as part of an ongoing dialogical relationship between storyteller, listener, and audience (Riessman 2008; Frank 2010).

Analyzing Narratives

Storied lives are far from autonomous constructions or reconstructions of standalone experiences. Rather, they should be understood in terms of relational and situational experiences, referencing existing social, political, and cultural events and encounters that forge an identity for the self. Individual personal narratives are

often located within these larger meta-narratives of social experience, in order to make meaningful sense (Atkinson and Rubinelli 2012).

Connelly and Clandinin (2006, pp. 477–487) have developed a conceptual framework consisting of narrative "temporality," "sociality," and "place" for supporting the analysis and reporting of others' narratives in a more precise way. *Temporality* can refer to the timeframe being described within a story and the way that time impacts on the telling of the story. Temporality also refers to the changing nature of stories as told, in order to make the most appropriate meaning of a person's life, for as Carr (1986, p. 76) indicates, "we are composing and constantly revising our autobiographies as we go along." In so doing, the future emerges from the understanding of various iterations of a person's narrative. Such temporality refers not only to the storyteller's life but also to the places, objects, and events in which encounters and experiences are contextualized.

Sociality refers to the way personal and social experiences interplay in the shaping and delivery of stories, while *place* or the context of stories takes us beyond the personal aspects of experience to the context within which the story took place and the social conditions most relevant to that context. Marmon Silko (1996) emphasizes the importance of recognizing that physical boundaries are inextricably linked to the experiences that take place within specified places and that there is a need to recognize physical boundaries surrounding the context within which events in stories unfold in order to fully understand the meaning of the narrative. Connelly and Clandinin (2006) describe all three concepts as important contributing factors that shape meaning-making and personal association.

Landman (2012, p. 30) sets out four levels of analysis in support of Connelly and Clandinin's (2006) conceptual framework. These are:

1. The linear level (relating to the sequential nature of storytelling and the basic structure of the narrative)
2. The relational level (what the story reveals about relationships between storyteller and listener)
3. The emotional level (conveying the storyteller and listener's feelings and subjective understandings of an event)
4. The analytical level (where the researcher reflects on the collected material)

Landman suggests that by underpinning the conceptual framework with these four levels of understanding, it is possible to guide researchers in making sense of how stories are told. Bruner (2003, p. 114) also illustrates this point when he says:

> . . .the ways of telling and the ways of conceptualising that go with them become so habitual that they finally become recipes for structuring experience itself, for laying down routes into memory, for not only guiding the life narrative up to the present but directing it into the future.

Bruner (2003) identifies that in analytical terms, obtaining knowledge through narratives that are told, heard, and recreated for research purposes is an inherently

interpretative process. Not only do the individuals providing their narratives "filter out" information that they consider superfluous, but the researcher then filters out information that is not clearly aligned with the research purpose, research question, or overall endeavor for the purpose of clarifying the complex experiences, perceptions, and events they are hoping to interpret.

Unraveling Epistemological Questions that Arise from Narrative Inquiry

We turn next to the necessity of unraveling wider epistemological questions that can be bound up with the researcher's appreciation of others' narratives, in order to pass on knowledge about the personal experiences of others from a health services research perspective.

By considering philosophical and epistemological questions linked to narrative presentation, we can consider not only the story as told but also the nature and scope of the knowledge acquired, resonating with the need to appreciate that there are multiple ways of not only studying but also knowing the social worlds of others.

The research participant and the researcher, who are actively engaged in a narrative inquiry, directly affect the scope of knowledge acquired and, as a result, our understanding of social phenomena and the construction at a macro level of our social worlds. Bruner, cited in Charon and Montello (Charon and Montello 2002, p. 4), speaks of the epistemology of narratives as follows:

> Telling stories is an astonishing thing. We are a species whose main purpose is to tell each other about the expected and the surprises that upset the expected, and we do that through the stories we tell.

Bruner claims that narrative inquiry is an instrument for collecting knowledge about experience, but that the collection of knowledge is not enough in and of itself. Rather, knowledge that is gleaned needs to be situated in a broader landscape to facilitate a deeper understanding of ontology, exploring truths embedded in what we come to know, as well as the way in which knowledge has been constructed. The cornerstones of narrative inquiry set out by Connelly and Clandinin (2006) may offer a framework for specifying certain dimensions of *new* knowledge to which Bruner refers, in as much as they support the passing on of knowledge and subsequently its reconstruction by others. Thus, re-situating stories and experiences within the social, political, and cultural fabric of society allows human endeavor to flourish.

The first cornerstone of Connelly and Clandinin's (2006) framework, "temporality," relates to the notion of continuity and change, envisaged within events, conversations, and individual thought, that then becomes the subject of the inquiry. As these are all in a state of flux, it is the job of the researcher to collect insights into the research participant's associated past, present, and future expectations, to develop a more comprehensive view of how they make sense of their world.

The second cornerstone, "sociality," focuses on the notion that in any given experience, event, or situation, people are always interacting with one another as well as simultaneously managing their own views and experiences of what is happening. Individuals' feelings, hopes, opinions, and moralities, suggests Connelly and Clandinin (2006), may influence how the researcher will interact with the participant and, as such, are as implicit in shaping understanding and the retelling of stories. Moreover, the role of the researcher cannot be removed from how the narrative develops in form and content, because the relationship becomes integral to the disclosure of the narrative and in turn informs the context of disclosure. Simultaneously, social conditions, as referred to by Connelly and Clandinin (2006), can influence both the storyteller and the listener and can shape what is said and what is heard.

Clandinin and Rosiek (2007) have considered the wider philosophical issues surrounding narrative inquiry by referring to the way the researcher can be affected by conflicting narratives in their own academic and personal worlds and how this can then impact on the research process. While the researcher can offer information to the narrative audience about their own views, experiences, and personal position, this nevertheless creates an additional challenge for how stories are managed, interpreted, and represented.

Connelly and Clandinin's (2006) final cornerstone, "place," refers to the abstract, imagined, and recalled places within which events take place. In epistemological terms, while the events being narrated are crucial to meaning-making, knowing the worlds of others, according to the positive or negative associations that the storyteller bestows on them, is also essential. Thus, meaning must be extracted to imply not only context-building but also symbolic association, and the researcher must consider the best way to contextualize the impact of place within the interpretation of the narrative and the way it is reported.

Not all researchers choose to, or indeed are able to, grapple with the more complex philosophical challenges of understanding the epistemological and ontological dimensions of narrative inquiry. However, Connelly and Clandinin's (2006) three cornerstones offer a useful framework for guiding researchers who wish to do so.

Pinnegar and Daynes (2007, p. 7) recognize that researchers must undertake a difficult personal transition if they wish to conduct philosophically appropriate and ethically driven narrative inquiries that take into account "the blurred genres of knowing." They refer to the need to explore "narrative turns" (2007, p. 7) in the timeline of a researcher's own conduct, suggesting that the "narrative turn" refers to:

> . . .a change of direction from one way of thinking or being toward another . . . in the process of designing, studying, and engaging in inquiries. (Pinnegar and Daynes 2007, p. 7)

In this respect, it is the researcher's own motivation that influences the stories being told and the stories being heard. As a result, the researcher needs to consider her/his own responsibility and ethical stance, beyond the mere collection of the

research participant's story and the writing and dissemination of outputs. The researcher is embedded in the research process, in the shaping of telling, in the recollection and retelling of others' lives, and in the context of health and illness narratives.

With this in mind, we turn to Frank's (1995) categorization of patients' illness narratives as:

1. *Restitution narratives*, in which the storyline or "plot" involves returning in the narrative presentation to a previous state of health. For researchers working with those living with chronic illness, listening to stories of a long recovery can be both inspiring and intimidating.
2. *Chaos narratives*, in which all life events are in a state of flux and subject to change. For researchers, others' illness narratives can reveal vulnerability, futility, and impotence (Frank 1995) and can be difficult to hear.
3. *Quest narratives*, in which illness is seen as a journey (both spiritual and secular). Quest narratives may serve to instill within the researcher an appreciation of a narrative ethic of illness or, as Ricoeur defines it, "living a good life" (Ricoeur 1992, p. 172) for all concerned.

The categorization of stories, according to Frank's typology above, the process of analyzing stories according to these categories, and the impression that these stories make on the researcher require the extrapolation of meaning attributed to the researcher's own perceptions of temporality, sociality, and place. Narrative inquiry, in other words, is affected by:

> the differences in people's experienced meaning, the stories they tell about this meaning and the connections between storied texts and the interpretation of those texts. (Polkinghorne 2007, p. 471)

Ethical Implications in Conducting Narrative Inquiry

In the pursuit of gaining new understanding in a health services research context, as the section above begins to explain, interesting ethical dilemmas can present themselves to the researcher. These relate to both the nature of understanding and the researcher's responsibility in the creation of new knowledge. Firstly, the notion that a narrative is socially constructed by both the research participant and the researcher, and then reconstructed by the researcher alone, indicates the existence of an unequal relationship and consequently an unequal distribution of power. Accordingly, it is important to acknowledge that while stories can become "research data," the dynamics of the situation and context of the collected narratives can change.

Secondly, because of the dialogical nature of storytelling (Frank 1995, p. 19), research participants may feel obliged to "say more" than they would ordinarily, in

order to uphold a conversation and build a trusting relationship with the researcher. The researcher must be aware of this and manage the situation carefully.

Thirdly, the researcher may respond to the dialogical nature of the relationship with the research participant, by becoming emotionally engaged not only with what they hear but also with the individual involved. As a result they may overlook or overstep the boundaries between the researcher and the researched, as laid out by the researcher's code of ethical practice. Some participants may be overcome by their emotions during narrative presentations, and a plan should be in place to manage such a situation should it arise (Wang and Geale 2015).

Fourthly, the researcher has a responsibility to the research participant to conduct research according to institutional codes of ethical practice and research principles (e.g., see British Sociological Association Code of Ethics, Academic Governance Protocols (British Sociological Association 2002)) and thus uphold an ethical relationship in studies involving narrative inquiry. This is pivotal to ensuring data are collected, analyzed, and disseminated appropriately. Finally, while it is well recognized that informed consent processes take place at the outset of a research study (Estroff 1995), in a narrative inquiry study, it is important for the researcher to appreciate the extent to which they are crafting the finished product and to consider whether there is a need to return to the research participants with their work before disseminating it more widely.

While attesting to the value of narrative inquiry, we also need to be aware of opposing arguments that question the value of narratives. Strawson (2004), for example, argues that the narrative model of meaning-making is not commonly experienced by individuals:

> there are deeply non-Narrative people and there are good ways to live that are deeply non-Narrative. (Strawson 2004, p. 429)

Strawson (2004) suggests that the explicit assumption that humans are happy to narrate by nature undermines those, and may even cause undue harm to those, who do not consider their experience of life in narrative form. Atkinson and Rubinelli (2012) suggest there are ways of overcoming such concerns, by offering alternative, "nonnarrative" approaches to meaning-making (such as creating art). These allow individuals to explore their personal and social experiences and give meaning to such experiences.

Narrative Inquiry in Action: The "Lives at Risk" Study

Having considered some of the ethical and philosophical issues associated with a narrative inquiry approach in a health services research context, the chapter now turns to narrative inquiry in action. The "Lives at Risk" study provides a case study to illustrate how the authors used the method reflectively, while taking into account the research participant's story, the context for storytelling, and the environment in

which patients' narratives were told. The case highlights both the specificities of the research study (its aims and research questions) and links the outcomes of the study through the concepts of "temporality," "sociality," and "place," the three cornerstones of narrative inquiry already outlined in this chapter (Connelly and Clandinin 2006, pp. 477–487).

The "Lives at Risk" study which, among its defined study objectives, aimed to improve the quality of service delivery for patients with breast cancer involved two cohorts of women: (1) women who had undergone treatment for one or more episodes of breast cancer and were being managed by an oncology unit and (2) women who were determined by clinicians to be at a genetic risk of breast cancer, but had never undergone treatment for breast cancer.

This study used bio-photographic qualitative data collection methods. These are methods of a visual and textual nature that can be considered as discrete elements, corroborative elements, or combined elements in a study. These methods used discretely or in combination can provide rich data that aims to add to our understanding of patient's personal experiences, and in this case textual and visual narrative materials were considered in terms of what it means for women to live with the "risk" of developing breast cancer or further cancer episodes. Study participants were asked to create "Books of Experience" as Stage 1 of the study. These were scrapbook-style, large, blank, spiral-bound books which were provided to enable women to present their narrative experiences in visual and textual formats. Participants were asked to present key aspects of their experience, such as being seen by an oncology consultant for risk assessment and risk presentation or being seen by a genetics specialist to discuss risk in relation to familial links with breast cancer and to relate these aspects of experience to the shaping of their own beliefs of what it means to be living a life "at risk." The research team provided a brief outline of six predetermined themes that the research sought to explore in order to support women's completion of the "Books": (1) risk assessment; (2) tests, treatments, and drugs; (3) impact on life; (4) support and care; (5) time periods; and (6) future expectations.

On completion, the "Books of Experience" were collected and considered by the research study team in accordance with the predetermined thematic areas of interest. Bio-photographic elicitation interviews were then conducted, as Stage 2 of the study, with each woman, who used the interview time to feel at liberty to interpret the Book's narrative for the researcher. Bio-photographic elicitation interviews also provided the opportunity for participants to convey the symbolism of described events, experiences, and images in more detail. The interview data encouraged the research team to explore how women responded to their own narrative presentations, once finalized.

The method of collecting bio-photographic data from the "Books of Experience" alongside the bio-photographic elicitation interviews provided women with the opportunity to clarify the disordered nature of real life for themselves and for the researcher and to highlight those areas that specifically related to their health and illness. The study findings illustrated that, irrespective of the independent nature of narrative creation, the researcher's role in the facilitation and collection of

narrative material enabled them to extract and forefront specific aspects of information. Furthermore, the research team's interpretation and involvement in the co-creation of these narratives was influenced by their impressions of the women's stories and the women's approach to narration, which would, in turn, influence how they engaged the reader or audience with the study outcomes.

Before reporting on the interpretative process and the findings generated from the "Lives at Risk" study, it is important to expand on the research study team's reflections of their influence in this narrative inquiry. It is also important to explore the associated impact of this on the representation of study participants' narratives. In this particular study, taking the notions of "temporality," "sociality," and "place" into account, the study team recognized the importance of the chosen context (a medical hospital environment) and the plethora of relationships within and beyond that context that shaped participants' experiences.

The research team identified that several aspects of the study researcher's own identity were significant in the collection and interpretation process. For example, the ethnicity of study participants was predominantly white Welsh, with only one patient from a non-white Welsh background, who, in addition, had recently relocated from England to Wales. The ethnicity of the study researcher, however, was Welsh-Asian, and the researcher thought that this might have influenced both the openness of the encounter and the development of rapport between researcher and study participant. Furthermore, the researcher clarified her role during the narrative inquiry and emphasized to study participants her nonclinical background and knowledge. The positioning of the researcher as a nonexpert among study participants contributed to developing equitable researcher-participant relationships and facilitated informal conversations around the retelling of experiences and events. The research team readily acknowledged that articulating a narrative can be influenced by the researcher's presence, the aims of the research itself, and how comprehensible the narratives might be.

The research team also acknowledged that the study researcher may have arrived at a different interpretation of the narratives in the "Books of Experience" and interviews than the team, following discussions during group analysis meetings. It is well documented (Bruner 2003; Riessman 2008; Maynes et al. 2012, p. 16) that narrative inquiry involves a process of shared meaning-making for the storyteller, research team, and audience in a social environment that is shaped by historical, cultural, and political or theoretical viewpoints. The "Lives at Risk" study team adopted a peer-group analysis approach to the "Books of Experience," which involved both study researcher and analysis team discussing their interpretations together through ongoing narrative analysis meetings during the whole study. Team meetings were extensive and undertaken to create an analytic research conversation by the wider team which, over time, helped hone down the data and clarify key issues arising.

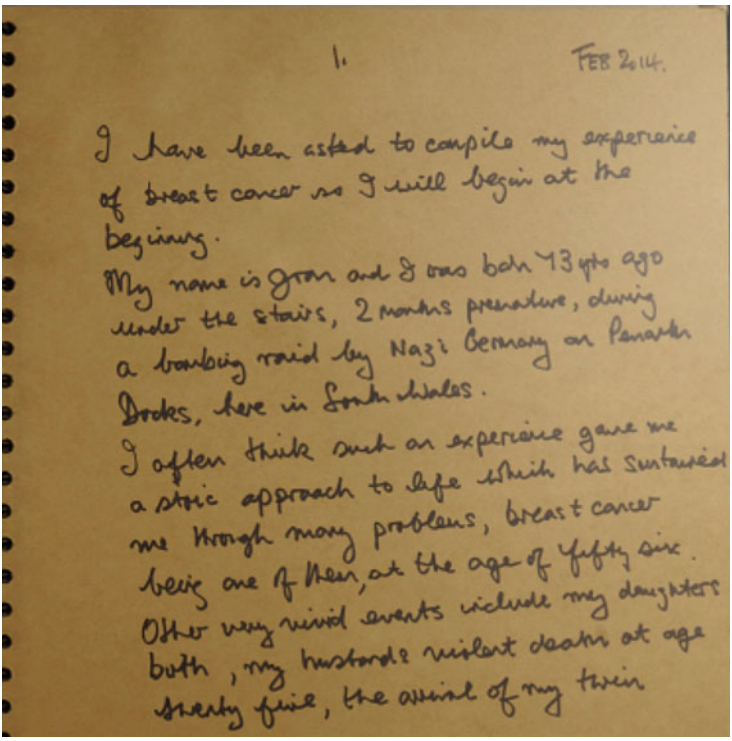
The following section outlines the study findings within the conceptual framework of Connelly and Clandinin's (2006, pp. 477–487) three cornerstones of narrative inquiry, supported by Landman's (2012, p. 30) four levels of analysis (linear, relational, emotional, and analytical).

Temporality

In the "Lives at Risk" study, women's narratives were strongly linked to their own personal and social discoveries and experiences, to which others have alluded (Maynes et al. 2012, p. 16), as well as their health and illness stories. There was a sense among study participants that they were struggling to take control of the transience of health and illness experiences, as they tried to ground their experiences in a more controllable, less quickly paced and less fleeting, reality. They did this by, for example, recording appointments meticulously, recording events in a diary, and counting the number of days taken off from work due to ill-health. The "Books" enabled women to create a longer-lasting legacy in some respects than might otherwise have been the case, bearing witness to their fear, pain, and regaining of health, and for this they were entirely grateful. The bio-photographic method allowed them to capture moments in time which enabled them to slow down the pace of their lives, securing moments that could, in effect, be bound and stilled between the pages of their "Books." The "Books," in essence, became not only active witnesses to these women's experiences but also evidence of moments in time and points of reflection, within which women could communicate their innermost thoughts about "who they were" and "how they lived" (Reissman 1993; Elliott 2005; Riessman 2008).

The "Books" became both creative elements and significant elements of their illness narratives. During their creation, they represented particular moments of ill-health and good health to which individuals attributed meaning. Women explored their own identities and considered how they could live with the risk of breast cancer and to what extent were they willing to renegotiate their identity to either accept or reject their current state of ill-health. According to Bruner (2003), outside the world of chronic ill-health, people are less likely to pause and take part in such conscious efforts of meaning-making.

The intersection between a life without cancer and a life with the possibility of cancer, or the confirmation of cancer, in any given moment in time, was one of the most striking features in these "Books." The temporality of women's sense of well-being was communicated clearly through this mixed media. One moment study participants illustrated and conveyed their wellness in terms of their health and daily routines, and the next they conveyed the chaotic array of emotions they were experiencing, following the news that they had been diagnosed with cancer, or told that they were "at risk" of developing breast cancer, through long textual diary-style transcriptions. Many study participants compartmentalized aspects and events into "before and after" cancer, acknowledging that they would not physically be the same person ever again, indicative of the importance of temporality in their lives. There was also a sense of loss, which was turned, by many, into an occasion for positive action. Women created new narratives of living, thinking, and acting, to fit in with their new identity (Ricoeur 1991). Women used their new-found knowledge, confidence, and skills to share their experience with others. Some aimed to make the most of life by traveling and experiencing new


1. FEB 2014.

I have been asked to compile my experience of breast cancer so I will begin at the beginning.

My name is Jean and I was born 73 yrs ago under the stairs, 2 months premature, during a bombing raid by Nazi Germany on Penarth Docks, here in South Wales.

I often think such an experience gave me a stoic approach to life which has sustained me through many problems, breast cancer being one of them, at the age of fifty six. Other very vivid events include my daughters both, my husbands violent death at age twenty five, the arrival of my twin

Extract 1 Temporally Meaningful Episodes in a Patient's Life

social worlds; others raised money for charity or undertook voluntary work. The temporal nature of time also brought on a new appreciation of life to many of these women and of the people they cared for and, in so doing, strengthened family and community bonds. However, this was not the experience for all women, and we would follow Atkinson and Rubinelli's lead (2012) by cautioning researchers against forming simplistic typologies of certain narratives (Atkinson and Rubinelli 2012).

For some women it was important to locate the experience of breast cancer as an episode within their whole life story, in order to gain an understanding of its cause and effect.

As shown in Extract 1 for one study participant, her "Book" helped her to reflect on her breast cancer experience by looking back on "temporally meaningful episodes" in her life (Polkinghorne 1988, p. 1). This study participant sidestepped the medical definition of cancer as a naturally occurring biological formation and proceeded to look for her own meaning for explaining the cause of her cancer. Later on in her narrative, she tried to locate some point in time when she thought her exposure to environmental risks could have caused her cancer.

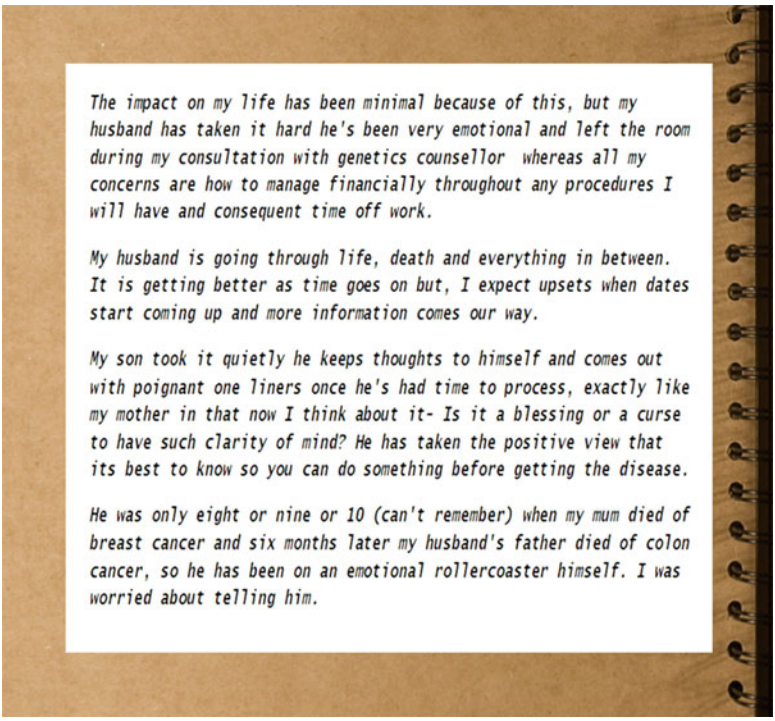
Sociality

To reiterate, the purpose of the "Lives at Risk" study was to explore meanings attributed to the notion of "risk" among women who had previously had an episode of breast cancer or had been assessed to be at risk of breast cancer as a result of a genetic predisposition. The research team were surprised that in some instances the topic of "risk" was altogether absent from participants' narratives. Indeed, the study researcher noted that some study participants found it difficult to recall any discussions around risk that had taken place with members of their medical team. In many cases women thought discussions around risk had been based on their relationship with the consultant rather than on any standard practice. In this respect, the consultant oncologist was identified as the individual controlling the disclosure of information, and clinicians were often described as being adept at using their own language, which conveyed both hope and more foreboding ambiguity, which often left women unsure about their future. This view aligns with the literature on the topic (Del Vecchio Good et al. 1994) which suggests that women seek definitive answers in relation to questions about their mortality, while clinicians appear unable to provide such clarity. Study participants often reported that the statistical probability of risk conveyed by clinicians was meaningless. Many understood it as "numbers and averages" and wanted to know if they had been cured or how long they had left to live. However, for the minority of participants, as the example below indicates, statistics were not meaningless, but meaningful and desired.

> I wanted statistics. I wanted figures. I wanted to know my chances . . . the oncologist spent quite a long time asking 'are you sure?' Lots of 'you don't have to, many people prefer not to know, there's no saying which group you'll be in.' I was adamant. He showed me the NHS'Predict' tool results.

The analysis of the bio-photographic "Books of Experience" and the bio-photographic elicitation interviews highlighted that participants' interpretations of living at risk depended not only on their understanding but also on others' perceptions of these issues, including family members, clinicians, and friends. Study participants reported that the dynamics within their own families and among their friendship groups changed according to the extent to which risk was understood, accepted, and, in some cases, rejected by individuals who shared their social worlds. Women acknowledged a range of reactions from friends and loved ones, some expected and some unexpected, which affected how they managed and lived with their "at risk" classification. For example, some women described wrestling with their own emotions as well as those of others, and acknowledged that this often posed additional distress, sometimes sending them into an isolated and lonely state, when they most needed the support of others.

The extract below, Extract 2, illustrates how this particular study participant played down her own emotional well-being and highlights the challenges she faced in managing the emotional well-being of her family, as they struggled to


The impact on my life has been minimal because of this, but my husband has taken it hard he's been very emotional and left the room during my consultation with genetics counsellor whereas all my concerns are how to manage financially throughout any procedures I will have and consequent time off work.

My husband is going through life, death and everything in between. It is getting better as time goes on but, I expect upsets when dates start coming up and more information comes our way.

My son took it quietly he keeps thoughts to himself and comes out with poignant one liners once he's had time to process, exactly like my mother in that now I think about it- Is it a blessing or a curse to have such clarity of mind? He has taken the positive view that its best to know so you can do something before getting the disease.

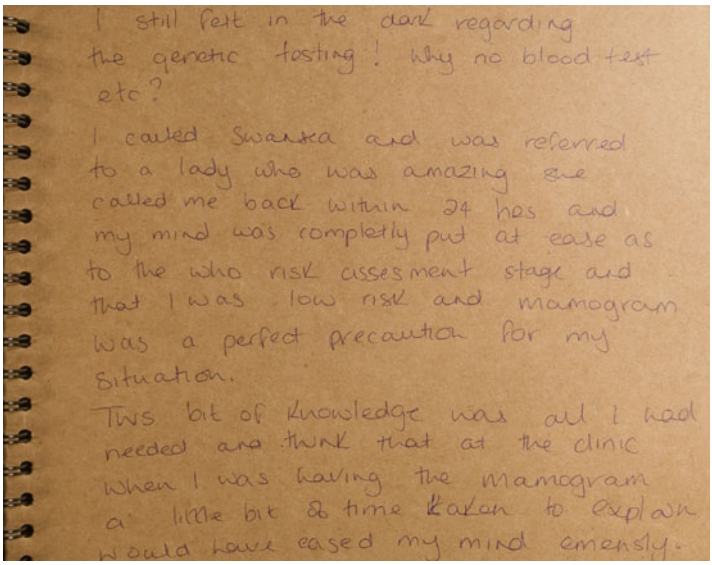
He was only eight or nine or 10 (can't remember) when my mum died of breast cancer and six months later my husband's father died of colon cancer, so he has been on an emotional rollercoaster himself. I was worried about telling him.

Extract 2 Patient's Concerns for Family Involvement in Ill-Health

come to terms with her risk of breast cancer. It is clear from the writing that the participant valued herself in terms of what she could offer others. The study participant used her "Book" to emphasize that social and financial aspects and her family's reaction to her ill-health took precedence over her medical story of ill-health. The ensuing medical interjections are given less importance and thought. This also illustrates the "sociality" of these narratives and provides a useful illustration of Frank's observation that "ill people are wounded, not just in body but in voice" (Frank 1995, p. 19). Within this one section of narrative, we get a glimpse into this woman's world and the interconnectedness of experiences of ill-health, juxtaposed by everyday concerns of life, work, and relationships she appeared to be grappling with.

Place

The "Lives at Risk" study participants frequently tried to associate their breast cancer experience with specific physical places they found themselves in, or thought about, or that were personally meaningful. Many study participants referred



Extract 3 Patient Expectation for Appropriate Professional Support

to their cancer as a "journey," both metaphorically and literally, in terms of mapping places associated with their illness and care. For example, one study participant described her journey by train to hospital to attend a breast screening appointment as the starting point of her care. The entry into her "Book of Experience" is deeply reflective, and she recollects a strong feeling of apprehension coupled with disappointment when the medical team unsuccessfully managed her anxiety. She explained that all she required was appropriate information during the care process to stabilize her emotions (see Extract 3).

For many, hospital waiting rooms and treatment rooms were places of dread. In some "Books" these places held silence; in others, silence arose from the emotional and physical impact of going through chemotherapy and radiotherapy treatment. The lack of detail in many narratives about the places where the treatment was conducted may suggest that this was not a salient issue for these women or that women may not have wished to relive the whole experience of treatment over again by writing about those areas. These places were often described in the interviews as painful to recall. Indeed, during one of the interviews, the researcher asked a participant about the treatments she had undergone, and in what appeared to be a desire to maintain agency and control, the women only briefly mentioned her treatment in response. On occasion, treatment and treatment places were redefined to appear manageable and under control. One woman called chemotherapy her "angel therapy" using her own religious beliefs to symbolize illness (Riessman 2000) in this case, in terms of the cure received from angels (see Extract 4).

Extract 4 Patient's Symbolization of the Illness Experience

Finally, study participants reported that they had recreated and revised their views about their future in light of the "philosophical truth" that they had discovered about the fleeting nature of life following their diagnosis. The study team recognized that the "Lives at Risk" narratives emulated the "modernist logic of the triumphant individual-one who has suffered, survived, and surpassed" (Plummer 1995, p. 34), but also that narratives were interwoven with the making of the self and provided women with the opportunity to consider their future and shape the stories they were yet to experience. Thus, narrative inquiry in this study supported the premise that "there is more than one story to tell" to enable "ethically conscious and thoughtful" conduct toward those experiencing ill-health (Carson 2001, p. 202).

This chapter has laid out a set of philosophical and ethical issues with respect to narrative inquiry, primarily with reference to Connelly and Clandinin's (2006) tripartite conceptual framework of narrative "temporality," narrative "sociality," and narrative "place," refracted through Landman's (2012) fourfold levels of analysis: the "linear," the "relational," the "emotional," and the "analytical." This form of narrative inquiry has been illustrated through a consideration of one particular case study that used narrative inquiry to ascertain what it means for women to live with the "risk" of developing breast cancer. The "Lives at Risk" study demonstrated how an attention to the interplay of narrative "temporality,"

narrative "sociality," and narrative "place" can reveal profound insights into women's stories of illness, health, and well-being. As a powerful approach to qualitative research, narrative inquiry has the potential to play an important role in contributing to more insightful, sensitive, and humane discourses in the context of health services research.

Summary Points

- Narratives contribute to meaning-making and a way of knowing the self.
- Narratives mirror the self and allow personal inner thoughts to be conveyed to others.
- The storyteller and the listener both influence the formation and interpretation of narratives.
- In health services research contexts, the process of narrating and listening to narratives needs to be managed and conducted within clear ethical boundaries.

Acknowledgement The research study referred to in this chapter was funded by Tenovus Cancer Care charity, Wales, for which the research team is entirely grateful.

References

Atkinson S, Rubinelli S (2012) Narrative in cancer research and policy: voice, knowledge and context. Crit Rev Oncol Hematol 84:S11–S16

British Sociological Association (2002) Statement of ethical practice for the British Sociological Association. Cambridge, Mass: Harvard University Press

Bruner JS (2002) Narratives of human plight: a conversation with Jerome Bruner. In. Charon R, Montello M (eds) Stories matter: the role of narrative in medical ethics. Routledge, New York

Bruner JS (2003) Making stories: law, literature, life. Cambridge, MA: Harvard University Press

Carr D (1986) Time, narrative, and history. Indiana University Press, Bloomington, Indianapolis

Carson AM (2001) That's another story: narrative methods and ethical practice. J Med Ethics 27 (3):198–202

Clandinin DJ, Rosiek J (2007) Mapping a landscape of narrative inquiry: borderland spaces and tensions. In: Clandinin DJ (ed) Handbook of narrative inquiry: mapping a methodology. Sage, Thousand Oaks, pp 35–75

Connelly FM, Clandinin DJ (2000) Narrative inquiry: experience and story in qualitative research. Educational Researcher, vol 6. Jossey-Bass, San Francisco, pp 94–118

Connelly FM, Clandinin DJ (2006) Narrative inquiry. In: Green JL, Camilli G, Elmore PB (eds) Handbook of complementary methods in education research. Lawrence Erlbaum, Mahwah, pp 375–385

Del Vecchio Good M, Munakata T, Kobayashi Y, Mattingly C, Good BJ (1994) Oncology and narrative time. Soc Sci Med 38(6):855–862

Elliott J (2005) Using narrative in social research: qualitative and quantitative approaches. Sage, London

Estroff SE (1995) Whose story is it anyway? Authority, voice, and responsibility in narratives of chronic illness. In: Chronic illness: from experience to policy. Indiana University Press, Bloomington, pp 77–102

Frank A (1995) The wounded storyteller: body, illness and ethic. University of Chicago Press, Chicago
Frank AW (2010) Letting stories breathe: a socio-narratology. University of Chicago Press, Chicago
Green B (2013) Narrative inquiry and nursing research. Qual Res J 13(1):62–71
Landman T (2012) Phronesis and narrative analysis. In: Flyvbjerg B, Landman T, Schram S (eds) Real social science: applied phronesis. Cambridge University Press, Cambridge, pp 27–47
Marmon Silko L (1996) Yellow woman and a beauty of the spirit: essays on Native American life today. Touchtone, New York
Maynes MJ, Pierce JL, Laslett B (2012) Telling stories: the use of personal narratives in the social sciences and history. Cornell University Press, Ithaca
Nettelbeck A (2008) The transfigured body and the ethical turn in Australian illness memoir. J Med Humanit 29(3):163–172
Pinnegar S, Daynes G (2007) Locating narrative inquiry historically: thematics in the turn to narrative. Sage, Thousand Oaks
Plummer K (1995) Telling sexual stories: power change and social worlds. Routledge, London
Polkinghorne DE (1988) Narrative knowing and the human sciences. State University of New York Press, Albany
Polkinghorne DE (2007) Validity issues in narrative research. Qual Inq 13(4):471–486
Reissman C (1993) Narrative analysis: qualitative research methods series 30. Sage, Boston University
Ricoeur P (1991) Life in quest of narrative. In: On Paul Ricoeur: narrative and interpretation. Routledge, London, pp 20–33
Ricoeur P (1991) From text to action: essays in hermeneutics, II (Studies in phenomenology and existential philosophy) (trans: Thompson JB, Blamey K). Northwestern University Press, Evanston, Illinois
Ricoeur P (1992) Oneself as another (trans: Blamey K). The University of Chicago Press, Chicago
Riessman CK (2000) Stigma and everyday resistance practices childless women in South India. Gend Soc 14(1):111–135
Riessman CK (2008) Narrative methods for the human sciences. Sage, Thousand Oaks
Skultans V (2000) Narrative illness and the body. Anthropol Med 7(1):5–13
Strawson G (2004) Against narrativity. Ratio 17(4):428–452
Wang CC, Geale SK (2015) The power of story: narrative inquiry as a methodology in nursing research. Int J Nurs Sci 2(2):195–198
Wiggs CM (2011) Mothers and daughters: intertwining relationships and the lived experience of breast cancer. Health Care Women Int 32(11):990–1008

Delusions: A Project in Understanding

35

KWM Fulford and Tim Thornton

Contents

Abstract

This chapter gives an illustrated overview of recent philosophical work on the concept of delusion. Drawing on a number of case vignettes, examples are given of the wide range of theories that has been advanced to explain this most challenging of experiences. Some have agreed with the philosophical founder of modern descriptive psychopathology, Karl Jaspers, that delusions are (empathically at least) "ununderstandable." The large majority, though, has sought to understand delusion in terms of aberrations of one kind or another either of beliefs (or related mental contents such as imaginings) or of the grounds or preconditions

K. Fulford (✉)
Faculty of Philosophy, University of Oxford, Oxford, UK
e-mail: kwm.fulford@philosophy.ox.ac.uk

T. Thornton
College of Health and Wellbeing, University of Central Lancashire, Preston, Lancashire, UK
e-mail: TThornton1@uclan.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_23

for beliefs. As a project in understanding, these theories offer helpful insights. A further group of theories focuses on the agential aspects of delusion as reflected, for example, in their role in the insanity defense. Agential theories converge with the person-centered approaches of contemporary empirical and clinical work on delusion. Such approaches bring an additional level of complexity in the form of delusions that are not pathological but adaptive in the life of the person concerned. As such they show the challenge of understanding delusion to be a project not just of understanding but of mutual understanding.

Introduction

The challenge of delusion, wrote the philosopher of mind and consciousness Naomi Eilan, is to "solve simultaneously for understanding and utter strangeness" (Eilan 2000, p. 97). Individually, many delusions are indeed strange – that "I have a nuclear reactor inside me," for example. Yet delusions collectively, the remarkable range and diversity of delusions, is stranger still.

It is with delusions collectively that we are concerned in this chapter. Through a series of case examples, some brief, others more extended, we show how the variety of delusions is matched by a corresponding variety of philosophical claims to understanding. Each of these adds useful insights. None, we believe, meet Eilan's double challenge in full. None solve simultaneously for understanding and utter strangeness across the variety of delusions as a whole. Taken together though, as we will see, these philosophical theories converge with contemporary clinical and empirical approaches in which the utter strangeness of delusions is embraced in a shared project of understanding between clinician and patient.

The chapter opens with the story of a real though biographically disguised person called "Simon" (from Jackson and Fulford 1997).

Simon's Story

Simon (40) was a senior, black, American lawyer, from a middle-class, Baptist family. Before the onset of his symptoms, he reported sporadic, relatively unremarkable, psychic experiences. These had led him to seek the guidance of a professional "seer," with whom he occasionally consulted on major life events and decisions.

He now faced a situation where his hitherto successful career was threatened by legal action from a group of his colleagues. Although he claimed to be innocent, mounting a defense would be expensive and hazardous. He responded to this crisis by praying at a small altar that he set up in his front room. After an emotional evening's "outpouring," he discovered that the candle wax had left a "seal" (or "sun") on several consecutive pages of his bible, covering certain letters and words. He described his experiences thus. "I got up and I saw the seal that was in my

father's bible and I called X and I said, you know, 'something remarkable is going on over here.'" "I think the beauty of it was the specificity by which the sun burned through. It was . . . in my mind, a clever play on words." Although the marked words and letters had no explicit meaning, Simon interpreted this event as a direct communication from God, which signified that he had a special purpose or mission.

From this time on, for a period of about 18 months, Simon received a complex series of "revelations" largely conveyed through the images left in melted candle wax. He carried photos of these, which left most observers unimpressed, but they were, for him, clearly representations of biblical symbols, particularly from the book of Revelations (the bull, the 24 elders, the arc of the covenant, etc.). At other times he described thoughts coming into his head: ". . . the things that come are not the things that I have been thinking about . . . they kind of short circuit the brain, and bring their message."

All these experiences meant nothing to Simon's family or friends but for him they signified that "I am the living son of David . . . and I'm also a relative of Ishmael, and . . . of Joseph." He was also the "captain of the guard of Israel." He found this role carried awesome responsibilities: "Sometimes I'm saying – O my God, why did you choose me, and there's no answer to that." His special status had the effect of "increasing my own inward sense, wisdom, understanding, and endurance" which would "allow me to do whatever is required in terms of bringing whatever message it is that God wants me to bring."

He expressed these beliefs with full conviction, "The truths that are up in that room are the truths that have been spoken of for 4000 years." When confronted with skepticism, he commented: "I don't get upset, because I know within myself, what I know."

Delusional Perception

The chapter starts with Simon's story because his experiences are paradigmatic not just of delusion but of a specific kind of delusion – called "delusional perception" – with particular diagnostic significance in psychiatry. One of us (KWMF) uses Simon's story in teaching sessions with trainee psychiatrists. In these sessions, trainees learn that "delusional perception" is defined in the Present State Examination (PSE) as a delusion that is ". . . based on sensory experiences . . ." and involves ". . . suddenly becoming convinced that a particular set of events has a special meaning" (Wing et al. 1974, symptom 82, pp. 172–173).

The PSE is a structured interview schedule covering psychiatric symptoms that are known from empirical studies to be both reliably identifiable and diagnostically significant. "Reliably identifiable" means that there is a high degree of agreement (from one observer to the next and from one occasion to the next) over whether the symptom in question is or is not present. The diagnostic significance of delusional perception is that it points to schizophrenia or some other psychotic illness. The ICD, for example, the World Health Organization's (1992) *International Classification of Diseases*, defines schizophrenia by the presence of delusional perception (or other

diagnostic symptoms) for at least 1 month. Other diagnostic possibilities in Simon's case would include hypomania (because of the grandiosity of his experiences) and a variety of possible organic disorders (e.g., a tumor in a part of the brain called the temporal lobe may produce psychotic experiences with religious or spiritual content).

But what exactly is it that marks Simon's experiences out as delusional? The sections that follow outline four main groups of philosophical answers to this question: that delusions (1) are mere nonsense, making no sense at all, (2) reflect some form of aberrant belief or belief structure, (3) arise from a disturbance in the grounds for or preconditions of belief, and (4) are characteristic failures of agency resulting from distortions in what the philosopher John McDowell (1994) has called "the space of reasons." To these are added a fifth and stand-alone section concerned with delusions that, contrary to widespread clinical as well as philosophical assumptions, are not pathological but adaptive in the life of the person concerned.

Clearly, in the space available, little more can be done than indicate exemplars of these five ways of understanding delusion each of which as will be indicated adds helpful insights. The accounts given of individual approaches are thus necessarily brief and there are inevitable omissions. The hope nonetheless is that in, as it were, summing across theories, the chapter will indicate the extent to which philosophical work on delusion is converging with clinical and empirical approaches in trying to meet Eilan's challenge of solving simultaneously for understanding and utter strangeness.

Delusions as Mere Nonsense

It is perhaps the utter strangeness of delusion that led no less a figure than one of the founders of modern descriptive psychopathology, the early twentieth-century philosopher-psychiatrist Karl Jaspers, to characterize delusions as "ununderstandable": "The most profound distinction in psychic life," Jaspers wrote (1913, p. 577), "seems to be that between what is meaningful and allows empathy and what in its particular way is ununderstandable, 'mad' in the literal sense, schizophrenic psychic life." Among many other contributions to modern accounts of delusion, Jaspers noted a number of features of a patient's experiences that signaled possible delusional thinking. Jaspers steered clear of using these to define delusion. (The closest he comes to a definition is, as described below, in terms of ununderstandability.) But the features he described have subsequently become the basis of textbook definitions: philosopher Rom Harré and psychologist Roger Lamb's (Harré and Lamb 1987) *Encyclopedic Dictionary of Philosophy and Psychology*, for example, draws on these features in defining delusion as ". . . a **false belief**, held **despite evidence to the contrary**, and one which is **not explicable in terms of the patient's educational and cultural background**. It is held with **complete conviction** and **cannot be shaken by argument**" (p. 142, emphases added).

As indicated, Jaspers considered these features more as pointers to or markers of delusion rather than as constitutive characteristics. Many delusions, indeed, as later

examples in this chapter will illustrate, lack one or more of these features. Simon, though, broadly satisfies the standard definition: his beliefs are (to all appearances) false; he is not swayed by evidence including the skepticism of his own (Baptist) peer group; and he holds his beliefs with complete conviction and is in no way swayed by argument.

Strange, then, to repeat Eilan's apt description, are Simon's experiences. Strange indeed to the point that Jaspers took delusions of the kind experienced by Simon as being beyond understanding. As such they represented a kind of limit case of his central thesis that psychopathology uniquely requires a twin-track "causes and meanings" approach. Psychopathology in general, Jaspers argued, demands in equal measure causal explanations and meaningful understanding. Causal explanations are established through the brain and other empirical sciences: there is, for example, a growing body of neuroscientific evidence of changes in the brain functioning in at least some kinds of delusional disorder (Sass and Byrom 2015, give a number of examples). Meaningful understanding, on the other hand, is acquired through empathic engagement and it is this that (on Jaspers account) breaks down in the case of delusion. Delusions are, he believed, beyond empathic understanding. They are, literally, "ununderstandable."

Contemporary defenders of Jaspers include the Finnish philosopher-psychiatrist Markus Heinimaa. Heinimaa has argued that delusions are in a particular but at the same time practically significant sense "incomprehensible" (2003). The historian and psychiatrist German Berrios has argued similarly that delusions are empty speech acts the "contents" of which are but random fragments of information "trapped" in the very moment that a delusion becomes crystallized (Berrios 1991, p. 12).

Such accounts, though, notwithstanding Jaspers' authority, are nowadays very much in the minority. Contemporary approaches, philosophical and empirical, are to the contrary aligned in seeking to prove Jaspers wrong. It is with one such approach, a group of approaches really, aimed at understanding delusion as being, in one way or another, a form of aberrant belief that the next section is concerned.

Delusions and Beliefs

With exemplars like Simon in mind, it is difficult to resist the idea that delusions have something to do with beliefs. Simon expresses his experiences primarily in terms of beliefs. His beliefs moreover, although indeed strange and hence difficult to understand, come across as anything but empty speech acts. He calls a friend to share his experiences; he is in awe of the revelations they bring; they impart "sense, wisdom, understanding, and endurance"; and he rejects well-intentioned but skeptical pushback from peers.

Focusing on this aspect of delusion, philosopher Brendon Maher, reversing Jaspers approach, argued that delusions are essentially normal beliefs arising in response to, and fully understandable as attempts to rationalize, anomalous experiences (Maher 1999). One such experience – thought insertion – is hinted at in Simon's story where he speaks of "... the things that come are not the things that

I have been thinking about . . . they kind of short circuit the brain, and bring their message." A widely cited example of thought insertion is included in a collection of first-hand reports by psychiatrist C.S. Mellor (1970).

A 29-year-old woman said: "I look out of the window and I think the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews (a celebrity at the time) come into my mind. There are no other thoughts there, only his . . . He treats my mind like a screen and flashes his thoughts onto it like you flash a picture."

Thought insertion is surely the epitome of an anomalous experiences: it involves having thoughts in your head that you are thinking (they are first personal thoughts) and yet which at the same time you experience as and believe to be *the thoughts of someone else*. As such they run contrary not only to everyday experience but also to traditional philosophical assumptions that first personal thinking and ownership of thought necessarily go hand in hand. Thought insertion has correspondingly been a focus of much recent work in philosophy and psychiatry (see, e.g., Stephens and Graham's 2000).

So just how would someone with thought insertion account for his or her experience? This is where Maher's normal belief account of delusion becomes relevant. Maher takes delusions to be rational (hence fully understandable) attempts to account for such experiences: you experience thoughts in your head as some other person's thoughts; it is understandable therefore that you should believe other person to be using your mind for his or her thinking.

There is though, as philosophers Max Coltheart and Martin Davies among others pointed out, an immediate difficulty with this approach, namely, the incorrigibility of delusions. Maher's account might explain why the delusional belief arises in the first place, but some further factor is needed to explain why it persists in the face both of counterevidence and of the availability of more plausible explanations. What is needed therefore, these and others have argued, is a two-factor rather than one-factor account of delusions (e.g., see Coltheart and Davies 2000; and Davies and Egan 2013, for a recent update).

A further point against Maher's one-factor account, as Davies and colleagues (2001) point out, is provided by occurrences of anomalous experiences in the absence of delusional belief formation. To the examples they cite might be added non-delusional examples even of thought insertion: the aura (prodromal symptoms) of temporal lobe epilepsy, for example, as the neurologist and psychiatrist Alwyn Lishman has described (personal communication), may include experiences of thought insertion. But patients with experiences of thought insertion arising from temporal lobe epilepsy, unlike those with thought insertion arising from psychotic disorders such as schizophrenia, readily accept their doctor's explanation that this is a symptom of something wrong with them.

That said, just what the requisite further factor in two-factor accounts of delusion is, has proved difficult to pin down. Without this a two-factor account simply doubles the challenge of understanding: it adds to the anomalousness of the initial experience an anomalous belief. Attention has thus moved back to the belief part of delusion with a range of theories exploring whether and if so in what way delusions may be understood as or as derived from aberrant beliefs.

A leading example of these doxastic theories, as they have come to be called, is philosopher Lisa Bortolotti's (2009) *Delusions and Other Irrational Beliefs*. Those opposed to the idea that delusions are beliefs point to differences between them and normal beliefs. Delusions, they say, echoing elements of the standard textbook definition (above), are unlike beliefs in being, for example, inferentially circumscribed (resistant to normal inferences), practically circumscribed (failing to issue in appropriate actions), and affectively circumscribed (lacking appropriate emotional coloring). The "other irrational beliefs" of Bortolotti's title thus captures a key insight, namely, that while some delusions may show one or more of these features, so also do beliefs of many other kinds, irrational and, indeed, normal. Political beliefs, for example, are often inferentially circumscribed, health beliefs (such as the belief one should take more exercise) are widely practically circumscribed, and beliefs about remote events (about poverty in another country, for instance) are typically affectively circumscribed.

Those opposed to doxastic theories, by contrast, have pointed out that features such as these are not exclusive to beliefs, delusional or otherwise. Philosopher Gregory Currie, for example, argues that such features show a better fit with acts of imagination or imaginings (Currie 2000). So an antidoxastic account starts with the idea that the subject imagines something. This accounts for the fact that delusions are practically circumscribed since people do not generally act on imaginings as they act on beliefs. But, on his account, the subject then misidentifies that mental act as a belief.

This it might be suggested makes Currie's account a second-order account of delusion as belief: delusion on his account is a mistaken belief about an imagining, which he calls a "cognitive hallucination." To account for the fact that delusions need not be completely practically circumscribed, Currie then suggests that this second-order belief can give rise to a first-order belief that in turn can motivate action. The resulting picture – from imagining, to a second-order belief, to a first-order belief – is thus rather complicated. Philosopher Andy Egan argues that delusions are a little like beliefs and a little like imaginings and hence are a kind of propositional attitude with aspects of both: sui generis "bimaginings" (Egan 2008). In yet a further ingenious variation on this theme, philosophers George Graham and Lynn Stephens have suggested that delusions are second-order "stances" toward any first-order mental state (belief or not) characterized in a sui generis way to match some key features of delusions (Stephens and Graham 2006).

How do doxastic theories and their antidoxastic opponents fare with respect to Eilan's dual challenge to philosophical accounts of delusion? Doxastic theories underline the inadequacy of standard textbook definitions of delusion. In marking the continuities between normal and pathological beliefs, they remind us that contrary to textbook definitions, beliefs in general (not just in the case of delusions), and other closely related contents of consciousness (such as imaginings), may be *held despite evidence to the contrary*, they may be *not explicable in terms of the patient's educational and cultural background*, and they are often held *with complete conviction and cannot be shaken by argument.* To this extent then these theories solve for Eilan's understandability. The same applies to antidoxastic rival attempts to

assimilate delusions to other propositional attitudes in order to capture aspects that do not accord with paradigmatic beliefs. If delusions really were misrecognized acts of imagination, that would shed understanding on why they did not directly guide actions. By the same token though, both doxastic and antidoxastic theories fail to solve for Eilan's "utter strangeness." In emphasizing the continuities between delusion and ordinary mental life in their contrasting ways, both simply fail to capture just how *extra*ordinary delusions really are.

To capture the strangeness of delusions, therefore, some have argued, it is necessary to dig deeper, turning attention from the form of belief (or related phenomena) to the underlying conditions for the very possibility of belief.

Delusions and the Grounds of Belief

Doxastic theories, perhaps not surprisingly, work best where delusions take the form of a well-systematized set of beliefs centered on a circumscribed topic (monothematic delusions as they are sometimes called). Much of the discussion of two-factor theories has been concerned with conditions like the Capgras syndrome, for example, in which the subject believes that people close to them have been replaced by look-alike fakes. Andrew Sims, in his now classic textbook of psychopathology, *Symptoms in the Mind*, gives the following example:

A woman who was at the time on an inpatient psychiatric ward asked about her husband: "Who is that person who drives my family up to the hospital every evening? It is a cheek. He stays at home and opens all my husband's letters. Anyway at least he pays the bills . . . He does look very like my husband only perhaps a little fatter" (Sims 1988, p. 97).

Those working in a more phenomenological tradition have however pointed out that these monothematic delusions are far from typical. The psychologist and phenomenologist Louis Sass, for example, connects the profound disturbance of "being in the world" shown by people with the more complex delusions of schizophrenia, to solipsism. Sass illustrates his thesis with the complex and wide-ranging delusions shown by the classic case of Paul Schreber (Sass 1994). The British phenomenologist and philosopher, Matthew Ratcliffe, argues similarly that the whole doxasticism versus anti-doxasticism debate misses the clinically important feature of delusions called delusional atmosphere (Ratcliffe 2009). Andrew Sims, again, describes delusional atmosphere thus:

> *"For the patient experiencing delusional atmosphere, his world has been subtly altered: 'Something funny is going on'. He experiences everything around him as sinister, portentous, uncanny, peculiar in an indefinable way. He knows that he personally is involved but cannot say how. He has a feeling of anticipation, sometimes even of excitement, that soon all the separate parts of his experience will fit together to reveal something immensely significant."* Associated with this is delusional mood in which "*The patient feels profoundly uncomfortable, often extremely perplexed and apprehensive.*" (Sims 1988, p. 89)

Delusional atmosphere may precede the emergence of delusions (including delusional perceptions of the kind shown by Simon, above). Ratcliffe connects delusional atmosphere to Heideggerian moods: deep conditions of possibility of thinking about the world. Shaun Gallagher, a North American phenomenologist, starting like Ratcliffe from delusional atmosphere, explores the idea of delusions in terms of alternative realities (Gallagher 2009).

Challenges to overly narrow conceptions of delusion, doxastic or otherwise, are to be expected from phenomenologists. The task of phenomenology, as Sass has put it, is to focus ". . . on delusion as a phenomenon, on its subjective or lived dimension: what it is like to have a delusion" (Sass and Byrom 2015, p. 164). This is why phenomenology was Jaspers' favored method for understanding psychopathological experiences. Challenges to doxasticism/anti-doxasticism however have come also from the analytic tradition of philosophy. The psychologist and philosopher, Richard Gipps, has argued that moving beyond a narrow focus on monothematic delusions shows the need for what he calls an "engaged" rather than an "estranged" epistemology (Gipps and Fulford 2004). That epistemological considerations are at the very least relevant to understanding delusions evident from the unusual (though by no means rare) cases of delusions as true beliefs:

> *Mr. A was seen by his general practitioner (GP) in connection with his wife's depression but turned out to have problems of his own. He complained of anxiety and his GP suspected that he had taken to drinking as a result. Pressed on this, Mr. A. suddenly announced that the real problem was that his wife was being unfaithful to him. He offered a wide range of reasons for believing this, mostly somewhat bizarre: for example, she had taken to doing her washing on a different day; and the pattern of cars parked in the street had changed.*
>
> *A psychiatrist confirmed the GP's diagnosis of Othello syndrome based on the presence of delusions of infidelity. Neither doctor had any doubt that Mr. A's beliefs about his wife were delusional. Yet both knew at the time they made their diagnosis that Mrs. A had become depressed following the break-up of an affaire.*

Delusions not uncommonly turn out to be true (the person with persecutory delusions who turns out to be being persecuted). The Othello syndrome, however, is regularly diagnosed as in this case when the belief in question (of infidelity in one's sexual partner) is known *at the time of making the diagnosis* to be true (Shepherd 1961). Recognizing this, delusions have sometimes been characterized not as "true" but as "unfounded" beliefs. In some instances there is, certainly, a suggestion of aberrant epistemology in the way those concerned seek to check the veracity of their beliefs. Mr. A (above) pointed to seemingly irrelevant facts such as that his wife *had taken to doing her washing on a different day* and that *the pattern of cars parked in the street had changed.* In other cases though, it is not clear just what kind of check would be relevant even in principle. The following unusual variant of hypochondriacal delusion – the paradoxical delusion of mental illness – is a case in point (like all the examples in this chapter, this is based on the story of a real though biographically disguised person):

Mr. MI was brought to A&E following an overdose. He had tried to kill himself, he explained to the duty psychiatrist, because he was "mentally ill and people who are mentally ill get put away". A second opinion confirmed a diagnosis of hypochondriacal disorder with delusions of mental illness and, given the evident risk of suicide, both doctors were ready to admit Mr A as an involuntary patient. In the event however, he accepted ordinary reassurance that people who are mentally ill do not get "put away" and arrangements were made for him to be seen as a psychiatric outpatient.

Delusions of mental illness are as has been indicated unusual. Hypochondriacal delusions are usually concerned with life-threatening physical illnesses such as cancer. But as philosopher Anthony Quinton first pointed out (Quinton 1985), the paradox they present is decisive (logical) evidence against delusions being essentially false beliefs: if delusions were indeed essentially false beliefs, then the delusion of mental illness would be a belief that if true is false and if false is true. So there would be no test by which the truth or otherwise of the delusion of mental illness could be checked even in principle.

John Campbell, a philosopher of mind who has worked extensively in philosophy and psychiatry, has taken the failure of normal checking shown by people with delusions to suggest that delusions involve a deviant version of what Wittgenstein called "framework" propositions (Campbell 2001). The Capgras subject, Campbell points out, does not carry out the canonical checking people would normally do if they believed that someone close to them was not the person they appeared to be. Mr. A (Othello syndrome) failed similarly in this respect. Mr. MI (delusion of mental illness) failed to offer (could perhaps not have offered) any checks at all. But such canonical testing, Campbell continues, would not be expected for Wittgensteinian framework propositions. For framework propositions are those largely tacit assumptions about the world that far from being open to checking people simply have to take for granted if they are to hold beliefs at all. Adapting one of Wittgenstein's examples, it would be entirely rational not to check whether this, holding up my right hand, is a hand and that this, holding up my left, is another. So delusions, Campbell suggests, are perhaps some kind of deviant framework proposition with a certainty beyond question. But unlike Wittgenstein's examples, these are not part of a world picture shared with others.

Just as the doxastic theories described in section "Delusions and Beliefs" solve for understanding while failing to solve for strangeness, so the "grounds of belief" theories described in section "Delusions and the Grounds of Belief" solve for strangeness while failing to solve for understanding. Disturbances in Ratcliffe's Heideggerian moods, in Sass' solipsism, in Gipps' engaged epistemology, or in Campbell's framework propositions, just to the extent that they sit beyond everyday awareness, are bound to be in Eilan's phrase "utterly strange." For the same reason, they are also necessarily beyond the reach of everyday empathic understanding.

The fourth and final group of theories of delusion to be considered here, however, focusing as these theories do on their agential nature, is concerned with an aspect of the strangeness of delusions that is in some respects at least accessible to everyday empathic understanding.

Delusions and Agency

One indication of the agential nature of delusions is their status as an excusing condition in law. The insanity defense as it is widely called, where a defendant is found not guilty by reason of insanity, has a long history. Philosopher and psychologist Daniel Robinson has traced it in various forms back to classical times and across diverse cultures (Robinson 1996). At the heart of the insanity defense is delusion. In the UK, for example, the McNaughton rules governing the admissibility of the insanity defense go back to a nineteenth-century case in which the eponymous Daniel McNaughton shot a stranger under the delusion that he was persecuting him. The story runs thus:

> *Daniel McNaughton was arrested by a police constable who witnessed him firing a pistol into the back of a man he had never met before. His victim, Edward Drummond, subsequently died of his wound and McNaughton was charged with murder. At his trial, Alexander Cockburn, argued in his defence that although he had indeed killed Edward Drummond McNaughton was "… the victim of a fierce and fearful delusion, which, after the intellect has become diseased, the moral sense broken down, and self-control destroyed, has led him on to the perpetration of the crime with which he now stands charged." Based on this argument the jury found McNaughton "… not guilty, by reason of insanity" and instead of being hung as a murderer he was admitted as a patient to Bethlehem Hospital.* (Based on West and Walk 1977)

The insanity defense illustrates the intuitive link between delusion and a very radical form of loss of agency. Delusion as the basis of the insanity defense is not merely a mitigating factor (e.g., "guilty but under duress") but a full-blown legal excuse ("not guilty (at all) by reason of insanity"). The person concerned, so this intuition goes, is simply not the agent of their action and hence cannot be held responsible.

A similar intuition lies behind the central place of delusion-defined psychotic disorders in involuntary treatment. The intuition justifying involuntary psychiatric treatment, like that justifying the insanity defense, is that the person concerned is not responsible for their actions. And like the insanity defense, this intuition is reflected in legislation in many countries across the world. The legal grounds of involuntary psychiatric treatment, it is true, are in general widely drawn requiring nothing more than, in one form or another, a combination of mental disorder and risk. In practice though, as a number of studies have shown, involuntary psychiatric treatment is used mainly for psychotic disorders the central symptom of which (as above) is delusion (Sensky et al. 1991).

It is worth emphasizing just how radical is the intuition of loss of agency in delusion. Shoot someone believing them (normal belief) to be persecuting you and you are guilty of murder: shoot someone believing them (delusional belief) to be persecuting you and you are not guilty of murder. This way of making the point challenges us to explain just why the deluded person should be considered not guilty.

This is an aspect of the strangeness of delusions. It is a strangeness moreover that is compounded by the fact that hard as it is to say just *why* the person who is deluded is not guilty, many people (including hard headed lawyers) actually *do* say this. Here, therefore, is an aspect of the strangeness of delusion that, strange as it is, people nonetheless seem to understand at least intuitively.

This combination of strangeness and intuitive understanding suggests a possible connection between delusions and practical reasoning, i.e., reasoning as in the reasons that people as agents have for their actions (Fulford 1989). Determining good or bad reasons in general, as philosophers John McDowell (1996), Jonathan Dancy (1993), and others have highlighted, is not a matter of having a universal theory of good reasons but is instead a context-dependent sensitivity. It is a matter of phronesis, of implicit or tacit understanding, rather than general explication. So if this is true of practical reasoning in general, it is perhaps no surprise to find that it is true of delusional practical reasoning in particular.

A further indication that delusions may have something to do with practical reasoning is provided by the observation that, again like reasons for action, delusions may take the form not only of beliefs about matters of fact but of value judgments (Fulford 1991). Delusions of guilt in depression are a case in point:

> *Mr. ED was admitted to hospital following a sudden deterioration in a long-term depressive illness. When asked if anything particular had happened recently he became tearful about the fact that he had forgotten to give his children their pocket money. His wife confirmed this adding that he had gone "completely over the top about it" saying that it was "some terrible sin", that he was "useless as a Dad" and that they "would all be better off if he was dead."*

Delusions of guilt are common in depression. They may take the form of beliefs as to matters of fact: one patient, for example, believed he had started a war, and, as most would, he felt guilty as a consequence. The delusional content of Mr. ED's thinking by contrast was the way he *evaluated* something trivial that he had actually done (forgetting his children's pocket money). Evaluative delusions expressing positive rather than negative values occur in the positive-affect counterpart of depression, hypomania. There is a suggestion of positive evaluative delusions in the grandiose content of some of Simon's beliefs (above).

Evaluative delusions, however, despite being commonplace, have been largely neglected equally in empirical and clinical as in philosophical accounts of delusion. This is because evaluative delusions, again like their counterpart reasons for action, carry the same practical implications as factual delusions. Delusions of guilt (pointing to depression) carry different implications from grandiose delusions (pointing to hypomania). But whether in a given case the delusion in question is factual or evaluative in form makes no difference at all to how the patient concerned is diagnosed and treated (including their treatment in law).

A potentially promising line of development of agential accounts of delusion would be by way of John McDowell's naturalistically enriched concept of the "space of reasons" (McDowell 1994). As the basis for such an account, McDowell's space of reasons (1) locates values alongside facts (as delusional values appear alongside

delusional facts) on an equal basis as features of the natural world, (2) shifts attention from the phenomenological forms of delusions to their context in what might be called the mental economy of an individual, and (3) connects legal intuitions about delusion by way of phronesis to judgments of rationality in general.

Non-pathological Delusions

Whether such an account plays out successfully remains to be seen. But as a route to understanding delusion, agential accounts connect with a further relatively neglected feature of delusions, namely, that notwithstanding their traditional identification as a central symptom of mental disorder, delusions may in some instances *not be pathological at all*. Simon's story, in the way that it actually worked out, illustrates this further and as will be seen, practically as well as theoretically important, feature of delusion:

> *Simon was empowered by his experiences to take on his accusers and being a lawyer he decided to run his own defence. In this his experiences not only empowered him but also gave guidance. They told him for example which books to turn to in the local law library and where he would find relevant cases. Just how this worked he couldn't say and no one else could read the messages as he received them. But the guidance was good. He won his case (it was shown to be a racially motivated attack on him); his reputation as a lawyer was enhanced; and his practice flourished to the point that he made enough money to establish a research trust for the study not of psychosis but of religious and spiritual experience.*

On the face of it then, Simon was not ill. Some might want to argue that he was indeed ill but that like other illnesses the course in his particular case was benign. Simon for one would have rejected such an interpretation and rejected it roundly: far from being pathological, benign or malign, his experiences were for him deeply religious in nature. Respecting Simon's understanding of his experiences is important, as will be described shortly, for contemporary person-centered approaches to healthcare. But Simon has a perhaps surprising ally for his self-understanding in psychiatry courtesy of the main international competitor to the ICD (above), the American Psychiatric Association's (2013) DSM (Diagnostic and Statistical Manual).

DSM differs from the ICD and other traditional approaches to psychiatric diagnosis in that it includes alongside the standard symptom-based criteria what it calls criteria of clinical significance. A diagnosis of schizophrenia in DSM thus requires that the patient satisfies *two* sets of criteria: first, a symptom-based criterion, its Criterion A, that is substantively the same as the criteria in ICD; second, an additional criterion of clinical significance, its Criterion B. Simon satisfies the symptom-based criteria in both classifications in virtue of his delusional perceptions (described above). But he *fails* to satisfy DSM's Criterion B. This is how in the latest edition of DSM (DSM-5) Criterion B for schizophrenia reads:

> *For a significant portion of the time since the onset of the disturbance, level of functioning in one or more major areas, such as work, interpersonal relations, or self-care, is markedly below the level achieved prior to the onset (or when the onset is in childhood or adolescence, there is failure to achieve expected levels of interpersonal, academic, or occupational functioning).* (American Psychiatric Association 2013 p. 99)

On the evidence then, of his functioning as a lawyer, Simon showed enhanced not diminished functioning. As just described he won his court case and his practice flourished. This in itself is not to endorse his understanding of what happened to him specifically as a religious experience. But in breaking the link between delusion and pathology, it at least provides space for non-pathological ways of understanding Simon's experience.

Simon, it is important to be clear, is not alone in exhibiting adaptive rather than pathological psychotic experiences. His story is one of a series collected by the British psychologist Mike Jackson (1997). Jackson and others have gone on to show that such experiences are indeed surprisingly common in the general population (see, e.g., Johns and van Os 2001). Contemporary authors were not the first to recognize this. Early twentieth-century philosopher and psychologist William James (1902) described delusion as religious experience turned upside down (i.e., delusions being identical phenomenologically with religious experience other than in having damaging rather than enhancing effects). Based on such observations, The British Psychological Society (2000) published a platform statement on psychosis arguing that it should be understood as a faculty concerned broadly with creativity and problem solving. Like other faculties it could go wrong and at one extreme going wrong meant a serious psychotic illness. But psychosis as such, the statement argued, is not pathological.

The challenges to understanding delusion in this way are considerable. Theoretically, the twin enabling/disabling nature of delusion adds a whole new level to the strangeness to be explained. Practically the distinction requires *inter alia* balancing a series of difficult value judgments: Criterion B requires not just a *change* in functioning but a change *for the worse.* Criterion B thus makes explicit an essentially evaluative element to the concept of delusion (Fulford and Radoilska 2012). Yet the challenges, theoretical and practical, are worth embracing to the extent not least that they reinforce contemporary person-centered clinical approaches to the management of delusion.

Once again, there is no space here to discuss such approaches in detail. The operative point is that the key features of contemporary person-centered approaches align closely with agential accounts of delusion in which the values of the individual concerned have a central place. Recent good practice guidance, for example, from the UK's Department of Health on values-based assessment in mental health gives central importance to the individual concerned being an active participant in rather than merely passive recipient of diagnostic assessment (The National Institute for Mental Health in England and the Care Services Improvement Partnership 2008). The intention here as in other areas of mental health is what has come to be called "coproduction" in which the clinician and patient are joint contributors. The aims of

care too are now more person centered in this specifically values-based sense. In traditional disorder-centered approaches, the aim is, straightforwardly, to get rid of the patient's symptoms (preferably by dealing with their underlying causes). The aim of the person-centered approaches of contemporary mental health practice is instead recovery of a good quality of life, where "good" means good as defined primarily by the values of the person concerned (Allott et al. 2005). Symptom control may well have a part to play in this, but only to the extent that it serves to support rather than prejudice recovery of the individual's quality of life.

Simon, whose story comes from over 20 years ago, avoided contact with mental health professionals. He feared, with good reason at the time, that his experiences, although at the top of his own scale of values, would be simply written off as pathological. It is not difficult to imagine how differently his story would have worked out had he been treated (quite possibly as an involuntary patient) with neuroleptic medication. His psychotic experiences instead of being adaptive and enhancing his quality of life could well have ended up maladaptive and deeply damaging to his quality of life. The British Psychological Society's platform statement (above) pointed to the risks of iatrogenically induced pathological psychosis arising from (well-intentioned) clinicians' assumptions of pathology. But of course in fighting shy of mental health professionals, Simon's story could have had other less positive outcomes. He might, for example, have turned out to have a brain tumor for which potentially lifesaving treatment could have been available.

So there is no quick way here. The challenges of understanding presented by delusion, no less for practice than for theory, are formidable. But what Simon's story and the stories of those like him show is that solving simultaneously for understanding and utter strangeness is a project of understanding not just of pathological delusions in their many forms, challenging as this is, but of non-pathological delusions as well.

Conclusions

In an early but still seminal contribution to philosophy and psychiatry, philosopher Anthony Quinton noted "Madness is a subject that ought to interest philosophers" and yet he continued, "they have had surprisingly little to say about it" (Quinton 1985, p. 17).

How differently matters stand now! Among many other burgeoning areas of enquiry in the still rapidly expanding field of philosophy and psychiatry, work on delusions has a central place. Of the theories covered in this chapter, some have mirrored Karl Jaspers' early twentieth-century account of delusions as simply ununderstandable. Most though have taken a more positive approach with interpretive analyses of delusion as, in one way or another, an aberration either of belief (or related mental contents such as imaginings) or of the grounds of or preconditions for belief. Yet other theories have focused on the agential aspects of delusion reflected in their role in such areas as the insanity defense. These agential aspects

in turn connect with the additional challenges presented by delusions that (as in Simon's story) are not pathological but enhancing. Other emerging lines of enquiry, not covered here, include work on the intersubjective (Fuchs 2015) and interpersonal (Ratcliffe 2015) nature of delusion and computational models, Bayesian (Mishara and Sterzer 2015) and non-Bayesian (Koralus and Mascarenhas 2013; Parrott and Koralus 2015).

The range of the philosophical approaches reviewed here corresponds with a broadly similar range of empirical and clinical approaches. Contemporary "multilevel" accounts of delusion, as the psychologist Philippa Garety has called them (Garety 2015), combine cognitive with emotional and behavioral factors. Garety, who as an empirical researcher was among the first to provide robust evidence of consistent cognitive biases in delusional thinking (Garety and Freeman 2013), describes these multilevel approaches in person-centered terms consistent with an agential understanding of delusion. Contemporary multilevel approaches, she says, involve "understanding the grounds for the person's belief – the unusual experiences and events underpinning it" – while at the same time "validating and empathizing with emotional distress" and "exploring with the patient, collaboratively, alternative possibilities, cognitive, emotional and behavioural, in the light of the person's history and social environment." This approach, she continues, works with "the process" "with the mode or manner of the (subject's) thinking rather than the content" (all from Garety 2015).

This chapter opened with Eilan's formulation of the double challenge presented by delusion as solving simultaneously for understanding and utter strangeness. Among the philosophical accounts reviewed here, those focusing on beliefs and other propositional attitudes, it has been suggested, solve (in part) for understanding but at the expense of strangeness, while those focusing on the grounds of belief solve (in part) for strangeness but at the expense of understanding. Agential accounts converging as they do with contemporary empirical and clinical person-centered approaches hold perhaps the best promise of solving simultaneously for understanding and strangeness though in a project that is not just of understanding but of mutual understanding.

Definitions of Key Terms

Delusion	The definition of delusion is controversial. Standard textbook definitions include a number of features: falsity of belief, incorrigibility (held despite evidence to the contrary and in the face of counter argument), and cultural dissonance (not explicable in terms of the patient's educational and cultural background). But many clinical examples of delusion lack one or more of these features.
Delusional perception	A particular kind of delusion in which the patient suddenly becomes convinced that some otherwise

	trivial observation or event has for them a special meaning.
Delusional atmosphere	A sense that something is not quite right, that something is going on, without being able to say just what. If persistent, it may resolve into a fully formed delusion (often by way of a delusional perception). May be associated with affective changes (ranging from fear through anticipation to excitement) sometimes differentiated as "delusional mood."
Monothematic delusion	A delusional belief or set of beliefs developed around a single theme and often with limited effects on the subject's other beliefs and actions.
Thought insertion	The experience of having thoughts that although first personal are nonetheless attributed to some other person or agency. Thought insertion occurs in delusional and non-delusional forms.
Capgras syndrome	A monothematic delusion in which the person concerned believes that people (usually one or more family members) have been replaced by look-alike doubles.
Doxastic	Of or related to beliefs.
Framework propositions	A term attributed to philosopher Ludwig Wittgenstein that refers to the wide range of largely tacit beliefs that are normally just taken for granted in finding our way around the world: for example, that "this is my right hand and this my left" or that "this chair will still be there when I sit down."
Agential	Of or related to agents usually where an agent is an initiator of actions.
Insanity defense	A form of legal excuse in which a plaintiff is found "not guilty by reason of insanity."
Involuntary treatment	Treatment that is given without the consent of the patient. Involuntary psychiatric treatment is particularly controversial.
Practical reasoning	A term derived from the classical philosopher Aristotle and used of the kind of reasoning that is characteristic of human beings as agents: in modern philosophy it is often thought to combine two elements variously described (e.g., as beliefs and desires or facts and values) although this is denied by Aristotelian philosophers such as McDowell.
Space of reasons	A phrase borrowed from philosopher Wilfrid Sellars by John McDowell and now used to capture the idea

	of an enriched concept of the "natural" encompassing besides the "facts" of natural science a range of normative elements (including but not limited to values).
Non-pathological delusions	Delusions occurring in the absence of evidence of pathology. Non-pathological delusions, as in some forms of spiritual and religious experience, may be positive and adaptive in the life of the person concerned.
Criterion B	A diagnostic criterion for schizophrenia introduced in the American Psychiatric Associations' Diagnostic and Statistical Manual (DSM). As a criterion of social and occupational functioning, Criterion B is additional to the standard symptom-based criteria (covered by Criterion A in DSM). Similar "criteria of clinical significance" are used throughout DSM.

Summary Points

- Delusion is a major focus of the rapidly expanding field of philosophy and psychiatry.
- Philosopher Naomi Eilan has described the interpretive challenge of delusion as solving simultaneously for understanding and for utter strangeness.
- Contemporary philosophical theories fall into three main groups exploring the nature of delusions, respectively, as (1) aberrant forms of belief (or related mental contents such as imaginings), (2) disturbances in the grounds of or preconditions for belief, and (3) defective practical reasoning or agency.
- Considered separately, these theories each contribute important insights for particular kinds or classes of delusion, but none solves simultaneously for understanding and strangeness across the variety of delusions as a whole.
- Taken together, contemporary philosophical theories of delusion converge with person-centered empirical and clinical approaches in which the project of understanding delusion becomes a project of mutual understanding between the clinician and the individual concerned.

References

Allott P, Fulford KWM, Fleming B, Williamson T, Woodbridge K (2005) Recovery, values and e-learning. Mental Health Rev 10(4):34–38

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (fifth edition, DSM-5). American Psychiatric Association, Washington, DC

Berrios G (1991) Delusions as "wrong beliefs": a conceptual history. Br J Psychiatr 159(Suppl 14):6–13
Bortolotti L (2009) Delusions and other irrational beliefs. Oxford University Press, Oxford
British Psychological Society (2000) Recent advances in understanding mental illness and psychotic experiences. The British Psychological Society, Division of Clinical Psychology, Leicester
Campbell J (2001) Rationality, meaning, and the analysis of delusion. Philos Psychiatr Psychol 8:89–100
Coltheart M, Davies M (eds) (2000) Pathologies of belief. Blackwell Publishers, Oxford
Currie G (2000) Imagination, delusion and hallucinations. Mind & Language 15:168–183
Dancy J (1993) Moral reasons. Blackwell, Oxford
Davies M, Coltheart M, Langdon R, Breen N (2001) Monothematic delusions: towards a two-factor account. Philos Psychiatr Psychol 8:133–158
Davies M, Egan A (2013) Delusion: cognitive approaches – Bayesian influence and compartmentalization. In: Fulford KWM, Davies M, Gipps R, Graham G, Sadler J, Stanghellini G, Thornton T (eds) The Oxford handbook of philosophy and psychiatry. Oxford University Press, Oxford
Egan A (2008) Imagination, delusion, and self-deception. In: Bayne T, Fernandez J (eds) Delusion and self-deception: affective and motivational influences on belief formation, Macquarie monographs in cognitive science. Psychology Press, New York
Eilan N (2000) On understanding schizophrenia. In: Zahavi D (ed) Exploring the self. John Benjamins, Amsterdam, pp 97–113
Fulford KWM (1989) Moral theory and medical practice. Cambridge University Press, Cambridge (reprinted 1995 and 1999)
Fulford KWM (1991) Evaluative delusions: their significance for philosophy and psychiatry. Br J Psychiatr 159:108–112 (Supplement 14, Delusions and awareness of reality)
Fulford KWM, Radoilska L (2012) Three challenges from delusion for theories of autonomy. Chapter 3. In: Radoilska L (ed) Autonomy and mental disorder. Oxford University Press, Oxford
Fuchs T (2015) The intersubjectivity of delusions. World Psychiatry 14(2):178–179
Gallagher S (2009) Delusional realities. In: Bortolotti L, Broome M (eds) Psychiatry as cognitive neuroscience. Oxford University Press, Oxford, pp 245–266
Garety P (2015) Therapeutic advances for people with delusions will come from greater specification and empirical investigation. World Psychiatry 14(2):180–181
Garety PA, Freeman D (2013) The past and future of delusions research: from the inexplicable to the treatable. Br J Psychiatry 203:327–333
Gipps RGT, Fulford KWM (2004) Understanding the clinical concept of delusion: from an estranged to an engaged epistemology. Int Rev Psychiatry 16(3):225–235
Harré R, Lamb D (eds) (1987) Dictionary of philosophy and psychology. Blackwells, Oxford
Heinimaa M (2003) Incomprehensibility. Chapter 14. In: Fulford KWM, Morris KJ, Sadler JZ, Stanghellini G (eds) Nature and narrative: an introduction to the new philosophy of psychiatry. Oxford University Press, Oxford
Jackson MC (1997) Benign schizotypy? The case of spiritual experience. In: Claridge GS (ed) Schizotypy: relations to illness and health. Oxford University Press, Oxford, pp 227–250
Jackson M, Fulford KWM (1997) Spiritual experience and psychopathology. Philos Psychiatr Psychol 4(1):41–66
Jaspers K (1913) Allgemeine psychopathologie. Springer-Verlag, Berlin. (trans: Hoenig J), Hamilton MW (1963) General psychopathology. University of Chicago Press, Chicago. New edition (two volumes, paperback), with a foreword by Paul R McHugh (1997) The Johns Hopkins University Press, Baltimore
James W (1902) The varieties of religious experience. Longmans, New York
Johns LC, van Os J (2001) The continuity of psychotic experiences in the general population. Clin Psychol Rev 21(8):1125–1141

Koralus P, Mascarenhas M (2013) The erotetic theory of reasoning: bridges between formal semantics and the psychology of deductive inference. Philos Perspect 27(1):312–365, 10.1111/phpe.12029/abstract
Maher BA (1999) Anomalous experience in everyday life: its significance for psychopathology. Monist 82:547–570
McDowell J (1994) Mind and World. Harvard University Press, Cambridge, MA
McDowell J (1996) Two sorts of naturalism. In: Hursthouse R, Lawrence G, Quinn W (eds) Virtues and reasons: Philippa Foot and moral theory, essays in Honour of Philippa Foot. Clarendon, Oxford, pp 149–179
Mishara AL, Sterzer P (2015) Phenomenology *is* Bayesian in its application to delusions. World Psychiatry 14(2):185–186
National Institute for Mental Health in England (NIMHE) and the Care Services Improvement Partnership (2008) 3 keys to a shared approach in mental health assessment. Department of Health, London
Mellor CS (1970) First rank symptoms of schizophrenia. Br J Psychiatry 117:15–23
Parrott M, Koralus P (2015) The erotetic theory of delusional thinking. J Cogn Neuropsychiatry 20 (5):398–415
Quinton A (1985) Madness, Chap. 2. In: Griffiths AP (ed) Philosophy and practice. Cambridge University Press, Cambridge
Ratcliffe M (2009) Understanding existential changes in psychiatric illness: the indispensability of phenomenology. In: Bortolotti L, Broome M (eds) Psychiatry as cognitive neuroscience. Oxford University Press, Oxford, pp 223–244
Ratcliffe M (2015) The interpersonal world of psychosis. World Psychiatry 14(2):176–178
Robinson D (1996) Wild beasts and idle humours. Harvard University Press, Cambridge, MA
Sass LA (1994) The paradoxes of delusion. Cornell, New York
Sass L, Byrom G (2015) Phenomenological and neurocognitive perspectives on delusions: a critical overview. World Psychiatry 14(2):164–173
Sensky T, Hughes T, Hirsch S (1991) Compulsory psychiatric treatment in the community, Part 1. A controlled study of compulsory community treatment with extended leave under the mental health act: special characteristics of patients treated and impact of treatment. Br J Psychiatry 158:792
Shepherd M (1961) Morbid jealousy: some clinical and social aspects of a psychiatric syndrome. J Mental Sci 107:687–704
Sims A (1988) Symptoms in the mind: an introduction to descriptive psychopathology. Bailliere Tindall, London
Stephens GL, Graham G (2000) When self-consciousness breaks: alien voices and inserted thoughts. The MIT Press, Cambridge, MA/London
Stephens GL, Graham G (2006) The delusional stance. In: Cheung Chung M, Fulford KWM, Graham G (eds) Reconceiving Schizophrenia. Oxford University Press, Oxford
West DJ, Walk A (1977) Daniel McNaughton: his trial and the aftermath. Gaskell Books, London
Wing JK, Cooper JE, Sartorius N (1974) Measurement and classification of psychiatric symptoms. Cambridge University Press, Cambridge, UK
World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

Impairments of Personal Freedom in Mental Disorders

36

Jann E. Schlimme

Contents

Abstract

The exercise of personal freedom involves decision-making capacity and behavior-actualization ability, both of which are subject to "inner" restrictions due to mental illness. In addition to changes in experience which may result, i.e., in preference reversal and hence unauthentic behavior, mental illnesses can also change one's habituality, putting certain habits out of play or trigger acquisition of new habits.

Furthermore, the exercise of personal freedom involves negative freedom, since "outer" restrictions can further impair one's decision-making capacity and behavior-actualization ability (i.e., inadequate information, stereotyping by relevant others, or non-barrier-free facilities). It is important to note that "outer" restrictions can become (the basis of) "inner" restrictions, especially by inducing certain habitualities or worldviews. On the other hand, "inner" restrictions caused by mental illness can be rendered less relevant for authenticity and personal freedom if counterbalanced by adequate support and/or circumstances. Therefore,

J.E. Schlimme (✉)
Department of Psychiatry and Psychotherapy, Charité University Medicine, Berlin, Germany
e-mail: schlimme.jann@gmx.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_24

"outer" restrictions can double the infringement of personal freedom already caused by "inner" restrictions due to mental illness. This calls for the empowerment of mentally ill persons and refers to the core of personal freedom, implying that no person can lose personal freedom completely simply because she is the active agent in her lifeworld.

Introduction

Personal freedom is taken as a precondition of a good life in the modern era. Mental disorders are considered to disrupt one's freedom, because they can impair a person's ability to make rational decisions and can lead to behavior which seems deeply out of character for that person. While in the early modern era this inability was deemed to be a rather general and lasting quality of insane (Mentally ill) persons, it emerged that such impairments are rather specific and usually correlate with episodes of acute illness.

During the last two decades, the debate surrounding autonomy and mental disorder attracted new interest in the wake of receding paternalism and an emerging recognition of the importance of informed consent in mental health care. Dehospitalization and development of outpatient and community mental health services (i.e., assisted housing, assisted employment) in the psychiatric field enhanced the negative liberty of persons with mental disorders in many (western) countries and fueled the necessity to improve insights into the connections between impaired autonomy and mental disorder. Nonetheless, future research needs to determine the specific, social-, and symptom-related impairments of individual freedom in persons with mental illnesses even further.

This debate draws on an understanding of autonomy which is informed by the decision-making capacity approach. Many authors argue for an enlargement of this understanding due to the role of values in human life. Some authors promote an understanding of a good life which is related to concepts of authenticity, while others argue for an operationalization of sub-capacities in order to achieve reliable and valid tests for easy clinical application (i.e., appreciation of information ability in depressed persons, Hindmarch et al. 2013). Other authors promote a concept of decision capacity detached from any conception of freedom of the will.

Most authors agree that the capacity for personal autonomy is independent of external restrictions on autonomy, such as inadequate information, stereotyping, or non-barrier-free facilities (i.e., the term "capacity" refers to an "ability of subjects," Charland 2011). However, in a given situation, such conditions (adequate information, absence of stereotyping, barrier-free facilities) must be fulfilled in order to adequately bring one's capacities into play. Nonetheless, inner restrictions are the focus of the debate on impairments of personal freedom in mental disorders such as addiction, depression, or psychosis. Hence, the entry will have the following structure:

- Concepts of autonomy in the recent debate on impaired freedom
  1. Impairments of personal freedom in addictive disorders
  2. Impairments of personal freedom in depressive disorders
  3. Impairments of personal freedom in psychotic disorders
  4. Outlook

Concepts of Autonomy in the Recent Debate on Impaired Freedom

Despite its importance, autonomy is not an unambiguous concept, yet most philosophers would agree that we should consider ourselves as autonomous if capable of giving good reasons for our behavior. In doing so, we manifest our values and (moral) guidelines into action and behave according to our fundamental moral principles. This concept of autonomy undergirds the ideal of an informed consent, in which the medical doctor follows four moral principles in his day-to-day practice, namely, respect for autonomy, non-maleficence, beneficence, and justice (Beauchamp and Childress 2011). These four moral principles are supposed to be applicable in any given situation regardless of the worldview (*Weltanschauung*) of the persons involved (so-called principlism). Hence, moral conflicts are understood as conflicts regarding the ranking of these four principles.

Critical debates point out the importance of understanding why a certain person selects a certain ranking of these principles – or other values – in a situation in which she is called upon to decide (Charland 2002; Wiggins and Allen 2011). One cannot simply infer that she ranked them autonomously or was not influenced in her ranking by her mental illness. Nonetheless, the principlist approach is considered to be the gold standard in applied ethics such as medical or clinical ethics (Appelbaum and Grisso 1995; Appelbaum 2007). This has influenced the understanding of autonomy which is drawn upon in recent debates regarding impairments of personal autonomy in mental disorders.

In these debates, personal autonomy is conceptualized as a capacity of a self-reflecting agent who is a moral agent. But no matter how the foundations of what it is to be a morally responsible agent are conceptualized, it has to be admitted that the reasons for an agent's behavior must be grounded in the mental structure this agent had in the moment right before he initiated his behavior and/or decided upon a specific behavior. This groundedness of decisions and actions in a person's mental life is the basis for the influence of mental disorders on one's personal autonomy. In extremely debilitating conditions, such as severe dementia, this influence can also impair a person's autonomy in an existential sense, disabling the person to explicitly say "yes" or "no" to his or her own existence as a living being in its own lifeworld.

This does not necessarily imply that one's ability to kill oneself is the highest expression of freedom, as existentialist positions often claim (Camus 1942), but basically supports arguments highlighting the role of values in decision-making and of authenticity in personal autonomy.

From a philosophical point of view, the influence of our deepest desires, such as those for other people or things we love or long-term life goals, is of special interest. If we cannot help but care for someone, this caring prescribes all our other implicit and explicit valuings, judgments, and decisions (Frankfurt 1999). From a phenomenological point of view, such passive qualities of moral agency can even refer to the manner in which our lifeworld is disclosed to us, calling upon us not only to take up responsibility for our deepest desires but also for the manner in which we disclose our lifeworld to us (Drummond 2010). However, positions which describe certain behavior as heteronomous due to its merely being "immoral" are contradicting the consensus within the relevant discourse. In other words, autonomy entails freedom to undertake immoral behavior.

According to the informed consent movement in (mental) health care – mirroring the necessity for persons to consent to their treatment on the basis of extended information which is in line with most contemporary approaches to moral agency – personal autonomy is related to the capacity of rational decision-making. Extended (empirical) studies regarding competence to make treatment decisions distinguished four sub-capacities, or abilities/skills, considered necessary for (autonomous) decision-making:

1. To express a choice
2. To understand relevant information
3. To appreciate one's situation and its consequences
4. To reason about treatment options/rationally manipulate information (Appelbaum and Grisso 1995)

These four cognitive skills have been used to develop a widely applied assessment tool, the MacCAT (MacArthur Competence Assessment Tool), offering at least some consensus on how to view and assess a patient's competence in clinical day-to-day practice (Appelbaum and Grisso 1995; Charland 2011; Meynen 2011; Owen et al. 2009; Vollmann 2008). Mental competence in this sense is, however, necessarily only valid with respect to a concrete decision in this very space and time (Buchanan and Brock 1989, pp. 18–20). It is furthermore unclear whether a person can only be deemed to make autonomous decisions if she can display all four sub-capacities in a given situation, even if the decision at hand is a very simple one, such as buying a cup of coffee. This so-called threshold quality of the capacity concept (Buchanan and Brock 1989) challenges concepts of the freedom of the will, since the latter cannot be given in "degrees." There is an open debate on whether or not a decision-making capacity can be conceptualized without referring to the concept of a freedom of the will (Meynen 2011).

On the one hand, decision-making capacity is claimed to substitute the concept of freedom of the will; on the other hand, it is argued that a meaningful understanding of personal autonomy requires acceptance of the "givenness" of freedom of the will, at least in the sense that it serves as a regulative idea. Furthermore, the role of values and valuing – both prima facie pre-reflective and reflective value apprehensions – remains unaddressed in this approach (Charland 2011; Schlimme 2012;

Tan et al. 2007). This point is of special relevance for psychiatry due to the importance of values in mental life (Fulford 2004). In combination with the problematic (or absent) role of values in the capacity concept, these debates call "the empirical validity of the concept of capacity embodied in a given test" into question ("dual nature of competence," Charland 2011). Nonetheless, the capacity concept, which originally started within a specific area of ethically complex treatment decisions, has thus developed into a model for autonomous daily life.

To summarize, the capacity approach depicts the moral agent as deciding rationally and meaningfully for itself in a first step. According to this idea of a moral agent, they afterward actualize their decisions in their life in a second step. According to this conceptualization, restrictions of personal autonomy might therefore arise:

(a) On the level of one's decision-making capacity
(b) On the level of one's behavior-actualization ability

Both capacities could be conceptualized as independent from certain qualities of the relevant person's environment in the sense of possible outer restrictions such as inadequate information, attribution of stereotypes by relevant others, or non-barrier-free facilities. However, in a given situation, such conditions (adequate information, social recognition, barrier-free facilities) must be fulfilled in order to adequately bring these capacities into play.

Nonetheless, with respect to mental disorders such as addiction, depression, or psychosis, inner restrictions are the main focus of interest. "Inner" restrictions are impairments which disable the person to use the granted (negative) freedom autonomously – that is, to display both her decision and behavior-actualization ability. Furthermore, debates on personal autonomy in mental disorders with a philosophical approach primarily focus on decision-making capacity, that is, on those conditions of mental disorders which specifically impair a person's decision-making capacity. Behavior-actualization abilities and outer restrictions might be of equal importance for the relevant person in the situation in which she is called upon to act but are usually addressed in a different discourse (i.e., psychiatric or therapeutic discourses on recovery, social psychiatry, or empowerment). Positions connecting both discourses often focus on the role of values and authenticity in personal freedom (Flanagan 2011; Schlimme 2012).

Impairments of Personal Freedom in Addictive Disorders

"The addict" is a classical example of a person with impaired personal freedom. She may intend, and voluntarily consent, to abstain from her drug of choice, articulating excellent reasons to do so, yet suffer relapse an hour later in a stressful or tempting situation. From a commonsense point of view, "the addict" displays a weakness of the will. In other words, her impaired behavior-actualization ability ("guidance control," self-control) corresponds with an impaired decision-making capacity.

"Addicts" typically display an impairment to actualize the behavior they have decided to perform (abstaining from taking drugs) in seducing situations. To put it differently, they encounter difficulties in adequately taking into account the strength of their compulsive habit and the momentary change of their goals and values toward a short-sighted and drug-oriented set when determining their choice.

The latter points out specific impairments regarding both her skill to appreciate her situation and its consequences and the ability to reason about behavioral options and rationally manipulate (relevant) information. From this point of view, the addict is indeed addicted to her drug of (involuntary) "choice" and impaired regarding all decisions (and behaviors) related to that drug. This addiction correlates with a specific impairment in her decision-making capacity (skills 3 + 4) and implies a specific impairment in her behavior-actualization ability (resembling a weakness of the will or impaired guidance control), leading to further drug consumption and so forth. All in all, this displays a specific, drug-related impairment of individual freedom.

This rather broad description is usually agreed upon both from medical and psychological as well as philosophical, ethical, and legal points of view. It is furthermore agreed that these impairments fluctuate in relation to the addictive cycle of intoxication and withdrawal. Persons with addictive disorders are hence, at least in those cultures that accept the concept of addiction as a disorder, usually deemed to display, or suffer, specific drug-related impairments of personal freedom. There is an ongoing debate on how to understand and conceptualize these impairments and regarding what kind of responsibility (legal, moral, clinical) is altered (i.e., whether they might even expel addicts from responsibility regarding drug-related behavior whatsoever) (Yaffe 2001; Poland and Graham 2011). There are three favored, not mutually exclusive, lines of argumentation on why persons with addictive disorders are specifically and fluctuatingly impaired in their decision-making capacity skills 3 and 4, resulting in an impaired actualization ability for these persons:

1. Directly drug-related impairments of attention, memory, or executive functions (i.e., intoxication, withdrawal)
2. An altered style of reasoning, developed and acquired during the process of developing and acquiring the drug habit (i.e., hyperbolic discounting to prefer short-term and neglect long-term effects of one's behavior, resulting in a preference reversal of goals: short-term goals, not corresponding with one's long-term interests, are preferred compared to long-term goals, corresponding with one's long-term interests; Elster 2000)
3. An altered set of values, developed and acquired together with the drug habit (i.e., overestimation of drug-related benefits or one's self-control, sometimes conceptualized as a specific manner of self-deception/irrational beliefs/irrational self-image; Charland 2002; Schlimme 2010)

Even a former drug addict who abstains from drugs for longer periods of time is often unable to use the drug of choice in a controlled and non-compulsive manner,

which correlates to his acquired and developed addictive habituality and extensive and lasting neurophysiological alterations acquired during his ongoing addictive behavior (Schlimme 2010; Schlimme and Voss 2017).

From a psychiatric point of view, underlying disorders (i.e., posttraumatic, personality, affective, or psychotic) might further fuel the addiction beside the acquired habituality due to positive effects of the drug of choice (so-called self-medication hypothesis; Khantzian 1985) as well as a "depraving" and demotivating environment (so-called rat park hypothesis; Alexander et al. 1978), which also minimizes negative freedom due to outer restrictions. Consequently, a person with an addictive disorder might not even try to discard the habit, if she successfully manages to maintain her social life. This is especially the case if her social life furthermore covers – or at least accepts – her habit, implying that she integrates her drug consumption as a meaningful and sustaining behavior into her life ("sober drunkard"; Schlimme 2010; Pickard 2012). This typically results in extensive "phases" or "moments of clarity," often also given if persons with (illegal) drug addiction are integrated in substitution settings, minimize consumption of other drugs, and develop and maintain an empowering social surrounding. From a philosophical point of view, it is debatable whether a manner of drug use allowing for such "phases of clarity" could be called responsible, at least responsible in the short term or within a specific context (i.e., a special setting or regarding one's coping with otherwise more severe symptoms/self-medication hypothesis).

This argument could be taken a little further, ending up in a right to be addicted claim opposing the war on drugs claim. The "right to be addicted" argument basically states that a person's addiction is her or his own problem as long as that person is sober enough to get along in life and not cause problems for others. The war on drugs argument claims that the addict is unable to decide against the drug he is addicted to, because he is addicted to it, and therefore, the drug as a disastrous agent should be banned (obsta principiis). Both arguments aim at the heart of the concept of addictive disorders.

If the birth of this concept around 1800 is reconsidered, one encounters an important distinction for developing this concept as well as all other modern concepts of mental disorders. This distinction separates well-being from the first-person perspective and having a disease from the medical perspective. It was Thomas Trotter (1760–1832) who demonstrated the often Janus-faced coexistence of personal well-being and constant alcohol consumption in the ideal type of a sober drunkard, a seemingly quite frequent manner of existence for British sailors and dock workers in Trotter's times (Trotter 1804).

Today this argument is mostly debated regarding the use of substitute drugs for illegal drug users. Some authors argue that substituted addicts might be unable to competently consent to substitution treatment because they are addicted (and hence intensively drawn) to the substituting agent (Charland 2002). Others argue that substituted addicts are typically highly competent and able to live their life authentically because they are substituted (Schlimme 2010).

Nevertheless, concrete impairments of decision-making capacity and behavior-actualization ability are not directly dependent on the addictive disorder but on the

effects of the acquired drug habit, on the level of drug intoxication or withdrawal, and on the stance and worldview the addict takes toward these. From this point of view, sobriety is more important than the simple fact of continuously using certain drugs (whether for self-medication, ritual, or recreational motives).

It seems to be relevant for philosophical (and ethical) judgment how addictive behavior is conceptualized, even if it is agreed that persons with addictive behavior have diminished responsibility at least some of the time. It is furthermore generally agreed that impairments of personal autonomy in addictive behavior cannot be conceptualized in a general manner. They can only be depicted and determined individually for which decision a person with addictive behavior (i.e., abstention or continuation of further drug use or the setting in which she uses her drug) can be held responsible. In this specific determination of responsibility, the actual stage of her addictive cycle (i.e., intoxication, clarity, withdrawal) has to be taken into account.

Impairments of Personal Freedom in Depressive Disorders

Depressive moods are experienced as being imposed. They are hence passively experienced phenomena, typically rendering the afflicted person unable to feel happy as well as diminishing the person's drive to pursue her own interests and thusly the person's interest to pursue her own happiness and autonomy. Prima facie "the depressed person" seems to be impaired in her personal freedom due to impaired behavior-actualization ability ("no drive") while seemingly having "normal" decision-making capacity. Yet this displays a rather poor understanding both of decision-making capacity and depressed moods.

A broader picture shows that "in depressed mental life everything is somehow at a devaluing loss ('The glass is half empty.')" (Schlimme 2013b). This negative or depressive manner of selecting and affectively responding to experienced non-axiological properties (oneself, circumstances and objects in one's lifeworld, one's life history, and future prospects) typically implies a devaluing, or negativistic, experience of oneself (low self-esteem) and one's (sensible) needs (hence serving primarily the needs of others), one's deeds in the past (feeling guilty, "If only I had. . ."), and one's behavioral options in the given situation and the future (anhedonia, helplessness, hopelessness) (Hindmarch et al. 2013; Meynen 2011; Rudnick 2002; Schlimme 2013b; Sullivan and Younger 1994).

Hence, depressed moods can correlate with specific impairments in decision-making capacity, namely, to appreciate one's situation and its consequences (skill 3), entailing possible preference reversal due to both a negativistic selection and devaluing of behavioral options or personal needs and aims.

This rather broad description accepts an intricate connection between mood and personal freedom in the sense of a "preintentional quality" of moods (see for this argument Rudnick 2002; Slaby and Stephan 2008; Ratcliffe 2010). Basically, this intricate connection argues that one's mental life is pre-reflectively prescribed by one's depressed mood and that the complete structure of lived experience is altered by the depression. This is a classical argument already given in the *Corpus*

Hippocraticum, a collection of medical papers from the fourth century B.C. in Alexandria. Here *melancholia* is deemed responsible for inducing delusions and a "certain desire to long for death as if it would be something good" (Hippokrates 1933–1940, V, XXIII/136ff).

It is furthermore a standard argument in the (clinically driven) psychiatric discourse, drawing on Karl Jaspers' distinction between mood (*Stimmung*) and affect/feeling (*Affekt*/*Gefühl*) in his influential *General Psychopathology*. Jaspers redefined the term "mood" as a complex state of feelings providing the background and color (*Färbung*) of actual mental life (Jaspers 1913, 62f). Importantly, impairments of personal freedom can be connected with depressive moods in two ways. On the one hand, the perceived or experienced (behavioral) options can be reduced in *scope*; on the other hand, the experienced options can be valued in a different ("negativistic," "devaluing") *style* (Schlimme 2013b).

The first connection impairs one's personal freedom via a reduced network of possibilities (Meynen 2011), which could be termed "inner restrictions of negative freedom" (correlating to skill 2 + 3). The second connection seems to result in impairments via a minimized concern for one's own welfare (Elliott 1997; Rudnick 2002), which could be termed "inner restrictions of positive freedom" (correlating to skill 3 + 4). Moreover, severe depression often implies lack of drive resulting in an impaired behavior-actualization ability.

Obviously, personal freedom can be impaired in depressive moods both in scope and style, resulting in the ambiguous intersubjective experience of persons with depressive moods mentioned above. Surveys on competence to consent to medical/psychiatric treatment demonstrate that moderately depressed persons are usually still capable of giving informed consent, even while already being impaired in personal freedom due to being depressed. On the other hand, severely depressed persons often lack the capability to give informed consent to treatment (Hindmarch et al. 2013; Lee and Gazini 1994).

This complex picture can also be found in persons with ongoing depressive disorders (dysthymia, double depression), who might achieve helpful and meaningful coping styles for their ongoing or recurring depressive symptoms. The latter typically entails a more complete personal freedom, while still being impaired in some ways, possibly resulting in a unique and mostly recovered way of living (Schlimme 2012). This does not, however, merely require a mental process in which one modifies one's demands and thereby achieves an altered worldview (i.e., fueled by psychotherapy). It also requires more negative freedom to arrange one's everyday life on a lesser activity level, including the social and economical dimension.

The extreme side of this connection might be found in modern societies, typically pursuing high aspirations for their members (i.e., regarding autonomy and authenticity). Postmodern societies even publicly support the claim of an "anything is possible" and thus the possibility of an authentic life. Naturally, this claim is just an illusion, since a person's negative and positive freedom is highly dependent on her social environment, individual life history, and resources. Similarly, authenticity cannot permanently be achieved or even owned like an object but is a demand in

itself. If persons identify themselves with and try to live up to their social environments' high demands, they are more easily overtaxed with simply trying to be themselves authentically (Ehrenberg 1998).

This might result in a fatigued self (Ehrenberg) with which comes a higher risk for depressive disorders. This argument claims that negative freedom might be impaired due to inner restrictions, since social demands are habitualized and adopted during socialization (whether during childhood or later on). From this perspective, depressive moods appear to mirror a missing internal negative freedom in the face of one's own demands.

This complexity fuels philosophical debates on suicide as well, since behavioral options are also typically altered in scope and style in suicidal mental states due to an "affective narrowing" (*Einengung*, hopelessness). This "affective narrowing" (*Einengung*, hopelessness) is usually given before committing or attempting suicide (being one of three crucial features of the so-called pre-suicidal syndrome, Ringel 1954; Beck 1987). In acute suicidal conditions, the behavioral options can indeed be effectively narrowed down to two choices: "staying alive" or "killing oneself," while one's prima facie valuing oscillates between valuing the given situation and mental life as "unbearable" or "just bearable" and the option of killing oneself as "last and only rescue/exit/escape" or "no exit at all," with the tendency to overestimate negative outcomes (see Schlimme 2013a).

Basically, these different concepts developed with respect to depressive moods and disorders can be transposed to manic moods and disorders, at least to a certain extent. In manic moods, an altered scope of one's experienced (behavioral) options and possibilities as well as an altered style of positivistic prima facie − +9 result in an overvaluing of one's abilities (inflated self-esteem) and needs (compared to the needs of others, correlating with a reckless behavior), one's deeds in the past, and one's behavioral abilities and options in the given situation and the future (euphoria, omnipotence).

Although, prima facie "the manic person" seems to be impaired in her personal freedom due to a massively inflated behavior-actualization ability ("too much drive") while seemingly having "normal" decision-making capacity, a closer look reveals that, in particular, the appreciation of one's situation and reasonable digestion of information in the light of one's values (or moral principles) is impaired. This impairment is due to the manic person's overly optimistic (sometimes even incorrigible/delusional) pre-reflective valuation of her individual abilities, sometimes even experiencing herself as being endowed with supernatural abilities (i.e., delusions of grandeur). In the light of her imagined abilities, a person values, reasons, and judges other things and circumstances as relevant for her decisions and behaviors than the same person would in a more sober mood (preference reversal).

Impairments of personal freedom in depressive and manic disorders draw attention to the correlation between mood and personal freedom, the intricate connection between social and personal demands and habitualization of these demands, as well as the influence of mood and demands on external as well as internal negative freedom. Internal negative freedom toward habitualized (self-)demands are of particular interest here. In order to achieve greater freedom, it might be necessary, both

during critical illness episodes and alongside ongoing depressive symptoms, to grant the depressed person a greater negative freedom from outer restrictions or social demands (i.e., as is granted in the social role of a patient or a chronically disabled person).

It is indeed relevant for understanding personal freedom to adequately conceptualize these intricate and complex connections. However, impairments of personal autonomy in depressive (and manic) disorders need to be depicted and determined specifically and individually, taking into account especially the possibility of preference reversal during the course of the mental illness (Lee and Gazini 1994; Rudnick 2002). In this regard, possible impairments of personal freedom in persons with a depressed (or manic) state of mind are, besides possible impairments of one's behavior-actualization ability (i.e., lack of drive, depressive stupor), usually conceptualized as an impaired decision-making capacity mediated via one's automatic (pre-reflective, subliminal) preintentional (implicit) and biased choice as well as explicit over- or underestimation of prospects and outcomes of behavioral options or personal needs and preferences.

Impairments of Personal Freedom in Psychotic Disorders

"The psychotic person" is the classical example of a person with impaired personal freedom: she is supposedly unable to develop a meaningful and reasonable intention and at the same time unable to coherently pursue her perhaps perceivedly peculiar and weird interests in the given situation. This picture of the unreasonable and unpredictable "lunatic" still fuels the stigma persons with psychotic disorders face in their communities. It is, however, outdated if taking a closer look at the discourse on impairments of personal freedom in persons with psychotic disorders.

Nonetheless, in florid psychotic states, the affected person may indeed act on delusions (i.e., delusional mood, delusional hallucinations, delusional convictions) possibly entailing threatening behavior or may be unable to adequately appreciate and digest information (skills 3 + 4). This might even lead to the inability to understand relevant information (skill 2).

Surveys with psychiatric inpatients demonstrate that poor appreciation as well as poor reasoning explain apparently poor decision-making capacity (skills 3 + 4) (i.e., due to magical thinking, formal thought disorders, delusions; Vollmann 2008, 114 F; Owen et al. 2009). In non-acute phases, decision-making capacity might be impaired to a lesser degree or fully given, and behavior-actualization abilities (i.e., lack of drive, lack of emotional engagement) might be more important regarding impairments of personal freedom.

This rather broad description claims the possibility of a clear distinction between inner and outer restrictions regarding personal autonomy in persons with psychotic disorders. It stresses the passive quality of psychotic experiences, which overwhelm the pertinent person to such an extent that she values herself in retrospect as "different" or "not herself" (in the sense of not authentic) (Bolton and Banner 2012, p 96; Moller and Zauszniewski 2011; Noiseux et al. 2010; Schlimme and

Brückner 2017). However, during psychotic experiences, it is usually not the person herself but the world that is experienced as changed and altered in the first place. Consequently, impairments of personal freedom often become overt in interpersonal conflicts fueled by an inability to adequately adopt a commonsensical point of view.

While the psychotic person might experience her behavior as meaningful and justified, other persons judge her behavior as unjustified and threatening. From a psychiatric point of view, such situations are often fueled by delusions, these delusions being the most important source of interpersonal conflicts in persons with psychotic disorders (Golenkov et al. 2011; especially delusions of grandeur, Ullrich et al. 2013). Delusions, occurring during most (75 %) psychotic experiences diagnosed as schizophrenic, often wax and wane during the course of the illness (Appelbaum 2007; Jorgensen 1994; Schlimme and Brückner 2015). The role of delusions regarding impairments of personal freedom is, due to the very individual nature of those delusions, very difficult to define and should be the object of further research (Schlimme 2013c).

While delusions often dominate the delusional person's experience during acute psychosis, they are typically "parked" as actually unrequired experience and interpretation at other times. They can actually become "integrated" into that person's lifeworld and worldview as a private, unshareable parallel reality in long-lasting psychosis (pseudo-solipsism: Sass 1994; also: Bock 1997; Schlimme 2013c; Schlimme and Brückner 2015 a. 2017). In both ways they do not – at least not in principle – restrict the person's potential to respect the rights and worldviews of those (potentially) afflicted by her behavior, even if her behavior is driven or informed by delusional experiences or convictions (Schlimme 2013c). Delusions that manifest themselves in this manner need not result in or correlate with unreasonable and irresponsible behavior.

Responsibility for one's behavior while having delusions seems to be easier to assume if the person:

(a) Is able to communicate her highly private (psychotic) experiences (i.e., in certain self-help groups, trialogue, psychotherapy), in spite of her ongoing psychosis
(b) Can maintain a robust amount of social integration, enabling her to adopt the commonsensical point of view more easily as a justified and (at least) parallel worldview in the given situation (islands of clarity, Podvoll 2003; Schlimme and Brückner 2017)

Furthermore, negative symptoms with a depression-like character might further impair one's personal freedom. Nonetheless, insight into the delusional (private, unshareable, unprovable) character of one's psychotic experiences correlates more often with better social integration but is not a necessary requirement (i.e., 40 % of persons with schizophrenia displaying full functional recovery did not show this kind of insight: Alvarez-Jimenez et al. 2012; see Bottlender and Hloucal 2010; Nixon et al. 2010). However, recovery with long-lasting psychotic experiences (and ongoing delusions) is often connected with an impaired personal autonomy

due to impaired stress tolerance, stamina, and cognitive functions, which implies impaired decision-making capacity (skills 2 + 3) and behavior-actualization ability after participating in strenuous situations for a prolonged period of time. These impairments can be further fueled by (rapidly) emerging self-disorders in such situations, overall implying the necessity to adequately "dose" these strenuous situations.

These highly diverse manners of impaired personal freedom in psychotic disorders have their common ground in a loss of "normality," more precisely in a loss of the taken-for-grantedness of one's "normality." This loss might be more overt in acute phases (i.e., delusions, hallucinations) but is often also given in non-acute phases during recovery (i.e., perplexity, hyperreflectivity). The concept of a "loss of taken-for-grantedness of one's 'normality'" (loss of commonsensical habituality) refers to the fact that we usually, that is, automatically and reliably, present our lifeworld in a more or less homelike manner to ourselves (Blankenburg 1971).

Moreover, the current "project" the person is involved in (i.e., buying a cup of coffee) automatically informs the manner of how we disclose our lifeworld to ourselves as an experiential workspace (i.e., automatically paring out irrelevant information in the given situation in which one is called upon to act) (Schlimme 2012; Schlimme and Brückner 2017; Schlimme and Voss 2017). It is this automaticity that is impaired in psychotic disorders. Hence, persons with psychotic disorders need to reflectively and actively select the relevant meanings even in the most common situations, while often being impaired in their cognitive functions as well, rapidly implying impaired decision-making capacity skills 2 + 3 (impaired ability to adequately appreciate and manipulate relevant (sic!) information).

Consequently, impairments of personal freedom in psychotic disorders call for a more intricate and complex conceptualization of "inner" and "outer" factors with regard to impairments of personal freedom. As persons with psychotic disorders and other long-lasting mental illnesses demonstrate, their personal freedom is not only dependent on personal abilities (i.e., decision-making capacity and behavior-actualization ability) but also highly dependent on the structure of the given situation (i.e., negative freedom, social support). Achieving normal goals is a paramount and demanding aim for persons with psychotic disorders. It often requires adequate assistance (i.e., assisted housing, assisted employment) (Davidson et al. 2009) and social support due to ongoing unusual behavior (Schlimme and Schwartz 2013). These insights call into question the adequacy of the abovementioned strict distinction between inner and outer restrictions of personal freedom.

Conclusion

Insights from appreciation of fine-grained understandings of mental life, as present in some specific disorders/mentally illnesses, challenge some normative qualities of autonomy concepts, especially regarding the self-image of profound independence both from one's situation (i.e., circumstances, social recognition, community,

empowerment) and from one's self (i.e., desires, habituality, life history). These insights highlight authenticity concepts and call for acknowledgment both within therapeutic discourses and discourses about diversity. They adopt Rudnick's critique of the standard notion of competence to consent (informed consent): "It seems that the four abilities noted above refer to the output (expression) and process (understanding, appreciation, and reasoning), but not to the input (information and preferences), of decision making. Input information is addressed within the broader doctrine of informed consent, but input preferences, which may be characterized as ends assumed by the individual, are largely ignored in this framework" (Rudnick 2002, p. 152).

Values and preferences as well as life history, in which these values and preferences are brought into play, are deemed to be those aspects most crucially missing in the current medico-ethical debate on personal freedom and mental illnesses (Wiggins and Allen 2011). This corresponds to the redefinition of many features of persons with long-lasting mental illnesses as handicaps which implies to move the relevant background of impairment from inner to outer restrictions, namely, to expect a more diverse normality in order to allow for a greater variety of lifestyles (including values, preferences, worldviews) which might be, to a relevant extent, fueled by mental illnesses (c.f. Convention on the Rights of Persons with Disabilities, United Nations 2006). Last but not least, the current debate rejects points of view claiming a given behavior as not autonomous simply because it is judged (from their stance, i.e., representing the commonsensical position of their culture) as immoral (i.e., drug consumption, suicide attempts) or abnormal (i.e., unusual behavior during long-lasting psychotic disorders).

Definitions of Key Terms

Authenticity	Behave and decide in accord with one's most cherished or cared for interests and preferences. Authenticity cannot permanently be achieved or even owned like an object but is a demand in itself for one's conduct of life.
Behavior-actualization ability	The ability to behave in accord with one's decisions. Being primarily a mental competence, in actual life it nonetheless depends on situational circumstances (i.e., social recognition, barrier-free facilities).
Decision-making capacity	The ability to decide in accord with one's (moral) values and principles. Being primarily a mental competence, in real life it nonetheless depends on situational circumstances (i.e., adequate information).
Freedom, negative	The situational or socially granted space to decide and especially behave the way one decides to do. It refers primarily to the concrete social situation in the sense of an "outer negative freedom" but could also be understood in the manner of an "inner negative freedom."

Freedom, personal	The concrete freedom a person can bring into play here and now in her lifeworld, in which she is called upon to act. On the one hand, it calls on decision-making capacity and behavior-actualization ability and is therefore potentially subject to "outer" and "inner" restrictions; on the other hand, it is an immeasurable component given of every human being as active agent in its lifeworld.
Habituality	The acquired, not easily unlearned or altered pre-reflective (subliminal, "passive") activity of the mind according to which our experience and the experienced is given in a manner we are already acquainted with. Habituality implies a complex set of pre-reflective anticipations concerning the manner of experiences and the experienced.
Lifeworld	The experienced situation and horizon a person is experiencing itself as being in as embodied self and in which it is called upon to act. Lifeworld is a key concept to every approach and concept drawing on the first-person perspective.
Mental disorder	A disorder affecting the mind and causing mental illness. Mental disorders are conceptualized as structural change of one's mind. The structural change can be located on the habitual (psychosocial), on somatic (neurophysiological), or on both (interacting) levels. The exact understanding of mental disorders varies over time and culture (i.e., mental disorders as natural entities versus mental disorders as useful guides to treatment).
Mental illness	To suffer from altered manners of experiences not open for intentional change (i.e., depressed mood, psychosis). These changes can be experienced as being given in the lifeworld (i.e., delusional hallucination), in oneself (i.e., loss of drive), or in one's experience itself (i.e., anxiety). Typically, these manners of experiences are evaluated as "unusual" or "altered" from the first-person perspective, but not necessarily as "ill" or symptom of a mental disorder.
Restrictions, inner	Impairments disabling the person to use the granted (negative) freedom autonomously and authentically and to display both decision-making capacity and behavior-actualization ability in the situation to act. Classical "inner" restrictions are mental alterations implying preference reversal (i.e., focus on short-term goals according

	to craving, different focus according to depressive (de) valuing or delusional convictions).
Restrictions, outer	Social or situational barriers for deciding or behaving the way one wants to. Classical "outer" restrictions impair one's negative freedom to behave the way one would like to do (i.e., inadequate information, givenness of stereotypes by relevant others, or non-barrier-free facilities). "Outer" restrictions can become (the basis of) "inner" restrictions especially by inducing habitualities or worldviews.
Values	Specific meanings of things and circumstances we find value in. The value of things and circumstances is principally distinguishable from these things and circumstances themselves. Therefore, values can be very concrete (i.e., the delicacy of food) or highly abstract (i.e., the concept of a delicacy of passions), immediately experienced (i.e., some food looks and tastes delicious), or reflectively addressed (i.e., a theater performance is judged as delicate in retrospect).
Worldview	The personal narrative concerning the given as a whole. It is a set of explicit and implicit meanings, interpretations, and values, ordering the experienced in relation to the whole. Every person has a worldview, more or less coherently corresponding to her experiences, life history, and prospects.

Summary Points

- Personal freedom can never be lost completely as long as the pertinent person is an active agent in her lifeworld.
- Personal freedom calls on decision-making capacity and behavior-actualization ability, both open for "inner" restrictions caused by mental illness, in order to live authentically.
- Personal freedom calls on negative freedom in order to live authentically, since "outer" restrictions can impair one's decision-making capacity and behavior-actualization ability (i.e., inadequate information, givenness of stereotypes by relevant others, or non-barrier-free facilities).
- "Outer" and "inner" restrictions are intricately connected: "inner" restrictions can be rendered less relevant if counterbalanced by adequate circumstances; "outer" restrictions can double the impairments caused by "inner" restrictions; and "outer" restrictions can become (the basis of) "inner" restrictions especially by inducing certain habitualities, demands, or worldviews in the long run.

- Mentally ill persons need to be empowered to grasp their personal freedom and live an authentic life.

References

Alexander BK, Coambs RB, Hadaway PF (1978) The effect of housing and gender on morphine self-administration in rats. Psychopharmacology 58:175–179

Alvarez-Jimenez M, Gleeson JF, Henry LP, Harrigan SM, Harris MG, Killackey E, Bendall S, Amminger GP, Yung AR, Herrman H, Jackson HJ, McGorry PD (2012) Road to full recovery: longitudinal relationship between symptomatic remission and psychosocial recovery in first-episode psychosis over 7,5 years. Psychol Med 42(3):595–606

Appelbaum PS (2007) Clinical practice. Assessment of patients' competence to consent to treatment. N Engl J Med 357:1834–1840

Appelbaum PS, Grisso T (1995) The MacArthur treatment competency study. I. Mental illness and competence to consent to treatment. Law Human Behav 19:105–126

Beauchamp T, Childress J (2011) Principles of biomedical ethics, 5th edn. Oxford University Press, Oxford

Beck AT (1987) Cognitive models of depression. J Cogn Psychother 1:5–37

Blankenburg W (1971) Der Verlust der natürlichen Selbstverständlichkeit. Ein Beitrag zur Psychopathologie symptomarmer Schizophrenien. Enke, Stuttgart

Bock T (1997) Lichtjahre. Psychosen ohne Psychiatrie. Psychiatrie-Verlag, Bonn

Bolton D, Banner N (2012) Does mental disorder involve loss of personal autonomy? In: Radoilska L (ed) Autonomy and mental disorder. Oxford University Press, Oxford, pp 77–99

Bottlender R, Hloucal TM (2010) Krankheitseinsicht. Eine Übersicht über das Konzept mit Fokus auf schizophrene Erkrankungen. Psychiatr Prax 37:59–67

Buchanan AE, Brock DW (1989) Deciding for others. The ethics of surrogate decision making. Cambridge University Press, Cambridge

Camus A (1942) Le mythe de Sisyphe. Gallimard, Paris

Charland LC (2002) Cynthia's dilemma. Consenting to heroin prescription. Am J Bioethics 2:37–47

Charland LC (2011) Decision-making capacity. Stanford encyclopedia of philosophy. Available via http://plato.stanford.edu/entries/decision-capacity. Accessed 3 Feb 2015

Davidson L, Rowe M, Tondora J, O'Connell MJ, Lawless MS (2009) A Practical Guide to Recovery-Oriented Practice. New York, Oxford University Press

Drummond JJ (2010) Self-responsibility and eudaimonia. In: Ierna C, Jacobs H, Mattens F (eds) Edmund Husserl 150 Years. Springer, Dordrecht, pp 411–430

Ehrenberg A (1998) La Fatigue d'être soi. Dépression et société. Odile Jacob, Paris

Elliott C (1997) Caring about risks. are severely depressed patients competent to consent to research. Arch Gen Psychiatry 54:113–116

Elster J (2000) Ulysses unbound. Cambridge University Press, Cambridge

Flanagan O (2011) What is it like to be an addict? In: Poland J, Graham G (eds) Addiction and responsibility. MIT Press, Cambridge, pp 269–292

Frankfurt HG (1999) Necessity, Volition, and Love. Cambridge University Press, Cambridge

Fulford KWM (2004) Ten principles of values-based medicine (VBM). In: Schramme T, Thome J (eds) Philosophy and psychiatry. De Gruyter, Berlin, pp 50–80

Golenkov A, Large M, Nielssen O, Tsymbalova A (2011) Characteristics of homicide offenders with schizophrenia from the Russian federation. Schizophr Res 133:232–237

Hindmarch T, Hotopf M, Owen GS (2013) Depression and decision-making capacity for treatment or research: a systematic review. BMC Med Ethics 14:54

Hippokrates (1933–1940) Die Werke des Hippokrates. Die hippokratische Schriftensammlung in neuer deutscher Übersetzung. In: R Kapferer (Ed and Trans). Hippokrates Verlag, Stuttgart

Jaspers K (1913) Allgemeine psychopathologie, 1st edn. Springer, Heidelberg

Jorgensen P (1994) Course and outcome in delusional beliefs. Psychopathology 27:89–99
Khantzian EJ (1985) The self-medication hypothesis of addictive disorders. Focus on heroin and cocaine dependence. Am J Psychiatr 142:1259–1264
Lee M, Gazini L (1994) The effect of recovery from depression on preferences for life-sustaining therapy in older patients. J Gerontol 49:M15–21
Meynen G (2011) Depression, possibilities, and competence. A phenomenological perspective. Theor Med Bioethics 32:181–193
Moller MD, Zauszniewski JA (2011) Psychophenomenology of the post-psychotic adjustment process. Archiv Psychiatr Nursing 25:253–268
Nixon G, Hagen B, Peters T (2010) Recovery from psychosis. A phenomenological Inquiry. Int J Ment Health Addict 8:620–635
Noiseux S, Tribble St-Cyr D, Corin E, St-Hilaire P-L, Morissette R, Leclerc C, Fleury D, Vigneault L, Gagnier F (2010) The process of recovery of people with metnal illness. BMC Health Services Research 10:161
Owen GS, David AS, Richardson G, Szmukler G, Hayward P, Hotopf M (2009) Mental capacity, diagnosis and insight in psychiatric in-patients. A cross-sectional study. Psychol Med 39:1389–1398
Pickard H (2012) The purpose of chronic addiction. AJOB Neurosci 3:40–49
Podvoll EM (2003) Recovering sanity. A compassionate approach to understanding and treating psychosis. Shambala, Boston and London
Poland J, Graham G (2011) Introduction. The makings of a responsible addict. In: Poland J, Graha G (eds) Addiction and responsibility. MIT Cambrige, Cambridge, pp 1–19
Ratcliffe M (2010) Depression, guilt and emotional depth. Inquiry 53:602–626
Ringel E (1954) Der Selbstmord. Abschluss einer krankhaften psychischen Entwicklung. Dietmar Klotz, Eschborn, Frankfurt/M
Rudnick A (2002) Depression and competence to refuse psychiatric treatment. J Med Ethics 28:151–155
Sass L (1994) The paradoxes of delusions. Cornell University Press, Ithaca
Schlimme JE (2010) Addiction and self-determination. A phenomenological approach. Theor Med Bioethics 31:49–62
Schlimme JE (2012) Lived autonomy and chronic mental illness. A phenomenological approach. Theor Med Bioethics 33:387–404
Schlimme JE (2013a) Sense of self-determination and the suicidal experience. A phenomenological approach. Med Health Care Philos 16:211–223
Schlimme JE (2013b) Depressive habit and altered valuings. A phenomenology of depressed mental life. J Phenom Psychol 44:92–118
Schlimme JE (2013c) Is acting on delusions autonomous? Philos Ethics Humanities Med 14:14
Schlimme JE, Brückner B (2015) Entaktualisierung und Orthostrophe. Zur phänomenologischen Psychopathologie des abklingenden Wahns Nervenarzt 86:872–883
Schlimme JE, Brückner B (2017) Die abklingende Psychose. Ein lebensweltliches Modell der Genesung. Unter Mitarbeit von Hans O, Hase B, Önal L, Palmer A, Peach P. Psychiatrie-Verlag, Bonn (in press)
Schlimme JE, Schwartz MA (2013) In recovery from schizophrenia. Regaining social cover. A phenomenological investigation. Psychopathology 46:102–110
Schlimme JE, Voss M (2017) Agency, accustomed behaviour, and autonomy in depression and addiction. In: Mishra AL, Corlett P, Fletcher P, Schwartz MA (eds) Phenomenological neuro-psychiatry. Bridging clinic with clinical neuroscience. Springer, New York (in press)
Slaby J, Stephan A (2008) Affective intentionality and self-consciousness. Conscious Cogn 17:506–513
Sullivan MD, Younger SJ (1994) Depression. Competence and the right to refuse lifesaving medical-treatment. Am J Psychiatr 151:971–978

Tan JOA, Stewart A, Fitzpatrick R, Hope T (2007) Competence to make treatment decisions in Anorexia Nervosa: thinking processes and values. Philos Psychiatr Psychol 13:268–282

Trotter T (1804) An essay, medical, philosophical, and chemical, on Drunkenness, and its effects on the human body [used reference: Thomas Trotter (1821) Ueber die Trunkenheit und deren Einfluß auf den menschlichen Körper. Eine philosophische, medicinische und chemische Abhandlung, 1810. Memersche Hofbuchhandlung, Lemgo]

Ullrich S, Keers R, Coid JW (2013) Delusions, anger, and serious violence. New findings from the MacArthur violence risk assessment study. Schizophr Bull. doi:10.1093/schbul/sbt126

United Nations (2006) Convention on the rights of persons with disabilities. Retrieved 06-22-16 http://www.un.org/disabilities/convention/conventionfull.shtml

Vollmann J (2008) Patientenselbstbestimmung und Selbstbestimmungsfähigkeit. Kohlhammer, Stuttgart

Wiggins OP, Allen AC (eds) (2011) Clinical ethics and the necessity of stories. Essays in honor of Richard M. Zaner. Springer, Dordrecht

Yaffe G (2001) Recent work on addiction and responsible agency. Philos Public Aff 30:178–221

Mental Capacity of Adult Patients in Health Care

37

Jeanette Hewitt

Contents

Abstract

Mental capacity is a fundamental determinant of an individual's ability to make autonomous decisions. Respect for autonomy is a legal and ethical requirement in health-care provision, which necessitates that a person's autonomous wishes be respected and informed consent validly obtained before therapeutic intervention is carried out. In Britain and many other Western jurisdictions, mental capacity legislation has developed with the aim of providing a framework for

J. Hewitt (✉)
Centre for Philosophy, History, and Law, Swansea University, Swansea, UK
e-mail: j.l.hewitt@swan.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_28

the assessment of mental capacity in health care, in a decision-specific context. Where a patient is judged to lack mental capacity with regard to a decision, the duty to respect autonomy is superseded by the duty to act beneficently and/or prevent harm which might otherwise occur due to the patient's lack of capacity. Mental capacity legislation typically provides procedural criteria for assessing task-specific competence in terms of comprehension, appraisal, and communication. Procedural criteria do not however specify a threshold for competency assessment, or provide guidance on evaluation of irrational belief systems. Procedural assessment of mental capacity may therefore provide only a partial indication of a person's autonomy, and further evidence in terms of instrumental rationality may be necessary to evaluation of capacity.

Introduction

Mental capacity is a term which denotes an ability to exercise autonomy with regard to discrete decision-making. Patients in many countries are legally entitled to make choices about the type of treatment they receive and to refuse treatment if they so wish, unless they have been shown to lack mental capacity (Buchanan 2004). Respect for autonomy is an ethical and legal tenet of health care, which requires that patients be allowed to make self-determined treatment choices, even where those choices conflict with medical opinion (Breden and Vollmann 2004). Informed consent is a crucial constituent of respect for autonomy, which requires that patients be given appropriate information necessary to make informed decisions and that clinicians assess the patient's understanding of that information before therapeutic intervention is performed. In order for consent to be considered valid, three conditions must be satisfied: Firstly, consent must be given voluntarily; it must be given freely, without coercion or unwarranted interference by others. Secondly, consent is only "informed" if it is given knowledgeably, after necessary information has been given that allows the patient to consider the risks, benefits, and alternative courses of action available, and thirdly, it must be given intentionally, with comprehension and reason (Beauchamp and Childress 2009). The assessment of mental capacity in health care is primarily concerned with this third condition – intentionality – the person's capability to comprehend information and to use that information to make an autonomous decision.

Not all decisions will be autonomous despite adequate protections from coercion and ignorance; patients may lack capacity to make informed choices due to internal conditions that affect cognitive capacity such as delirium, disease, or injury (Beauchamp and Childress 2009). Mental incapacity to make autonomous decisions may be enduring (e.g., severe intellectual disability, brain damage, advanced dementias), temporary (e.g., sepsis, states of altered consciousness brain trauma), or fluctuating (e.g., serious mental illness). In cases where mental capacity is impaired, the question of best interests then arises in determining beneficent action. What counts as beneficent action may be relatively clear in cases where the previous wishes of the patient are known or where action is directed toward immediate

life-sustaining treatment of a previously healthy adult (although there may be exceptions to this where religious/cultural prohibitions apply). More complex questions can arise in determining beneficence where capacity is fluctuating, where the person's previous decisions are unknown, where there are risks that treatment may cause harm, or where clinical judgment is clearly at odds with the current beliefs and wishes of the patient.

The Department of Health (2005) has led to a formalization of the law in England and Wales with regard to treatment of persons without mental capacity in relation to their physical needs. Under sections "The Importance of Mental Capacity to Health Care" and "Assessment of Mental Capacity: Theoretical and Practical Considerations" of the Mental Capacity Act, a person is said to lack capacity where he or she is unable to make a decision in relation to a matter because of an impairment of or a disturbance in the functioning of the mind or brain, which may be either permanent or temporary (Department of Health 2005). Being incapable of making a decision is manifested by an inability to understand or retain relevant information, an inability to weigh that information as part of a decision-making process, and an inability to communicate a decision by any means (Dimond 2008). In a similar way, the US case law has addressed the assessment of mental capacity through the development of a model which identifies four criteria for demonstrable capacity in relation to treatment: the ability to express a choice, the ability to understand relevant information, the ability to appraise such information in relation to self, and the ability to reasonably weigh treatment options (Okai et al. 2007).

Legal frameworks expressed in health-care policies provide discrete criteria for procedural assessment of mental capacity with regard to decision-making in health care. Procedural assessments do not however necessarily provide evidence of overall mental capacity, but have a task-specific remit. Where substituted judgment for a person who lacks mental capacity is likely to have enduring consequences for the person in terms of future capabilities or where acting in the person's best interests involves deprivation of liberty, procedural assessment of task-specific decisions may provide insufficient information in guiding judgments about actions to be taken in the person's best interests. Mental capacity as defined in the task-specific sense may be seen as one feature of a person's autonomy but not necessarily as an indication of the person's global capacity for autonomy in all senses. Some philosophical and psychiatric accounts of mental capacity have therefore debated whether overall rationality, beliefs, value systems, desires, and emotions should be taken into account in assessing the person's capacity to exercise autonomy in health care (e.g., Craige 2013; Charland 2014).

In the context of mental health care, the assessment of mental capacity may be complicated by the absence of "insight" – the acceptance by the patient that he or she is suffering from a mental illness that warrants treatment (Nordenfelt 2007). Capacity to comprehend, retain, and process information necessary to a decision may be retained in mental illness (Hindmarch et al. 2013), which under the conditions stipulated by mental capacity legislation would seem to indicate that mental capacity is unaffected. However, for some persons with serious mental

illness, underlying belief systems may be false due to disorders or distortion of mood or perception. Procedural assessments of mental capacity have therefore been less widely adopted in mental health care, where separate mental health law exists to permit the detention and treatment of people who suffer from a serious mental illness and pose a risk to themselves or others (Dawson 2008). The ethical problems of a status approach to capacity assessment in psychiatry (lack of mental capacity conferred by a diagnosis of serious mental illness) have however been increasingly debated in contemporary literature (e.g., Hotopf 2005; Hewitt 2010).

A further controversy highlighted in mental capacity accounts pertains to the issue of risk. There is some evidence to suggest that in practice, clinicians raise the threshold for proof of mental capacity according to the degree of seriousness associated with the consequences of the patient's decision (Buchanan 2004; Hotopf 2005). Although the factors taken into account will depend on how much the patient has to lose, competence may only be seen as an issue when a patient decides contrary to what others regard as in their best interests (Buchanan 2004). Central to the risk-threshold debate is therefore whether this practice constitutes inappropriate paternalism or whether a sliding scale of competence is a necessary safeguard to vulnerable patients where much is at stake.

In this chapter, the concept of mental capacity is explored in relation to the key debates identified in contemporary literature: the role and remit of procedural assessments of mental capacity, the fundamental constituents of decisional capacity for informed consent, the relevance of broader conceptions of autonomy and rationality to clinical assessments of decisional competence, and the contested place of value systems in assessment criteria. The relationship between mental illness and mental capacity is examined with reference to the limitations of mental capacity legislation for mental health practice. The remit of paternalism in deciding best interests for patients considered to lack mental capacity is discussed in the context of the risk-threshold debate.

Decision-Making Capacity

Accounts of mental capacity within the literature can be broadly grouped into philosophical, empirical, and legal (Charland 2014). Philosophical accounts of mental capacity have been concerned with the necessary characteristics of persons who possess decision-making capacity, i.e., autonomy, and those qualities which constituent autonomous reasoning – rationality, free will, liberty, values, and goals (e.g., Culver and Gert 1982; Buchanan 2004; Beauchamp and Childress 2009). Empirical accounts have examined the methods of assessing mental capacity in health care (e.g., Etchells et al. 1999; Hotopf 2005), the prevalence of mental capacity in different populations (e.g., Cairns et al. 2005; Okai et al. 2007), and the role of clinicians in applying mental capacity legislation in practice. Legal and medicolegal accounts have been concerned with legal competence, deprivation of liberty, substituted judgment, and agents who may decide on best interests (Cairns et al. 2010; Jacob et al. 2013). Health-care policy sets out the operational definition

of mental capacity and guidance for the assessment in clinical practice, the use of advanced directives, the roles of responsible others, and the scope and limitations of mental capacity legislation (e.g., Department of Health 2005).

Mental capacity for decision-making can be temporarily compromised in a usually competent person by factors that negatively influence physical or mental health. Loss of consciousness due to, for example, brain insult or injury, cerebrovascular accident, serious cardiac events, hypoglycemic coma, anesthetic, or intoxication, will at least temporarily prevent a previously competent person from possessing mental capacity. Extreme pain, sepsis, physical trauma, serious injury, and drug or alcohol intoxication may cause levels of consciousness to fluctuate. Psychoactive substances such as psychotropic drugs or alcohol may influence reasoning capability and judgment although consciousness is not lost. Mental illness or disorder may at times influence the person's capacity for reasoned thought or action; the experience of hopelessness, depression, and suicidal thoughts may, for example, prevent the person from appreciating the potential for alternative futures through changed action or successful treatment. Psychotic disorders are characterized by delusions (fixed, false ideas) and hallucinations (false perceptions) usually accompanied by a spectrum of thought disorder. Mental capacity for persons experiencing psychotic disorders may vary according to the type of decision in question – whether likely to be directly influenced by delusional thinking or unrelated to delusional content, the severity of psychotic phenomena at the given time of decision-making, and the level of awareness that the person has of possessing a mental disorder. Contextual variables such as unfamiliar or threatening clinical environments, type and complexity of information which needs to be processed, and anxieties which may accompany discussion of disease, illness, or treatments may also influence the capacity of persons to make competent, self-determined decisions.

Lack of Mental Capacity

Indicative signs of mental incapacity may include disorientation or delirium, severe behavioral abnormality (although behavior in itself is not a reliable indicator), a history of cognitive impairment, concerns raised by others that the person's mental function exhibits signs of gross impairment, or refusing recommended treatment without being able to give a rationale (Nicholson et al. 2008). Mental incapacity is not associated with gender, educational level, social classification, or any sociodemographic variable apart from advancing age (Kim et al. 2007; Okai et al. 2007; Dawson 2008), although age in itself is not a reliable indicator of mental incapacity. Memory impairment may influence mental capacity, but an ability to retain information relevant to a decision for only a short period of time does not necessarily bar a person from being regarded as competent to make the decision (Department of Health 2005). Dementias such as Alzheimer's disease are likely to impact on the ability for comprehension and retention of information necessary to make an autonomous decision. However, disease progression in

terms of the extent of cognitive impairment is highly variable in patients diagnosed with dementia (Dresser 1996; Tschanz et al. 2011), and therefore the diagnosis alone should not be taken as an indicator of mental incapacity.

The contributory factors which may result in mental incapacity may therefore be diverse, and the existence of certain conditions or diagnoses (e.g., advanced age, Alzheimer's disease) is not a reliable indicator of mental incapacity in itself. A person may lack mental capacity if he or she is unable to make a self-determined decision due to any temporary or enduring impairment or disturbance of cognitive functioning. A robust assessment of mental capacity is therefore likely to be dependent on evidence obtained in a context-specific, case-by-case basis rather than by reference to particular physical or mental states of being.

The specific evidence of mental incapacity relevant to health-care provision is typically obtained through procedural assessment; health-care policy protocols and validated psychometric tools are used to measure capacity in relation to specific criteria. Although specific measures will vary according to jurisdictions, the decision-specific approach to the assessment of capacity typically judges incapacity to be evidenced where a person may be unable to understand and retain information relevant to the decision at hand, and the decision-making process fails to understand the likely consequences of the decision and is unable to communicate by any means an understanding and appreciation of the same (Department of Health 2005). By means of procedural assessment, evidence of capacity is obtained through evaluating the process by which the decision has been reached, rather than by simply evaluating the decision itself or the condition presented by the patient.

Mental Illness

Mental illness is viewed as an influencing factor which may potentially impair mental capacity and sometimes render a person incapable of making an autonomous decision with regard to treatment (Nordenfelt 2007). People with mental illness are typically seen to lack mental capacity for treatment decisions due to two factors: The first factor relates to a potential lack of voluntariness due to the illness state. Voluntariness is usually taken to mean the degree to which a person controls his or her actions without undue external influence (Beauchamp and Childress 2009). Mental illness is seen as being capable of diminishing or voiding voluntariness (Beauchamp and Childress 2009), due to the significant influencing forces of the illness. These forces may interfere with rational perception and reasoning processes, leading to a "prolonged inability to know and deal in a rational and autonomous way with oneself and one's social and physical environment" (Edwards 1982: 70). Mental illness has to do with the volitional and emotional machinery of action, with the person's intentions and reasons, and with his or her moods and emotions, which in turn determine the person's intentions and reasons (Nordenfelt 2007: 88). Reasons can be defective in the sense that they are inefficient; they do not result in the action in question or are defective in the sense that they do not give good reason for the action in question (Nordenfelt 2007: 99).

Unjustified or delusional beliefs are typically characteristic of serious mental illness associated with psychotic symptoms and may give rise to actions which are not fully intentional, in the sense that they do not stem from rational beliefs. Voluntariness may therefore be impaired where the basis for decision-making is influenced by irrational beliefs, where these beliefs are upheld even in the face of overwhelming evidence to the contrary.

The second factor that is perceived to reduce the potential for mental capacity in persons with serious mental illness is a lack of insight into the presence of the illness itself and the ways in which that lack of insight influences reasoning abilities and judgment (Grisso and Appelbaum 1995; Grisso et al. 1997; Nordenfelt 2007). The term "insight" in the context of serious mental illness is typically taken to refer to a person's awareness that he or she is suffering from a mental illness, that certain thoughts and perceptions are directly related to psychotic phenomena, and that psychiatric treatment is required to remedy these abnormal mental events (Carroll et al. 1999; Mintz et al. 2003). Clinical evaluations regarding the person's insight are significant in relation to task competence, where a lack of insight is seen to indicate an inability to comprehend treatment options and to choose or refuse treatment autonomously. Having awareness and acceptance of one's illness state is therefore taken to be an important indicator of mental capacity (Cairns et al. 2005).

It is however easier to assess mental capacity in patients with chronic but stable conditions such as severe intellectual disability or dementia than in those with an acute mental illness, in which fluctuations in capacity are the rule rather than the exception (Chiswick 2005). Historically, a diagnosis of serious mental illness was seen to equate to characteristic irrationality, which inevitably led to an alleged inability to engage with treatment decisions. The idea that serious mental illness necessarily constitutes irrationality of a global nature, which continuously impairs functioning, has now been contested; isolated irrationality (limited to discrete beliefs) is more consistent with delusional thought and complete disconnection from reality is rare (Cholbi 2009; Hewitt 2010). A diagnosis of mental illness therefore does not necessarily mean that a person is incompetent to consent to or refuse treatment (Cherry 2010); global mental incapacity is not an inevitable consequence of serious mental illness but incapacity is likely to vary according to the decision at hand and whether that decision is directly influenced by the content of delusional thought.

The Importance of Mental Capacity to Health Care

Respect for Autonomy: Informed Consent

Autonomy is the capacity for self-government, whereby an individual who has the ability to reason is free from external controlling forces and is capable of controlling actions and working toward achieving higher goals or intentions (Beauchamp and Childress 2009).

Autonomy is generally held as an important good for people, associated with ideas of individual liberty, self-determination, freedom of choice, and the right to make independent decisions for the self and the acceptance of responsibility for one's moral positions.

In the Western world, respect for autonomy is typically considered to be an accepted entitlement for persons who are deemed to possess the cognitive capacity to deliberate and act on those deliberations without internal constraint (Beauchamp 1986). The concept of autonomy is central to the area of applied moral philosophy in the biomedical context, where all discussions of the nature of informed consent and its rationale refer to patient autonomy (Dworkin 1988).

The historical origins of decisional capacity are embedded in the moral principle of respect for autonomy (Charland 2014). Respect for autonomy includes obligations to acknowledge a person's right to hold views, to make choices, and to take actions based on personal values and beliefs (Beauchamp and Childress 2009). Respecting this moral principle involves more than simply acknowledging that people are entitled to their opinions or noninterference with their decisions. It entails, in some circumstances, enhancing a person's capacity to make autonomous choices by providing information or access to such means as to enable the enactment of autonomous choices (Hewitt and Edwards 2006). However, it is generally accepted that there are limits to respecting a person's autonomy; a person's claims for autonomy should not cause serious harm or incur undue costs to other persons, should not arbitrarily infringe on the rights of others, or should not conflict with other moral obligations that may rightly supersede personal desires.

Different meanings have been ascribed to autonomy within the literature, whereby the term is often used interchangeably to describe related concepts and ideals:

> It is used sometimes as an equivalent of liberty. . .sometimes as equivalent to self-rule or sovereignty, sometimes as identical with freedom of the will. It is equated with dignity, integrity, individuality, independence, responsibility, and self-knowledge. It is identified with qualities of self-assertion, with critical reflection, with freedom from obligation, with absence of external causation, with knowledge of one's own interests. . . It is related to actions, to beliefs, to reasons for acting, to rules, to the will of other persons, to thoughts and to principles. About the only features held constant from one author to another are that autonomy is a feature of persons and that it is a desirable quality to have. (Dworkin 1988: 6)

The capacity to be fully autonomous is not generally taken to be a static or all or nothing state of being in most cases. It is potentially variable, wherein persons may fluctuate in their abilities or capacities to be autonomous at any given time. Different individuals at different phases, in different circumstances of their lives, can occupy different locations on the continuum of autonomy. If a person's beliefs concerning some matter are false, ill informed, or inconsistent, then he or she is not autonomous with respect to that matter, at that moment (Varelius 2003). For example, a person may fail to possess the relevant knowledge or skills required for an informed decision with regard to some action, thereby curtailing the person's choices and consequently the capacity for autonomy.

Autonomy, understood in the broad terms described above, requires a range of optimum capabilities and circumstances. This means that an ideal state of autonomy is likely to be somewhat elusive, since complete freedom from external influences is largely impossible, and full comprehension of all information available in all circumstances is unlikely to be realized (Blackburn 2005). The notion of autonomy as a complex and multifaceted phenomenon has long been viewed as problematic by health-care providers who have a legal and ethical requirement to respect patient autonomy but who also have a duty of care toward patients who lack autonomy and/or are vulnerable to exploitation or coercion. While mental health law existed in most Western countries (e.g., the UK Mental Health Act 1983) to stipulate the remit of care and treatment for patients who were deemed non-autonomous due to mental disorder, treatment for patients who lacked capacity to make decisions about their physical needs was dealt with under common law or legislated by the courts (Nicholson et al. 2008). Mental capacity legislation in much of the Western world therefore developed with the aim of assisting clinicians and the courts to assess the task-specific competence of patients in procedural terms. The aim of such legislation was to attempt to reduce misuse of paternalism while ensuring that the best interests of patients who lacked mental capacity were served (Craige 2013).

Paternalism: Best Interests

The term paternalism is used to describe action that involves some type of interference with another's actions or preferences on the grounds that such interference is necessary to observe the principles of beneficence or non-maleficence (Beauchamp and Childress 2009). Benjamin and Curtis (1986: 57) argued that only three conditions can justify paternalism, all three of which need to be satisfied at the same time: The first condition is one of an impairment of autonomy, where in present circumstances the person is "irretrievably ignorant of relevant information, or his or her capacity for rational reflection is significantly impaired." The second condition is that predicable and significant harm is likely to occur unless intervention takes place, and the third is the ratification condition, where it is assumed that at some future time, upon the resumption of autonomy, the person will ratify the decision taken to intervene. The prediction of harm is particularly important to the justification of paternalism, wherein the consequences of a person's acts must be serious enough to defend intervention (Culver and Gert 1982).

Beauchamp and Childress (2009) distinguished between the concepts of weak and strong paternalism. In weak paternalism, intervention is permitted to prevent the substantially non-voluntary conduct of non-autonomous actions, whereby the person is protected from harms that may be brought about by conditions beyond their control. For example, a person with advanced cognitive impairment consequent to dementia may act in ways which pose a danger to the self due to severe memory impairment (e.g., wandering from home in the middle of the night), but such actions are not intentional, i.e., they are not actions arising from reasoned

decisions. Paternalistic intervention which prevents unintentional acts in this sense is weak, in that it involves no conflict between autonomy and beneficence since unintentional acts are not autonomous. Strong paternalism however involves beneficent intervention that interferes with another's actions, even though those actions are informed, voluntary, and autonomous (ibid). Justification of paternalism would seem the easiest where it may be shown that the person substantially lacks autonomy. However, Beauchamp and Childress (2009) argued that beneficence can sometimes provide grounds for intervening even in autonomous actions; even autonomous persons might temporarily lose their ability to rationally reflect on their conduct and as such they should be prevented from an act that is dangerous and irreversible.

The term paternalism often has negative connotations in its depiction in contemporary health-care literature; it is typically linked with unwarranted force and coercion, medical dominance, and disempowerment of patients. Western emphasis on self-determination and choice in health care has led to a shift in what patients now expect from their relationships with health-care providers. Medical authority to decide what is best for patients is no longer accepted unquestioningly, and partnership and patient centeredness are seen as more appropriate values in present-day accounts of ethical health-care practice (Jensen and Mooney 1990). One of the challenges presented by this paradigm shift is how to best avoid unwarranted interference with the self-determined decisions of competent patients while protecting the interests of those patients whose capacity for decision-making has been compromised. Decisions which run contrary to medical advice are not evidence in themselves of a lack of mental competence – competent patients can refuse treatment, even where the potential risks are great. Mental capacity legislation in many Western countries has evolved within the context of these concerns, attempting to provide procedural assessment criteria for the evaluation of decisional competence, which is intentionally limited in scope. The limited scope of decision-specific assessment is aimed at reducing unjustified paternalism, however, well intended, while providing for those whose actions are not fully intended and who might otherwise be harmed by a lack of medical intervention (Hotopf 2005).

Mental capacity legislation concerned with decision-making competence usually provides for decisions and actions undertaken on behalf of a person who lacks capacity in terms of best interests. Treatment without consent is only permitted where there is clear evidence that a person's capacity to competently consider treatment options is compromised (Department of Health 2005). The UK Department of Health (2005) stipulates that decisions made for people lacking capacity must be the least restrictive option for their rights and freedom of action. It should be recognized that incapacity may be a temporary state, and competence to reclaim self-determination should be reassessed over time. Where decisions and actions are undertaken in the best interests of the person, these should, where practicable, optimize the person's ability to participate in decision-making and aim at returning autonomy to the person where possible (Beauchamp and Childress 2009). Consideration of best interests should take into account where possible the

person's past and present wishes, feelings, beliefs, and values that would be likely to influence decision-making if mental capacity was retained (and, in particular, any relevant written statements made when capacity was retained) (Department of Health 2005).

Mental Health Care: Justice

Lack of mental capacity in psychiatric inpatients has been variously estimated at between 20 % and 45 % and has been particularly associated with delusions, mania, and hypomania (Cairns et al. 2005; Dawson 2008). However, despite a closer correlation between delusions and mental incapacity, studies undertaken on mental capacity in psychiatric inpatients appear to show that psychosis is not invariably associated with a lack of decisional competence (e.g., Kemp et al. 1997; Belhouse et al. 2003; Owen et al. 2007, 2008).

Assessment of competence in psychiatry is typically concerned with a patient's expressed willingness to consent to treatment considered warranted by a diagnosis of mental illness. Where a diagnosis of serious mental illness has been made, clinicians have a tendency to equate treatment refusal with mental incapacity and treatment acceptance with capacity rather than adopting a procedural approach to the assessment of treatment decisions (Okai et al. 2007).

Most Western countries have mental health laws which legislate for the use of involuntary inpatient treatment (Jarrett et al. 2008). Mental health law in the UK, USA, Canada, Australia, and New Zealand typically has a different agenda and scope to legislation addressing task-specific mental capacity. Whereas mental capacity assessment concerns itself with a person's ability to demonstrate competence with regard to specific decisions in relation to treatment, mental health law concerns itself with particular conditions which, if present, may allow treatment to be given without patient consent (Okai et al. 2007).

The conditions that together are necessary to allow compulsory admission to hospital (evaluation and treatment) defined in Western mental health law can be roughly summarized as follows: (1) the person has a mental disorder; (2) as a result of mental disorder, the person is a danger to self or others; and (3) the person would likely benefit from treatment, which is available for such a mental disorder (Mental Health Act, 1983, 2007 (UK); Mental Health Services Act, 2004 (USA); MHA, 1992 (NZ)). The length of permitted detention and definition of mental disorder varies from country to country, but all five countries permit mental health detention on the basis of certain general conditions, without the necessity of decision-specific competency assessment.

In the UK, the Mental Health Act (1983) has a different agenda and scope to the Mental Capacity Act (2005). A person subject to detainment under the MHA may possess mental capacity, but can still receive psychiatric treatment against his or her will. The Mental Capacity Act is used to assess competence with regard to specific decisions and does not have the wide-ranging powers of compulsion that are conferred by the Mental Health Act. Psychiatrists may use either Act, but in

practice, the Mental Health Act (1983) is typically used for patients with a status diagnosis of serious mental disorder.

The use of status approaches to capacity assessment has numerous implications for psychiatric patients: Under mental capacity legislation, treatments are only provided in the patient's best interests (with particular attention paid to previously expressed wishes, including advance directives, which have legal weight) (Nicholson et al. 2008), whereas under mental health legislation, best interests are considered alongside potential harms to others. A status approach is taken where a person, having reached a diagnostic threshold, would be described as lacking capacity for all decisions, which means that the patient can be given a range of treatments, even if he or she might have the capacity to refuse one or more of these (Okai et al. 2007). Where a person is judged as meeting the criteria for involuntary treatment, decision-specific capacity need not be assessed (Hotopf 2005).

The differences in mental capacity legislation as compared with mental health law have been seen as ethically problematic (Hotopf 2005). Although a competent person with a physical illness can reject treatment that is clearly in his or her best interests, mental health legislation can override psychiatric patients' decisions to withhold consent for treatment of their mental disorders. Respect for autonomy is therefore not absolute in the same way as in legislation for the treatment of physical illnesses even where it can be shown that the person possesses decisional competence in procedural terms (Cairns et al. 2005). The Mental Capacity Act (2005) may be used only when a person lacks the capacity to consent. The Mental Health Act (1983), in contrast, can be used regardless of a person's capacity to consent if the Act's different criteria of mental disorder, risk of harm, availability of treatment, and so on apply (Department of Health 2015). Dawson (2008) observes that many detained patients will recover their capacity after initial treatment or their capacity may fluctuate. If incapacity principles were strictly applied, such patients would have to be swiftly released from involuntary treatment whenever they regained their capacity, an outcome that might preclude the provision of sustained treatment (Dawson 2008).

The underpinning arguments for adopting a status approach to capacity assessment typically refer to the capacity for voluntariness – the ability to act without undue influence. Serious mental illness may potentially impair voluntariness due to irrational beliefs which are persistently held despite available evidence to the contrary and from which arise irrational desires and actions (Beauchamp and Childress 2009). Beliefs may be seen as irrational if they are deficient in certain ways or if they stem from imprudent desires. Some desires may be irrational because they are damaging to the person (e.g., deliberate self-harm) or work against his or her best interests (e.g., addictions). They may be unreasonable, conflicting, incongruous, mistaken in their founding assumptions, or partially or wholly self-deceiving (Graham 1998). Decisions are founded on belief systems – a person's appraisal of what things are likely to be good or bad to have in terms of one's self-interests. Therefore, even if process reasoning remains intact, where there are irrational beliefs arising from serious mental illness, these may negatively influence perception, desires, and appraisal. The fundamental issue is then not simply

whether a person is able to demonstrate the capacity to make a decision but whether such a decision can be truly autonomous given the influence of mental illness (Charland 2014).

Despite the plausibility of arguments that suggests that delusional beliefs constitute characteristic irrationality, there is some evidence to suggest that delusions are typically isolated to one theme in psychotic disorder – people with serious mental illness are not deluded about everything or apparently at random, and complete disconnection from reality is extremely rare (Owen et al. 2007, 2008). Therefore, decisional approaches to mental capacity assessment are of significant use in psychiatry, and it may be that status approaches to competency determination unjustly miss opportunities to involve patients in care decisions.

Assessment of Mental Capacity: Theoretical and Practical Considerations

Clinical research into the assessment of mental capacity in general hospital patients indicates that approximately one third of patients lack decisional capacity, typically caused by cognitive impairment subsequent to delirium or dementia (Hotopf 2005). A lack of mental capacity is the inability to make a particular decision because of an impairment or disturbance of the mind or brain, whether temporary or permanent, assessed at the time the decision needs to be made (Department of Health 2005).

The factors considered relevant to an assessment of mental capacity include the person's general intellectual ability, memory, attention and concentration, reasoning, and information processing – how a person interprets what they are told, verbal comprehension, forms of communication, cultural influences, social context, and overall ability to communicate (Department of Health 2005). As capacity may fluctuate, treatments which are extended over a lengthy period of time may require that mental capacity is repeatedly assessed (Buchanan 2004).

Assessment of mental capacity should include evidence that the person is able to understand information relevant to the decision, retain that information, use or weigh the information in the process of making the decision, and communicate that decision by any means, including nonverbal methods where verbal communication is not possible (Department of Health 2005).

Competent patients are entitled to make unwise decisions and to refuse treatments even where these decisions conflict with clinical judgments about best interests. It is not the decision itself but the process by which the decision is reached that determines if capacity is absent, however risky or grave the outcome of that decision (Nicholson et al. 2008).

Charland (2014) describes four constituents of decisional capacity relevant to health-care assessments developed in the USA by Grisso and Appelbaum (1995) and Grisso et al. (1997), which have been widely discussed in contemporary literature: understanding, appreciation, reasoning, and choice, and a fifth constituent of values which has received varying support. The four constituents proposed by Grisso and Appelbaum (1995), Grisso et al. (1997) have generally received support

in clinical literature, due in part to the availability of validated psychometric tools incorporating these constituents, which are accessible for clinical assessments (e.g., MacArthur Competence Assessment Tool – Treatment (MacCAT-T) (Hotopf 2005) and offer measurable outcomes for practical evaluations.

Understanding

The mental capacity to make a competent decision depends on a person's ability to understand and retain essential information relevant to the decision, including the costs and benefits of deciding one way or the other and the ability to reason with that information and use it to exercise a choice (Szmukler and Appelbaum 2008; Beauchamp and Childress 2009). Comprehension of information is however likely to be a matter of degree – the threshold for minimum comprehension necessary for decision-making is likely to be influenced by the type and significance of the decision in question, the associated risk, and to some extent the judgments set by the assessor. Although ostensibly the assessment of mental capacity with regard to discrete decision-making is easier to measure than overall autonomy since it is primarily concerned with the ability to follow a logical reasoning process within the context of particular parameters, threshold questions are likely to involve uncertainty and in some cases errors in accurately assessing necessary understanding.

Appreciation

Grisso and Appelbaum's (1995), Grisso et al. (1997) theory of mental capacity includes a requirement for insight (referred to in their work as appreciation). Appreciation is defined as recognition that one is suffering from a disorder and that the generally accepted risks and benefits of treatment apply to one's own situation. Appreciation is concerned with the person's realistic perceptions of current events and circumstance, the significance of the decision to be made, and their capacity to choose alternative actions.

Although lack of insight into one's illness state is typically seen as a problem associated with mental disorder, it is also seen in patients who have been given a poor prognosis subsequent to serious physical illness (Fried et al. 2003). Although mental capacity literature has been largely concerned with cognitive processes necessary to decision-making, affective states may have considerable significance in terms of the person's preparedness for information gathering, processing, and communication. For example, anxiety, which may accompany decision-making in stressful situations, can impair the ability to consider alternative courses of action and lead to selective attention of information gathering (Keinan 1987; Beck and Clark 1997). The significance of emotions to appraisal is therefore not confined to mood disorders; emotions (both positive and negative) are likely to play a significant role in hearing, understanding, and relating information to the self.

Reasoning

Reasoning capacity is a function of rational agents. Rational agents perform intentional acts, i.e., acts that are not random but are undertaken for a discernible reason. The behavior of rational agents can be explained by means of reason explanations; it is possible to pick out practical reasoning processes that cause agents to act as they do (Fay 1996). The ability to evidence reason explanations is important to the procedural assessment of mental capacity, since evaluation of the process by which a person reaches a decision is the central concern.

Practical reasoning is a reflective process by which a person arrives at a decision to act; it is typically engaged in seeking the best course of action in a set of circumstances given all potential alternatives for action. Practical reasoning is not concerned with arriving at objective truths, but rather with considering what decisions seem best, all things considered, in relation to the self (Wallace 2014). Theoretical reason by contrast is engaged in seeking the truth of claims or beliefs; it is concerned with the formation of "true beliefs" or what one ought to believe on the basis of available evidence (ibid) and can be taken as equivalent to saying that a person is acting rationally if he or she has an undistorted view of reality.

Mental capacity assessment is ostensibly concerned with the assessment of practical rather than theoretical reasoning – the ability to comprehend and appraise information in relation to the self and use that appraisal to reach a decision about the costs and benefits in relation to a particular course of action. However, assessment of practical reasoning does not allow for irrational belief systems that may underpin reflection and give rise to irrational desires (Culver and Gert 1982). Decisional-capacity assessments which only focus on process may potentially overlook irrational thoughts, desires, or actions that arise from mood disorders or disorders of perception that can accompany mental illness.

Mental capacity legislation typically allows for some decisions that may appear to be irrational by some others (e.g., refusal of life-saving treatment due to religious reasons (Charland 2014) which are nevertheless not to be seen as indicative of mental incapacity, but rather as stemming from culturally significant belief systems. Since assessment of mental capacity is concerned with process reasoning rather than beliefs per se, belief systems should be irrelevant to evaluations. However, in practice, assessment is likely to be less straightforward – choices which stem from compromised belief systems (e.g., in the case of acute mental illness) may indicate a temporary lack of decisional capacity, even though comprehension of information and so forth is evidenced. Clinical evaluations of mental capacity must therefore accept some decisions which seem at odds with rationality while rejecting those which arise from unhealthy belief systems (Charland 2014).

Choice

Choice is constrained by a person's capacity for voluntariness. Restrictions upon voluntariness have been broadly grouped in terms of influence, lack of control, and

unintentional action (Nelson et al. 2011). External influence includes inducements, implicit or explicit coercion, deception, and nondisclosure or manipulation of information. Control may include the preceding factors and may also denote threats, physical restraint, and force (Beauchamp and Childress 2009). Not all influence is malign or controlling; necessary advice given by a clinician in order to assist decision-making will likely influence a decision, but is not necessarily a constraint to voluntariness (Nelson et al. 2011). Constraints upon voluntariness may be difficult to assess in practice and may include a range of factors such as unwarranted pressures from others or internal conditions such as fear, addiction, mental disorder, or pain (ibid). Comprehension and appreciation of relevant information in relation to the self and potential actions are necessary for intentional action, and these factors may also present difficulties in assessment in practice, in terms of the threshold required for competence.

As previous discussion has shown, mental capacity legislation typically stipulates that the nature of the choice that a patient makes about treatment is not immediately relevant to an assessment of competence; even choices which conflict with medical advice must be respected, where patients evidence mental capacity in the *process* of decision-making. However, in practice, two qualitative factors are likely to have significance for the assessment of capacity: the threshold for comprehension and appreciation of risk.

Competence to make a decision is always likely to be a matter of degree (Hotopf 2005) – the ability to comprehend all necessary information and appreciate fully all relevant aspects of one's situation, and the gravity of the decision at hand is liable to be variable even in patients who possess overall autonomy. The challenge then is one of determining where the threshold for the necessary degree of comprehension and appreciation should be set and whether the setting of such a threshold should depend on the degree of appreciation generally accepted to be acceptable or the degree of risk associated with the decision.

In practice, clinicians are likely to increase the threshold for decisional capacity, depending on the seriousness of what is at stake (Hotopf 2005). The standard explanation is that respect for patient autonomy is being balanced against their best interests – the more serious the decision, the greater the duty of care to determine that sufficient capacity is held (Buchanan 2004). Where gravity is extreme, doctors and courts may in some circumstance allow their disapproval of what the patient proposes to outweigh the patient's wishes, whatever the patient's capacity (Buchanan 2004). In terms of logical consistency, this type of risk-threshold approach to the assessment of capacity has been disputed; capacity to make a particular decision is either in evidence or it is not, and the riskiness of the decision of hand cannot in itself alter the determinants of competence (Charland 2014). However, the ethical requirements of clinical decision-making in practice may justify the desirability of a more in-depth assessment of capacity where the outcomes for the patient are serious and irreversible (Hotopf 2005). Raising the threshold level of capacity required for competence when the anticipated harm is the greatest stems from a clinician's wish to be more certain, leaving a greater margin for error when the consequences are grave. However, this practice also

increases the potential number of instances in which people may be incorrectly assessed as incompetent (Buchanan 2004).

Values

Autonomy has two particular features which have been discussed in relation to the assessment of mental capacity in health care: agency and authenticity. Agency is the primary concern of procedural assessments of capacity and describes the capabilities of a person to reason and deliberate, appreciate one's circumstances, understand information, and communicate a choice (Buchanan 2004). Authenticity, which is concerned with the capability to reflect on the appropriateness of desires, values, and goals, has a more disputed place in discussion of capacity assessments (Brudney and Lantos 2011).

The concept of authenticity derives from the theory of instrumental rationality, which describes rationality in terms of acting so as to maximize the fulfillment of one's own desires and goals (Stanovich 2011). Our ability to be rational in this sense is not solely dependent on a basic capacity to exercise intelligent judgment or adherence to systems of externally set rules but to also form intentions in accordance with personal goals (Nordenfelt 2007). Instrumental rationality may therefore be displayed in logical consistency in regulating one's actions, plans, intentions, and affective states in accordance with one's own value systems (Svavarsdottir 2008). Congruence and coherence of thought are displayed in the ways in which beliefs, desires, perceptions, intentions, decisions, and actions fit together in ways that are not antagonistic but are consistent with the same goal or set of goals (Guttenplan 1999). Instrumental rationality is therefore concerned with a cluster of personal identity conditions, rather than a single belief or decision.

Instrumental rationality has been considered to be a more contentious factor in mental capacity assessments (Charland 2014). It has proved to be more difficult to define in procedural terms, since a person's motivation and intentions are largely private and not easily verifiable for others (Beauchamp 1986). Both agency and authenticity may however have relevance to assessments of capacity in practice. Where much rests on the outcome of a decision, clinicians may be more likely to consider the additional questions of whether the decision is part of a stable and coherent set of beliefs and values. For example, in the case of a Jehovah's Witness decision to refuse a life-saving blood transfusion, decision-specific capacity is considered in the context of wider belief systems and values (Brudney and Lantos 2011). A person's end-of-life decisions or refusal of life-sustaining treatment may well include consideration of current and expected quality of life as this relates to personal values and vital goals. What may seem like unwise or irrational decisions may take on different meanings where viewed in the context of wider belief systems. The inclusion of instrumental rationality may therefore have an important role to play in the assessment of mental capacity in some circumstances.

Conclusion

Mental capacity is the capability to demonstrate competence in decision-making. In health care, assessment of mental capacity is typically concerned with a patient's ability to demonstrate competence in the process of decision-making, rather than with the content of a decision itself. The essential elements of process competence are evidenced in comprehension, appraisal, and communication of a choice. Patients are entitled to refuse treatment and make unwise choices and have their wishes respected unless they demonstrably lack mental capacity. The law in the UK, USA, and Canada similarly includes a right for people to choose and refuse treatment and a requirement to assess mental capacity specific to the decision at hand.

Not all patients will demonstrate the capacity to make competent decisions, and the question then arises as to which actions constitute best interests. Best interests may be relatively straightforward where there is an immediate threat to the life or safety of the person, but may be less easy to determine where capacity is fluctuating, such as in the context of mental illness or where best interests involve deprivation of liberty and/or where best interest judgments seriously conflict with the expressed desires of the person.

Mental capacity assessment in mental health settings typically takes a status approach to competency determination and is concerned with the level of insight the patient demonstrates in relation to the illness state. The status approach to capacity assessment in mental health care is ethically contentious, since it assumes irrationality on the basis of a diagnostic threshold and may involve deprivation of liberty and treatment against a person's will. Decisional capacity is not necessary lacking in people with serious mental illness, but underpinning beliefs may be faulty or unhealthy and work against the prudential interests of the person. Where these conditions pertain, voluntariness to make decisions may be impaired until the person regains autonomy to make decisions unimpaired by mental illness.

The assessment of mental capacity in practice generally considers four determinants of capability: understanding, appraisal, reasoning, and choice. The role of emotions, desires, and value systems has received less attention, but may well be relevant, for instance, to end-of-life decisions and/or where treatment will significantly impact on a person's ability to pursue vital goals. The capacity for autonomous reasoning is dependent on a variety of capabilities and contextual opportunities, and mental capacity is liable to be a matter of degree at any given time. Assessment of discrete decision-making competence only gives an indication of capacity in a task-specific context. However, even within this particular domain of assessment, evaluation of the necessary threshold for decisional competence remains elusive.

Definitions of Key Terms

Insight	Awareness and acknowledgment of an illness state in relation to the self.
Instrumental rationality	Acting so as to maximize the fulfillment of one's desires and goals.

Mental capacity	An ability to exercise autonomy with regard to discrete decision-making.
Practical reasoning	A reflective process by which a person arrives at a decision to act.
Procedural assessment	Protocols used to measure mental capacity in relation to specific criteria, which is typically concerned with the process elements of decision-making.
Status approach/status diagnosis	Lack of mental capacity conferred by a diagnosis, usually of serious mental illness.
Theoretical reason	Seeking the truth of claims or beliefs.

Summary Points

- Mental capacity is the capability to make reasoned decisions.
- The assessment of mental capacity in health care is concerned with the ability to demonstrate competence in the process of decision-making rather than with the substance of the decision itself.
- Where a person demonstrably lacks decisional capacity, health-care professionals have a duty of care to consider best interests while avoiding unwarranted paternalism.
- Ethical challenges are posed by the degree of competence necessary for decision-making and the relevance of risk to competency judgments.
- The status approach taken to assessing mental capacity in mental health practice is contentious and raises questions about the relevance of beliefs and desires to evaluations of competence.

References

Beauchamp TL (1986) Suicide. In: Regan T (ed) Matters of life and death, 2nd edn. Random House, New York, pp 77–124

Beauchamp T, Childress J (2009) Principles of biomedical ethics. Oxford University Press, New York

Beck AT, Clark DA (1997) An information processing model of anxiety: automatic and strategic processes. Behav Res Ther 35:49–58

Belhouse J, Holland A, Clare I et al (2003) Capacity-based mental health legislation and its impact on clinical practice: 2. treatment in hospital. J Ment Health Law 24–28

Benjamin M, Curtis J (1986) Ethics in nursing, 2nd edn. Oxford University Press, New York

Blackburn S (2005) Oxford dictionary of philosophy, 2nd edn. Oxford University Press, Oxford

Breden TM, Vollmann J (2004) The cognitive based approach of capacity assessment in psychiatry: a philosophical critique of the MacCAT-T. Health Care Anal 12(4):273–283

Brudney D, Lantos J (2011) Agency and authenticity: which value grounds patient choice? Theor Med Bioeth 32(4):217–227

Buchanan A (2004) Mental capacity, legal competence and consent to treatment. J R Soc Med 97:415–420

Cairns R, Maddock C, Buchanan A, David A, Hayward P, Richardson G, Szmukler G, Hotopf M (2005) Prevalence and predictors of mental incapacity in psychiatric in-patients. Br J Psychiatry 187:379–385
Cairns R, Richardson G, Hotopf M (2010) Deprivation of liberty: mental capacity act safeguards versus the mental health act. Psychiatrist 34:246–247
Carroll A, Fattah S, Clyde Z, Coffey I, Owens DG, Johnstone EC (1999) Correlates of insight and insight change in schizophrenia. Schizophr Res 35:247–253
Charland LC (2014) Decision-making capacity. In: Zalta EN (ed) The Stanford encyclopedia of Philosophy, Fall 2014 edn. forthcoming. http://plato.stanford.edu/archives/fall2014/entries/decision-capacity/. Accessed 20 Aug 2014
Cherry MJ (2010) Non-consensual treatment is (nearly always) morally impermissible. J Law Med Ethics 38(4):789–798
Chiswick D (2005) Commentary: test of capacity has little practical benefit. Br Med J 331:1469
Cholbi M (2009) Tonkens on the irrationality of the suicidally mentally ill. J Appl Philos 26 (1):102–106
Craige J (2013) Introduction: mental capacity and value neutrality. Int J Law Context 9(SI 01):1–3
Culver CM, Gert B (1982) Philosophy in medicine: conceptual and ethical issues in medicine and psychiatry. Oxford University Press, New York
Dawson J (2008) Mental capacity and psychiatric admission. Br Med J 337:5–6
Department of Health (2005) Mental capacity act. HMSO, London
Department of Health (2015) Mental Health Act 1983 Code of Practice. HMSO, London
Dimond B (2008) Legal aspects of nursing, 5th edn. Pearson Education, Harlow
Dresser R (1996) Mentally disabled research subjects: the enduring policy issues. JAMA 276 (1):67–72
Dworkin G (1988) The theory and practice of autonomy. Cambridge University Press, New York
Edwards RB (1982) Mental health as rational autonomy. In: Edwards RB (ed) Psychiatry and ethics. Prometheus, New York, pp 68–78
Etchells E, Darzins P, Silberfeld M, Singer PA, McKenny J, Naglie G, Katz M, Guyatt GH, Molloy DW, Strang D (1999) Assessment of patient capacity to consent to treatment. J Gen Intern Med 14:27–34
Fay B (1996) Contemporary philosophy of social science. Blackwell, Oxford
Fried TR, Bradley EH, O'Leary J (2003) Prognosis communication in serious illness: perceptions of older patients, caregivers, and clinicians. J Am Geriatr Soc 51:398–1403
Graham G (1998) Philosophy of mind: an introduction, 2nd edn. Blackwell, Malden
Grisso T, Appelbaum PS (1995) The MacArthur Competence Treatment Study: III. Abilities of patients to consent to psychiatric and medical treatments. Law Hum Behav 19:149–174
Grisso T, Appelbaum PS, Hill-Fotouhi C (1997) The MacCAT-T: a clinical tool to assess patients' capacities to make treatment decisions. Psychiatr Serv 48:1415–1419
Guttenplan S (1999) A companion to the philosophy of mind. Blackwell, London
Hewitt J (2010) Rational suicide: philosophical perspectives on schizophrenia. Med Health Care Philos 13:25–31
Hewitt J, Edwards SD (2006) Moral perspectives on the prevention of suicide in mental health settings. J Psychiatr Ment Health Nurs 13:665–672
Hindmarch T, Hotopf M, Owen GS (2013) Depression and decision-making capacity for treatment or research: a systematic review. BMC Med Ethics 14:54
Hotopf M (2005) The assessment of mental capacity. Clin Med 5(6):580–584
Jacob R, Gunn MJ, Holland A (eds) (2013) Mental capacity legislation: principles and practice. RCPsych Publications, London
Jarrett M, Bowers L, Simpson A (2008) Coerced medication in psychiatric inpatient care: literature review. J Adv Nurs 64:538–548
Jensen UJ, Mooney G (eds) (1990) Changing values in medical and healthcare decision-making. Wiley, Chichester

Keinan G (1987) Decision making under stress: scanning of alternatives under controllable and uncontrollable threats. J Pers Soc Psychol 52:639
Kemp R, Chua S, McKenna P, David A (1997) Reasoning and delusions. Br J Psychiatry 170:398–405
Kim S, Appelbaum P, Swan J, Stroup S, McEvoy J, Goff D, Jeste D, Lanberti S, Leibovici A, Caine E (2007) Determining when impairment constitutes incapacity for informed consent in schizophrenia research. Br J Psychiatry 191:38–43
Mintz AR, Dobson KS, Romney DM (2003) Insight in schizophrenia: a meta-analysis. Schizophr Res 61(1):75–88
Nelson RM, Beauchamp T, Miller VA, Reynolds W, Ittenbach RF, Luce MF (2011) The concept of voluntary consent. Am J Bioeth 11:6–16
Nicholson T, Cutter W, Hotopf M (2008) Assessing mental capacity: the Mental Capacity Act. Br Med J 336:322–325
Nordenfelt L (2007) Rationality and compulsion. Oxford University Press, Oxford
Okai D, Owen G, McGuire H, Singh S, Churchill R, Hotopf M (2007) Mental capacity in psychiatric patients. Br J Psychiatry 191:291–297
Owen G, Cutting J, David A (2007) Are people with schizophrenia more logical than healthy volunteers? Br J Psychiatry 191:453–454
Owen G, Richardson G, David A, Szmukler G, Hayward P, Hotopf M (2008) Mental capacity to make decisions on treatment in people admitted to psychiatric hospitals: cross sectional study. Br Med J 337:448
Stanovich K (2011) Rationality and the reflective mind. Oxford University Press, New York
Svavarsdottir S (2008) The virtue of practical rationality. Philos Phenomenol Res 77:1–33
Szmukler G, Appelbaum PS (2008) Treatment pressures, leverage, coercion, and compulsion in mental health care. J Ment Health 17:233–244
Tschanz JT, Corcoran CD, Schwartz S, Treiber K, Green RC, Norton MC, Lyketsos CG (2011) Progression of cognitive, functional and neuropsychiatric symptom domains in a population cohort with Alzheimer's dementia the cache county dementia progression study. Am J Geriatr Psychiatry 19(6):532
Varelius J (2003) Autonomy, subject-relativity, and subjective and objective theories of well-being in bioethics. Theor Med 24:363–379
Wallace RJ. Practical reason. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, Summer 2014 edn. http://plato.stanford.edu/archives/sum2014/entries/practical-reason/. Accessed 2 Sept 2014

Patients' Responsibility for Their Health

38

Martin Langanke, Wenke Liedtke, and Alena Buyx

Contents

M. Langanke
Faculty of Theology, Systematic Theology, Ernst Moritz Arndt University of Greifswald, Greifswald, Germany

Institute for Ethics and History of Medicine, University Medicine Greifswald, Greifswald, Germany
e-mail: langanke@uni-greifswald.de

W. Liedtke
Faculty of Theology, Systematic Theology, Ernst Moritz Arndt University of Greifswald, Greifswald, Germany
e-mail: wenke.liedtke@uni-greifswald.de

A. Buyx (✉)
Institute of Experimental Medicine, Christian Albrechts University of Kiel, Kiel, Germany
e-mail: medizinethik@iem.uni-kiel.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_22

Abstract

Personal responsibility for health has been a topic of debate and analysis for decades. Most publications on the topic are devoted to the ethical and regulatory implications of holding individuals responsible for their own health status. This chapter introduces and discusses a philosophical definition of personal responsibility for health, as part of a logical and semantical analysis of the concept. Responsibility is defined as a relation among six variables U, V, W, X, Y, and Z: *U is responsible to V for W to an extent of X and with regards to the time frame Y because of certain normative standards Z*. Each of the variables is explained and discussed throughout the chapter. The chapter closes by drawing out the ethical and political implications of such a philosophical model of patients' responsibility for their own health.

Introduction

Personal – or individual – responsibility for health has been a topic of debate and analysis for several decades now. Most of the publications on the topic are devoted to the ethical and regulatory approaches to, arguments for and against, and implications of, holding individuals, be they patients or healthy citizens, responsible for their own health status. At least since the later 1970s, influential authors from various fields, including practical medicine, health policy, philosophy, and epidemiology, have analyzed whether and how such responsibility could, and indeed should, be ascribed (see for example Wikler 1978a, b; Crawford 1977; Watkin 1978; Veatch 1980; Knowles 1977). Today, the literature analyzing and assessing the ethics and practice of personal responsibility for health is so wide-ranging that any attempt to select only a few citations for an introduction is nigh impossible.

By contrast, publications on individual or personal responsibility for health from a perspective of philosophy of medicine are few and far between. In most writings on philosophy of medicine, the term of responsibility does not play any role at all. And even a volume such as George Agich's "Responsibility in Health Care" (1982), which came out in the Springer Series Philosophy and Medicine, is devoted mostly to the legal, political, ethical, social, and cultural aspects and implications of responsibility (Agich 1982, p. 5). Books specifically on the philosophy of medicine – itself arguably a small field dwarfed by the ever burgeoning area of biomedical ethics – rarely devote dedicated chapters to personal responsibility for health (see, however, Maier and Shibles 2010, Chap. 12; Engelhardt and Jotterand 2008, Chap. 10; Cherry 1999, Chap. 6). Personal responsibility also does not play a significant role in published curricula of philosophy of medicine (e.g., Rudnick 2004).

This does not mean that philosophers of medicine have not written about personal responsibility of health; indeed, in this article, several such works are referenced. However, it does mean that overall, patients' responsibility has, at least

so far, not been a topic of central or even sustained attention in the philosophy of medicine, unlike, as mentioned, in biomedical ethics. This chapter therefore aims to summarize the work that has so far been done on patients' responsibility for health. In addition to this, it also actively extends the discussions on personal responsibility by introducing, explicating, and ultimately applying a philosophical definition of personal responsibility for health. This is done as part of a logical and semantical analysis of the concept of responsibility (section "The Concept of Responsibility: Logical and Semantical Analysis"). Following this, a number of general considerations regarding personal responsibility for health are presented and discussed in section "Personal Responsibility for Health: General Considerations." Co-responsibility for health is discussed in section "Co-Responsibility for Health." The chapter closes by drawing out the ethical and political implications of such a philosophical model of patients' responsibility for their own health in section "Ethical and Political Implications."

"The Concept of Responsibility": Logical and Semantical Analysis

A variety of meanings are associated with the term "responsibility" (cf. Werner 2011, 2013, 2016). Each depends on the context, for instance, "responsibility" as "causal responsibility" (for this concept cf. Feinberg 1977; Putnam 1982) or as a "normative relation." Regarding the context of patients' responsibility for their own health, the second usage of the term is of particular interest. In this usage, responsibility refers to *the demand on a person or an institution to justify its action or actions towards another person or institution*. Often, this happens because that person seeks to receive a service or financial compensation, or not to lose certain entitlements. It can also happen outside of any regulatory environments, for example, when individual responsibility is demanded by particular religious or moral teachings (for the general discussion of responsibility in biomedicine see, for example, Buyx 2008; Schicktanz and Schweda 2012; for legal aspects, for example, Krpic-Mocilar 2003; various ethical and political aspects and implications are discussed, e.g., by Minkler (1999), Steinbrook (2006), Brownell et al. (2010), Wikler (2002), Rohr and Schade (2000), Yoder (2002), Cappelen and Norheim (2005), Pearson and Lieber (2009), Schmidt (2008, 2009a, b), Buyx and Prainsack (2012), Eyal (2013), Brown (2013), Bærøe and Cappelen (2015), Resnik (2014), Nielsen and Andersen (2014), Wiley et al. (2013), Fleck (2012), Lewis and Rosenthal (2011), Yang and Nichols (2011), Bringedal and Feiring (2011)).

It is worth noting that this requirement for justification emerges only in settings where the compliance of the person or institution with certain rules or requirements is in question (for the philosophical discussion of the responsibility of supraindividual entities cf. French and Wettstein 2006); if a person is in compliance with a relevant normative setting and the setting itself is not in question, no need for justification arises.

Defining Responsibility

This general understanding of responsibility can be sharpened and specified by constructing a formula, according to which responsibility is a relation among six variables U, V, W, X, Y, and Z:

> U is responsible to V for W to an extent of X and with regards to the time frame Y because of certain normative standards Z.

This formula, further developed from earlier work in Langanke and Fischer (2012) and Langanke et al. (2012, 2013), takes into account both the results of a predominately German, long-term discourse on the normative function of the relational concept of responsibility, as well as different contributions from Anglo-American philosophy that both have ancient roots in Aristotelian Ethics (Aristotle 2011). In 1919, Max Weber prominently introduced the concept of responsibility into academic and public debate (Weber 1919). After the Second World War, the German debate on responsibility as a relational normative category was mainly continued by contributions from the ethics of technology. The discussion then markedly changed its character following the widely received, but highly controversial, publication of an "ethics of responsibility" by Hans Jonas (1979). Influenced, among others, by the philosophy of language and the German tradition of Discourse Ethics, authors from the field of ethics of technology, such as Lenk and Maring (1991), Ropohl (1993), Ott (1997), and Grunwald (1999) argued for a relational, logically clarified understanding of "responsibility." They proposed different relational logical formulas and schemata, akin to the formula presented above. Reconstructions of responsibility range from three-digit formulas to six-digit relations (e.g., Ropohl 1993). Most suggestions are based on four-digit formulas, covering the variables U, V, W, and Z of the formula above (e.g., Hillerbrandt 2006; Werner 2011, 2013, 2016). The introduction of a relational logical understanding into the discourse of medical ethics and its application to the problem of patients' responsibility of health was prominently proposed, for example, by Marckmann et al. (2004).

The six-digit formula introduced here includes "extent" and "time frame" of responsibility as relevant relational aspects. In contrast to the six-dimensional concept published by Ropohl (1993), the aspect of time is not applied to a retrospective or prospective understanding of "responsibility" in the formula here, but instead to the temporal dimension of the preconditions of accountability (see below). Furthermore, Ropohl's dimension of "results" or "consequences" of acts is covered in our formula by W, based on the decision to interpret acts and their results as one single aspect. Instead it includes the implementation of the "extent" of responsibility as a separate sixth dimension or relational aspect.

Within the equation proposed in this chapter, U and V are placeholders for single persons or institutions (for institutional responsibility cf. French and Wettstein 2006). W stands for certain *results* of an action or an omission, but often the respective action or omission itself is inserted into the position of W. X defines to

Table 1 The relation of responsibility – variables, terms, and suitable insertions (Modified from Langanke and Fischer 2012; Langanke et al. 2012, 2013)

	Variable	Term	Suitable insertions
1	U	Subject of responsibility	Persons or institutions
2	V	Authority of responsibility	Persons or institutions
3	W	Object of responsibility	(Results of) actions or omissions
4	X	Extent of responsibility	Specifications like "fully" or "partially" or percentage
5	Y	Time frame	Time intervals
6	Z	Normative standard	Rules, regulations, prohibitions, permissions, or systems of those norms

which extent a person or institution U is responsible for W, for example, whether U is *partially* or *fully* responsible for W. Y determines the period of time for or during which U is responsible to V. Z stands for particular norms, including rules, regulations, commandments, prohibitions, permissions, etc., or even larger normative systems (cf. Table 1).

Intentionality

In this definition of responsibility, only results of *actions* or *omissions* may have to be justified (variable W in the formula). In line with broad intuitions, common sense, and indeed legal understandings of responsibility, it does not make sense to demand the assumption of responsibility for a state of affairs that cannot be influenced by a person or institution, e.g., because this state of affairs cannot be modified *intentionally*. The concept of responsibility is therefore strongly connected with the concepts of intentionality and susceptibility to intentional modifications, affecting the extent of responsibility (variable X in the formula). In this, the intention of an action or omission is to be understood as the actor's objective or, more precisely, the purpose of her act. Because of this character of intentionality, actions and omissions differ from other human behaviors that are not subject to intentional and purposeful control by the actor. Such involuntary behaviors are, at least from a theory of action as well as an ethical perspective, no, or only marginal, cases of objects of responsibility (for the following cf. Fischer and Ravizza 1998).

Intentionality can only become effective if particular actions or omissions are connected with the purpose of the action in such a way that through them, the purpose can be achieved *reliably*. This means that it is likely, normal or even inevitable that it can be achieved in this way. This necessary connection between purpose and act can in turn be mediated by causal relations that are initiated by a person through certain actions or omissions, when the person is taking intentional advantage of them for her purpose. Hence, an intended result or state of affairs for which factors – including causal relations – guarantee or reliably allow for its achievement can be considered "intentionally modifiable."

For some intended states of affairs, none, or not all, factors are known that can contribute to bringing them about. Such states of affairs are under limited, or even no, control of an acting individual. Therefore, and this is a crucial point, they cannot be wittingly and willingly brought into existence. It follows from this that all states of affairs that cannot be intentionally modified in the sense described cannot correctly be criticized or sanctioned by referring to them in terms of responsibility (cf. critical Frankfurt 1969; in line with the concept presented here Fischer and Ravizza 1998). "Shall," as it is said traditionally, implies "can." In other words, intentionally not modifiable states are no candidates for the attribution of responsibility regarding an action or omission of an individual. The latter cannot be expected or requested, and no rule can be formulated that would classify a certain act or omission as violating some normative standard with regard to that state of affairs. Instead, in cases of this type, the circumstances responsible for the deviation from the desired target have to be considered – such as, for example, the genetic causes of a particular monogenetic disease, that, as a deviation from the desired state of health, simply befalls someone (for the discussion of the term "to befall" cf. Kamlah 1972; Marx 2010).

It also follows from this how to consider states of affairs that might be partially modifiable: The smaller the degree of intentional modification possible, and the bigger the uncertainty and/or lack of knowledge regarding causal factors, the less it is correct to discuss a relevant act or omission in terms of (lacking) responsibility.

Implications

The terminological decisions described so far have several consequences that are pertinent to the discussion about patients' responsibility for their health (for the following Fischer and Ravizza 1998; Marckmann et al. 2004; Werner 2011, 2013, 2016):

(a) An individual or institution is only responsible for an action or omission with respect to the results of that action or omission. This is the case if an event or situation can be understood as being caused or generated by the action or omission and if the event or situation can be judged to be desirable or undesirable by virtue of applying the rules, regulations, values, etc., governing in a particular setting.
(b) If it is uncertain which of two or more potential acts causes a desirable or undesirable situation, it will necessarily be problematic to claim or attribute any responsibility regarding either act.
(c) Responsibility concerning a certain action can be understood prospectively as well as retrospectively (cf. Zimmerman 2001). Therefore, it is possible to distinguish between "responsibility of competence" and "responsibility of accountability":
   – Responsibility of competence (prospective): An individual or an institution bears responsibility for something which must be executed in the future (for problems of prospective responsibility cf. Hart 1949).

– Responsibility of accountability (retrospective): An individual or institution is held responsible for the results or consequences that have occurred as a result of an action or omission of that individual or institution in the past.

In light of this distinction, we can specify the logical relationship between responsibility of competence and of accountability as follows:

> Only if someone, on the basis of normative standards, bears prospective responsibility for a situation this person can be held retrospectively accountable. (Marckmann et al. 2004, p. 716)

Preconditions

From these terminological considerations it further follows that there are certain requirements that have to be fulfilled when claiming responsibility or attributing it to a particular person or institution (for the following cf. Fischer and Ravizza 1998; Langanke and Fischer 2012; Langanke et al. 2012, 2013):

1. Voluntariness: Pressure and coercion decrease the degree to which a person can intentionally modify a particular state of affairs. An action by a person or institution needs to be performed with no coercion and at most minimum external pressure in order for responsibility to be attributed both prospectively and retrospectively to this person. In cases of undue pressure or coercion, depending on its degree, responsibility is shifted fully or partially from the coerced person towards the coercing individual or institution (this criterion was already presented and discussed in Aristotle's Nicomachean Ethics, Book III, cf. Aristotle 2011).
2. Availability of alternative options: If someone is to be held accountable for an action, a feasible alternative action needs to have been available at the time of acting. Responsibility reaches its logical limit if the possibility to act otherwise neither existed, nor exists, and it is diminished in proportion with decreasing options to act.
3. Level of information: To hold an agent accountable for an action or actions, prospectively as well as retrospectively, that agent needs to have had reasonable knowledge regarding possible results of both the action and the alternative options or at least needs to have been in a position to have this knowledge in principle (cf. Marckmann et al. 2004). This would be the case if the knowledge that a certain action or omission leads to an undesirable result can be reasonably regarded as part of common knowledge (obviously, some pragmatic difficulties can arise from exactly determining "common knowledge" and "reasonable" itself).
4. Self-determination and accountability: The ability of persons to be self-determining – that is, to be autonomous – when considering and performing certain actions or omissions is a transcendental precondition of responsibility.

This simply follows from the terminological decision that persons can be, at most, responsible for their *actions* – or, more precisely, for the *results* of their actions (cf. Fischer and Ravizza 1998). As suggested above, states of affairs or situations that cannot be brought about or at least modified intentionally are not potential objects of responsibility. In legal and other practical contexts, the possibility that human beings can lose and regain their ability for self-determination regarding certain actions or omissions is taken into account. Consequently, the responsibility of a person for certain states of affairs can be limited (always, or temporarily) in a fundamental sense.

Limitations of Responsibility

It is indeed important to note that responsibility can be full or limited. In view of the thematic focus of this chapter, two types of limitations in particular require a more detailed investigation:

1. Extent of responsibility: Limitations with respect to the extent of responsibility can result in at least two different cases:
   - A particular state of affairs or situation cannot be understood as the result of a single person's or institution's actions or omissions, but as the result of the cooperation or at least coinfluence of multiple agents. In this case, the responsibility for the respective state or situation is typically distributed among the different persons or institutions involved and depends on the extent of their respective involvement(s).
   - If a state of affairs or situation is only partially modifiable through intentional acts, a person or institution can, at most, be held co-responsible for it.
2. Temporal limitations: A person's or institution's responsibility for a certain state of affairs or situation can be temporally restricted. All four aspects mentioned in the section *Preconditions* – that is, voluntariness, availability of alternatives, level of information, and self-determination – as well as the extent of responsibility have, or can have, a temporal dimension. It follows that a person or institution can be retrospectively held (co-)responsible for a certain situation with respect to the time interval $[t_1; t_2]$ in which, e.g., the condition of voluntariness was fulfilled, but not for a later interval $[t_3; t_4]$, in which relevant circumstances changed fundamentally and she was, e.g., under duress.

Having defined responsibility in this way, it can be assessed how responsibility and, more specifically, individual or personal responsibility should be defined within the general context of medical care. In this chapter, personal responsibility for health is taken to mean that *one's health is the object of one's own responsibility*. Alternatively, individual responsibility could be interpreted as a responsibility in which *the subject and the authority of responsibility are the same*. In the following paragraphs, this alternative interpretation will not be the prominent one. The debates about personal responsibility for health are mostly framed in the way that a person, e.g.,

practices an unhealthy lifestyle and is therefore responsible *to others* for her diminished health, or, broader, for her way of life. The alternative interpretation of personal responsibility as mentioned above does not fit such a concept. (It will be shown below that subject and authority of responsibility can indeed coincide.)

Personal Responsibility for Health: General Considerations

The results of the logical and semantical analysis in the preceding section can be applied to the context of both healthy individuals and patients. This section discusses three major premises that often underlie the assumptions that individuals themselves are the *subjects of responsibility* in the way introduced above. These are not always made explicit, although they are very pertinent to the debate. All three have already been mentioned above, as part of the formula: (1) the susceptibility of health to intentional influences, such as health behavior including eating, exercise, etc. (affecting extent of responsibility, X in the formula), (2) the existence of a normative standard, such as regulatory demands, that allow for the attribution of responsibility in the first place (Y in the formula), and (3) the existence of an authority of responsibility (Z in the formula), which decides on assertions of responsibility. In other words, for a person to be responsible for her state of health that state had to have been intentionally modifiable; there had to have been a clear normative standard in place that expected or demanded responsibility for health, or certain behaviors regarded as responsible health behaviors; and there had to have been a person or, more likely, an institution the person was responsible to. The first premise – intentional modifiability – has already been unpacked regarding responsibility in general in section "The Concept of Responsibility: Logical and Semantical Analysis." In the following, it will be analyzed in the particular context of health, as will the other two. Even if the three premises are fulfilled, questions of the extent of responsibility and temporal limitations remain to be answered; they will be examined in section "Co-Responsibility for Health."

Diseases as Potential Objects of Responsibility

Following the logical and semantical analysis on the links between action and responsibility, the personal health status of an individual may be an object of responsibility if, and only if, it can be modified through actions or omissions of patients or healthy individuals (cf. Rohr and Schade 2000; Yoder 2002; Schmidt 2008, 2009a; Pearson and Lieber 2009). However, it has already been described that in many cases, a state of affair cannot be clearly or fully ascribed to the person in question in this way, for several reasons. In such cases, the person would be at most co-responsible for them. In the context of illness and disease, therefore, we can specify that some conditions are potential candidates for being objects of responsibility or co-responsibility (for the definition of the term "co-responsibility" cf. Schmidt 2008, 2009a), because their emergence, development, or course is

considered to be susceptible to personal choices and individual actions (cf. Langanke and Fischer; 2012; Langanke et al. 2012, 2013) (Note that this does not mean that in each case, the person would be responsible for that health status. In order to come to that conclusion, other elements – such as the relevant normative standard – have to be considered as well, see below):

(a) This includes diseases that can be prevented, entirely, largely, or to some degree, through *protective or health-promoting behaviors* such as appropriate physical activity or healthy nutrition.
(b) Furthermore, diseases which are directly related to *chosen risky behaviors* are potential objects of (co-)responsibility.
(c) Under the additional premise that the respective agents have access to healthcare services, diseases which can be avoided by making use of *preventive medical interventions* like vaccinations or preventive surgery could be discussed as potential objects of persons' (co-)responsibility for their own health.
(d) Assuming the same premise of access to relevant healthcare resources, diseases that can be modified in their course and severity by *screenings and preventive health assessments* could also be regarded as objects of patients' (co-) responsibility, even though they are not affected by lifestyle, health behaviors, or preventive medical interventions (cf. Dabrock 2006).

In general, individual responsibility for health could be construed prospectively and retrospectively (cf. Marckmann et al. 2004). If, and only if, all premises discussed above are met, a patient who has developed, e.g., a disease as a direct result of unhealthy behaviors could be held responsible retrospectively, once treatment for that disease required medical resources. Real-world examples of being held responsible retrospectively would be paying higher premiums for insurance, higher co-payments, or exclusion from some services (cf. Pearson and Lieber 2009). (Note that this does not include a judgment whether holding the person responsible would be ethically and/or politically appropriate.) Likewise, if, and only if, premises were met, an individual could be held responsible prospectively for a particular behavior, if that behavior lead to the development of diseases that would require treatments. Real-world examples would be sanctions if a person did not keep up with certain prospective requirements of health promotion (e.g., regular visits to the gym, keeping to a certain weight, etc.) or for being noncompliant in other ways.

Since, as explained in section "The Concept of Responsibility: Logical and Semantical Analysis," persons can be held responsible only if they could have chosen otherwise, prospective responsibility is indelibly linked to access to relevant information – in the healthcare context, as well as in general. Only a person who is informed about and aware of health risks that result from a certain behavior, or from refusing healthcare interventions, can *choose* not to engage in the relevant behavior, or decide differently. In other words: Only if patients or healthy individuals have access to adequate information about a health risk can they become subjects of responsibility regarding that health risk, in the sense that they are accountable for failing to avoid risky behavior or take relevant preventive measures if these, indeed,

exist (cf. Paul 2010). In the context of this section, the criterion of a sufficient level of information can be reconstructed as a premise under which diseases that belong to one of the categories, (a) to (d), are candidates for being objects of responsibility.

It is obvious from the analysis so far that the potential for assigning personal responsibility for health does not include *all* diseases (cf. Düngen 2009; Resnik 2014). Since some diseases cannot, according to current knowledge, be modified through actions or omissions of the individual, they must be treated as unfortunate situations that occur through no action or omission of persons, i.e., which befall individuals (cf. Kamlah 1972; Marx 2010). Here are a few clear-cut cases:

(a) Conditions of this type include those with *unknown etiology*. In cases in which it is unknown which out of two potential acts causes a desirable or undesirable situation, it is problematic to claim or attribute responsibility with respect to either act. All diseases with unknown or partially known etiology are therefore excluded from being the object of responsibility. It is important to stress that this holds even where most – or even all – etiological factors of a multifactorial condition have been described, but it is unclear which factors have contributed to the illness *in each individual case*. If it cannot be determined clearly what the etiology in an individual case has been, retrospective responsibility cannot be attributed for the reasons mentioned. This is the case in many chronic, so-called "lifestyle-related" illnesses, including diabetes, coronary heart disease, back pain, and depression.
(b) Diseases that are determined *entirely through genetic causes* are no potential objects of patients' responsibility. This includes genetic disorders (cf. Düngen 2009; Resnik 2014) such as single-gene disorders, autosomal and x-linked illnesses, as well as forms of trisomy, etc.
(c) It also seems to be illegitimate to attribute responsibility for diseases that are caused by *living conditions*, such as environmental or working conditions which cannot be directly and/or immediately influenced or changed by the concerned people. Illnesses of this kind include, e.g., lead poisoning from drinking contaminated tap water or occupational illnesses such as hearing loss in certain professions.

These diseases clearly do not fit into the category of illnesses that can be objects of individual responsibility for health, because they cannot be intentionally modified. Another group of illnesses are difficult to assess because they are caused partially or entirely by behaviors and actions that are, at the same time, risky and health-promoting. Many forms of sports and exercise belong into this group of *ambivalent behaviors*. Skiing, for example, is an outdoor activity that may have positive effects on the physical and psychological health of an individual. On the other hand, every winter season many skiing accidents happen which often require surgery. Do individuals who practice skiing as a winter sport behave responsibly regarding their health, by being athletic and active outside, training their muscles, and improving their cardiovascular fitness? Or are they responsible for their broken bones, because they have engaged in behaviors that put them in harm's way?

Summing up the observations in this section, it becomes apparent that a significant number of disease entities are not suitable as potential objects of responsibility. In addition, there are also lifestyle and behavioral factors, such as having certain professions or choosing certain forms of exercise that can contribute to illnesses, because they have positive as well as negative influences on health. Many of these, in turn, are affected by social determinants such as education and income, which makes the attribution of individual responsibility for related illnesses even more complicated. (Note that the debate around the social determinant of health is very extensive and cannot be summarized here; a good overview can be found in, e.g., Venkatapuram (2011).)

Normative Standards

Another component of the attribution of responsibility must be discussed. According to the understanding of responsibility presented in section "The Concept of Responsibility: Logical and Semantical Analysis," responsibility cannot be attributed if there is no relevant normative standard; that is, *a particular set of values within a particular context, providing the normative background for assessment* (cf. Ott 1997; Grunwald 1999; Werner 2011, 2013, 2017). This standard allows classifying certain acts or omissions as desirable or undesirable. Only on the basis of its rules, regulations, prohibitions, commandments, permissions, etc., do certain behaviors become objects of judgments. Hence, if we attribute responsibility to a person in the full sense of the formula introduced in section "The Concept of Responsibility: Logical and Semantical Analysis," we render a value judgment on this person's actions or omissions in accordance with the normative standard.

In the discourse on patients' responsibility for their own health, this connection between normative standards and judgment is regularly ignored or not made explicit enough. In consequence, the normative standard for sanctioning certain health-related behaviors is often not made transparent.

However, the normative standard is of significant importance for any discussion of responsibility for health. For example, national health care systems vary greatly. In countries with full, or partial, publicly funded health care systems, the question of the relevant normative standard can have a regulatory or quasi-regulatory dimension.

There are various terms in use to denote that a health care system is paid for from common resources, such as tax, income, or insurance contributions, instead of private out-of-pocket payments, and administered through state-controlled or even state-run actors, instead of industry and private enterprises. This article follows established usage in that "publicly funded health care system" (and short, "public health care system") is used to describe a system of the first kind. Obviously, systems vary in many details that cannot be discussed here.

Responsibility may be ascribed, e.g., against the background of a particular legal framework that includes rules and regulations on responsibility and health behaviors. In countries without a publicly funded health care system, the ascription of responsibility will vary greatly, depending on the relevant normative standard, e.g.,

as stated in private health insurance policies (cf. Pearson and Lieber 2009). Where all health care is organized privately, individuals' responsibility for their health will, at least in practice, be delineated to a significant degree by their ability to pay.

It is important to note that in addition to regulatory normative standards, relevant normative standards can also be part of the private morality of a person. Given, e.g., a family depending on one family member's working and earning abilities, this person can attribute to herself the responsibility for her own health, because all other family members may depend on her. Irrespective of any regulatory frameworks, in this scenario, the normative standard is a more or less private principle of family care, which generates a purely moral obligation. Accordingly, in such a scenario, the subject and the authority of responsibility can (but obviously do not have to) coincide.

Such purely moral obligations and private principles are also relevant in countries with a fully or partially publicly funded health care system. In such countries, however, private moral principles are not the sole normative standards that can be applied to a person's health related behavior. Germany, for example, with its mixed public/private health care system, is an illustrative example of this. In Germany, there is a legal obligation for almost every resident to have health insurance. Only civil servants, those who are self-employed, or earn above a certain threshold, can choose to have private insurance; everyone else is automatically part of public statutory insurance, with a (limited) choice among various sickness funds. This includes the unemployed, retired, or those not able to work. Statutory insurance is explicitly based on the principle of solidarity (cf. German Social Security Code [Sozialgesetzbuch] V, § 1). The contribution every member of a statutory sickness funds has to pay does not depend on individual risks, but typically on the level of income, and is subtracted directly from each person's pay. Family members such as children and spouses that do not work are automatically insured in the earning person's sickness fund. For those who cannot contribute, insurance is covered by social care.

In a system of this type, the overall community of members of statutory sickness funds shares an interest in ensuring that the fund exercises thrift in the use of available resources and that preventable costs are avoided (for the following cf. Werner 2006). Indeed, this is explicitly part of the legal framework regulating statutory insurance and applies not only to the administration of the fund itself, but also to the reimbursement of expenses.

The German normative standard appears to include, at least in principle, the option that, e. g., the costs of treatments originating from illnesses that are considered avoidable are regarded as the personal responsibility of the individual who did not avoid them. Note, however, that the criteria developed above regarding modifiability, alternative options, knowledge of causality, etc., all apply; that is, the question would arise whether the respective diseases are fully or partially individually caused and/or modifiable, and whether the individual actually had the option to avoid them. If these criteria were fulfilled, it would be appropriate to discuss whether the reimbursement of costs resulting from such diseases could be withheld or only

partial reimbursement provided, etc. (Potential practical, ethical, and political implications are discussed further below).

In the German case, the relevant normative standard could be reconstructed as the abovementioned aim to avoid the misuse of resources – if resources are understood as misused, when they are spent on treatments for avoidable, individually caused diseases. This can be rephrased in the following rule (cf. Langanke and Fischer 2012; Langanke et al. 2012, 2013):

> All members of a public health-care system that carries costs for all its members are obliged to prevent costs of treatment for avoidable illnesses, in order to use resources as effectively as possible and to avoid misuse.

In Germany, a normative standard of thrift has indeed quasi-regulatory power. It is tied very directly to personal responsibility; the fifth book of the German Social Security Code claims in its preamble explicitly that all members of the statutory health insurance are co-responsible for their own health and that they are therefore obliged to strive towards preventing illness (cf. German Social Security Code V, § 1).

The example shows that responsibility for health cannot be attributed to an individual person without revealing the relevant normative standards. These standards have to be made explicit. If they are not clearly defined and/or made explicit, the concept of individuals' responsibility for their own health can easily be manipulated for political or other goals in public debates. Obviously, the justifications of the relevant normative standards can be questioned. The outcomes of such debates will in turn alter whether personal responsibility can (still) be attributed within a given system or context (see below).

Authority of Responsibility

As noted above, responsibility depends on some authority (V in the formula). Without authority, it would be difficult to ensure individuals behaved in line with their responsibilities. But what kind of authority applies to the context of health care? To answer this question, it is helpful to return to the distinction between countries with a publicly funded, or partially publicly funded, health care system, and those without such a system.

Where there is no public system, or no obligation at all to participate in the public elements of a mixed system, a wide range of authorities may come into play, from individual conscience to societal institutions like churches, from private insurance companies to metaphysical entities like a god or gods. But in these cases, the binding power of these authorities results from the individual decision to acknowledge or join them.

In cases where a public system does exist, the question of the authority of responsibility has a public, or even political, dimension, particularly if individual responsibility is codified in regulation, such as, e.g., in Germany. Nevertheless,

which institution wields this authority is often not clearly addressed in relevant official statements even where explicit normative standards of personal (co-) responsibility exist. For instance, the fifth book of the German Social Security Code does only provide an implicit answer to this question (cf. German Social Security Code V, § 1): The publicly funded statutory health insurance (through sickness funds), which is representing the community of those insured, is functioning as the relevant authority of responsibility.

At this point, the value of a relational logical reconstruction of the concept of "responsibility" becomes clearly apparent (for the following cf. Werner 2006). The reconstruction, as developed in the formula introduced in section "The Concept of Responsibility: Logical and Semantical Analysis," allows for the examination of the authority of responsibility, and this in turn can lead to follow-up questions of significant ethical and political importance at least in countries where the concept of individuals' responsibility for their own health is charged with regulatory or quasi-regulatory power. Whenever personal responsibility for health plays a role within a health care system, the question of the authority for responsibility becomes highly pertinent, both for ethical and political reasons.

Co-Responsibility for Health

It follows from the discussion in the preceding sections that in most cases, individuals' responsibility for their own health can at best be conceptualized as joint or co-responsibility.

Most health risks are the result of combinations of factors that can be controlled by individuals, such as health behaviors and the utilization of preventive measures on the one hand, and those that are beyond the control of individuals, such as genetic and environmental factors on the other. There are very few phenomena in human health and illness that can be attributed in a straightforward way to a single and, moreover, intentionally modifiable cause and only a few more that can be attributed to multiple factors that can all be controlled intentionally. Hence, it is only possible to connect patients' responsibility for their health to those elements of risk that are regarded to be susceptible to personal choices.

The situation gets even more complex because most preventive strategies or healthy behaviors only have a, statistically speaking, limited or even minimal influence on the prevalence and/or course of those diseases that are considered to be modifiable by prevention or healthy behaviors. It was argued above that when it is uncertain which out of two potential acts causes a desirable or undesirable situation, it is problematic to attribute responsibility for either act. In such cases, even a co-responsibility of individuals for their health can be doubted, regarding the compliance or noncompliance with preventive strategies or lifestyle-related suggestions (cf. Schmidt 2008, 2009a).

Other problems with the concept of responsibility for health arise from the fact that the *accountability* of persons can be narrowed by biological and therefore medically relevant reasons. The phenomenon of addiction is a good example to

demonstrate how a disease can reduce someone's accountability to the extent that it diminishes the possibility to attribute responsibility to that person. In cases of addiction, a person's ability for self-determination is lost or at least temporally limited. For example, while a person might be held responsible for her alcohol consumption at nontoxic, nonaddictive levels, once alcohol consumption tips into addictive behavior, her control is significantly diminished. Therefore, it has to be doubted whether this person is responsible for her actions, or the results of her actions, at least in times when her alcohol consumption corresponds with addictive behavior patterns. At this point, the advantage of the integration of the time frame Y into the relational-logical formula of responsibility in paragraph 2 becomes obvious: Addiction is generally seen as a medical condition, which has as one important feature that someone's co-responsibility for his or her health can be temporally, or permanently, limited. It should be noted that addictions are conditions with very complex etiology and progression. Behavior that would qualify as full blown addiction in one person might yet be under full voluntary control in another; and depending on family history, genetics, environment, and a host of other factors, some people might find it easier than others to retain or regain control over their substance abuse. This obviously does not make the ascription of responsibility any easier – on the contrary.

The dimension of time also plays an important limiting role within the context of pediatrics. Depending on their age, children are not or not yet fully accountable for their health-related behaviors. In consequence, the concept of individuals' responsibility for their health has logical limits in pediatrics. Age-related limits to responsibility for health are particularly important, because much of health-related behavior, including eating behavior, level of physical activity, psychological coping mechanisms, level of impulse control, and many others, are learned – some would say ingrained – in early (and later) childhood (e.g., Brown and Roberts 2011). This in turn affects the level of accountability for such behaviors in adulthood.

Another complicating factor needs to be exemplified: it was established above that responsibility is diminished in proportion with decreasing options to act. In other words, a person who has many unconstrained options can, all things being equal, be held responsible or co-responsible, as opposed to a person who does not have these choices open to her. This condition of alternative options for action is very relevant to the prominent debates concerning the social determinants of health (cf. Venkatapuram 2011; and many others). If individuals live in so-called "obesogenic environments" of the kind that do not allow for a consistently healthy lifestyle (Egger and Swinburn 1997) – for example, if individuals do not have reasonably easy access to healthy food options and/or their built or work environments preclude physical activity – then an attribution of responsibility becomes less appropriate. Indeed, it becomes less appropriate the fewer alternative options individuals in such environments have to make healthy choices.

Finally, personal responsibility for health can be constrained through the influence of other agents, such as healthcare providers, for example, regarding the recovery from illnesses. It is necessary for a layperson to trust that whatever measures chosen and executed by their healthcare providers will contribute to a return to health with, ideally, minimal adverse effects. In such cases, the patient is only responsible for those adverse effects of medical treatment that he or she could have deliberately modified or avoided. In such a scenario, the individual and medical staff share co-responsibility for the patient's health, again without clarity on how much each contributed to a given health status.

In sum, based on a consideration of these factors alone, it is very difficult to assign responsibility – even co-responsibility – to patients for their own health in a justified way, if all conditions of the definition introduced in section "The Concept of Responsibility: Logical and Semantical Analysis" are taken into account. The final section of this article will draw out a few additional problems and challenges of an ethical and political nature.

Ethical and Political Implications

In this chapter, through the philosophical analysis and discussion in the previous sections, it has become obvious that there is very limited scope to assign personal responsibility for health to individuals in a legitimate way. Moreover, even if there were cases that corresponded with all conditions set out above, other considerations come into play. If the ascription of personal responsibility for health has any practical consequences, for example, by being sanctioned via malus systems in health insurance; higher co-payments; or the exclusion from certain treatments or by being incentivized, for example, through bonus systems, etc., these consequences can be analyzed for their ethical and political implications (cf. Apel 1988; Bayertz 1995; Lenk and Maring 2003). (It is often difficult to clearly distinguish between ethical and political implications in this context, since they are regularly intertwined, see below). The literature on this analysis is, as mentioned in the introduction, significant (see references in the section "Introduction"). This being a chapter on the philosophy of medicine approach to examining personal responsibility for health, the following is a very short summary of the most important ethical and political issues.

Firstly, and following on directly from the formula introduced in section "The Concept of Responsibility: Logical and Semantical Analysis," both the authority of responsibility (V in the formula) and the normative standard applied (Z in the formula) can be controversial (if, as mentioned above, the subject of responsibility and the authority are not one and the same). The political legitimacy of a given authority of responsibility – which, in real life and as mentioned, sometimes is not even clearly specified – can be contested, particularly if it is empowered to sanction

undesired and incentivize desired health behavior. The same holds for the normative standard, if not more so. A normative standard has to be justified, and recent cases have shown that where new normative standards regarding personal responsibility for health have been introduced, these have often lacked sufficient justification (Steinbrook 2006; Bishop and Brodkey 2006; Prainsack and Buyx 2011, 2012). To return to the example from section "Personal Responsibility for Health: General Considerations": While the German Social Security Code that requires co-responsibility from all members of statutory insurance has legitimacy by law, it could be challenged on ethical grounds. For example, it could be argued that if it was applied in a strict interpretation (which it currently is not), it could lead to the exclusion of individuals from care who could face stigmatization as a consequence. An (currently hypothetical) example would be the "reckless," fully informed, unconstrained in her decision-making, skier, who suffered a complicated hip fracture in a skiing accident and could not fund optimal care for the fracture resulting in a highly visible gait defect. Others would describe this as an unethical example of "victim blaming" (Crawford 1977) or as an example of state actors evading their duties of care towards their citizens (Minkler 1999; Schmidt 2009b). Some would argue that any exclusion from medical treatment, or any financial penalty in health care based on behavior instead of need, even if fully responsible according to the conditions set out above, was unethical, based on the wrong values and principles (Wikler 2002). Finally, even those who would argue that the normative standard demanding responsibility for health in the German case was based on the value of solidarity have been challenged. The understanding of solidarity as reciprocity, as enshrined in the Social Security Code (cf. Buyx 2008), has been shown to be at least controversial (cf. Buyx and Prainsack 2012; Prainsack and Buyx 2017).

In addition to challenges to the authority and normative standard of personal responsibility for health, it could also be argued that personal responsibility for health contradicts values that are relevant in other spheres of life. For example, while thrift might be a key value in statutory health insurance, in other areas of life, abundance, joy, indulgence, etc., are important values, and these could come into conflict with thrift. Most obviously, the value of freedom and the related principle of personal autonomy over one's private actions and behaviors, supremely important in modern pluralistic societies, are seen to conflict with incentivizing and sanctioning personal responsibility for health (Wiley et al. 2013; for many others).

And finally, even if all these implications are not deemed relevant, complex problems regarding the practical implication of incentives and sanctions remain: for example, which behaviors exactly would be incentivized/sanctioned, and based on which criteria, how behaviors would be monitored, and what kind of investigative powers authorities would have (Langanke and Fischer 2012; Langanke et al. 2012, 2013; for many others). Again, these issues highlight that an ascription of personal responsibility, if it becomes actionable, may conflict very directly with personal rights and liberties usually regarded as fundamental, both from an ethical as well as a political perspective.

Definitions of Key Terms

Responsibility	Responsibility is defined as a relation among six variables U, V, W, X, Y, and Z: U is responsible to V for W to an extent of X and with regards to the time frame Y because of certain normative standards Z.
Personal responsibility for health	Personal responsibility for health is taken mainly to mean that one's health is the object of one's own responsibility.

Summary Points

- Personal responsibility for health has been a topic of debate and analysis for decades.
- Most publications on the topic are devoted to the ethical and regulatory implications of holding individuals responsible for their own health status.
- This chapter introduces and discusses a philosophical definition of personal responsibility for health, as part of a logical and semantical analysis of the concept.
- Responsibility is defined as a relation among six variables U, V, W, X, Y, and Z: *U is responsible to V for W to an extent of X and with regards to the time frame Y because of certain normative standards Z.*
- Each of the variables is explained and discussed throughout the chapter.
- If all conditions of the definition are taken into account, there is very limited scope to assign personal responsibility for health to individuals in a legitimate way. Both the authority of responsibility and the normative standard applied can be controversial, for a number of reasons.
- There are also a number of important ethical and political issues to be considered.
- The ascription of personal responsibility may also conflict very directly with personal rights and liberties usually regarded as fundamental from an ethical as well as a political perspective.

Acknowledgments This chapter is partially based on considerations, which appeared elsewhere (Langanke and Fischer 2012, Langanke et al. 2012, Langanke et al. 2013).

References

Agich GJ (ed.) (1982) Responsibility in Health Care. Kluwer Publishing, Dordrecht

Apel KO (1988) Diskurs und Verantwortung: Das Problem des Übergangs zur postkonventionellen Moral. Suhrkamp, Frankfurt a.M

Aristotle (2011) Nicomachean ethics. A new translation by Bartlett RC, Collins SD, with an Interpretive essay, notes, and glossary, University of Chicago Press, Chicago

Bærøe K, Cappelen C (2015) Phase-dependent justification: the role of personal responsibility in fair healthcare. J Med Ethics 41:836–840
Bayertz K (1995) Verantwortung: Prinzip oder Problem? Wissenschaftliche Buchgesellschaft, Darmstadt
Bishop G, Brodkey AC (2006) Personal responsibility and physician responsibility – West Virginia's Medicaid Plan. NEJM 355:756–758
Bringedal B, Feiring E (2011) On the relevance of personal responsibility in priority setting: a cross-sectional survey among Norwegian medical doctors. J Med Ethics 37:357–361
Brown RC (2013) Moral responsibility for (un)healthy behaviour. J Med Ethics 39:695–698
Brown H, Roberts J (2011) Exploring the factors contributing to sibling correlations in BMI: a study using the panel study of income dynamics. Obesity 20:978–984
Brownell KD et al (2010) Personal responsibility and obesity: a constructive approach to a controversial issue. Health Aff 29:379–387
Buyx A (2008) Personal responsibility for health – why we don't like it and why maybe we should. J Med Ethics 34:871–874
Buyx A, Prainsack B (2012) Lifestyle-related diseases and individual responsibility through the prism of solidarity. J Clin Ethics 7:79–85
Cappelen AW, Norheim OF (2005) Responsibility in health care: a liberal egalitarian approach. J Med Ethics 31:476–480
Cherry MJ (ed.) (1999) Personen and their Bodies: Rights, Responsibilities, Relationships. Springer, Berlin
Crawford J (1977) You are dangerous to your health: the ideology and politics of victim blaming. Int J Health Serv 7:663–680
Dabrock P (2006) Public health genetics and social justice. Community Genet 9:34–39
Düngen M (2009) Genetische Disposition und medizinische Behandlung. Available via Bundeszentrlae für politische Bildung. http://www.bpb.de/themen/J6B7SV,1,0,Genetische_Disposition_und_medizinische_Behandlung.html. Accessed 22 Feb 2011
Egger G, Swinburn B (1997) An " ecological" approach to the obesity pandemic. BMJ 315:477
Engelhardt HT Jr, Jotterand F (2008) The philosophy of medicine reborn. A pellegrino reader. Notre Dame IN, University of Notre Dame Press
Eyal N (2013) Denial of treatment to obese patients – the wrong policy on personal responsibility for health. Int J Health Policy Manage 26:107–110
Feinberg J (1977) Handlung und Verantwortung. In: Meggle G (ed) Analytische Handlungstheorie, vol 1, Handlungsbeschreibung. Suhrkamp, Frankfurt a. M., pp 186–224
Fischer J, Ravizza M (1998) Responsibility and control. A theory of moral responsibility. Harvard University Press, Cambridge, MA
Fleck LM (2012) Whoopie pies, supersized fries. "Just" snacking? "Just" des(s)erts? Camb Q Healthc Ethics 21:5–19
Frankfurt HG (1969) Alternate possibilities and moral responsibility. J Philos 66:829–839
French PA, Wettstein HG (eds) (2006) Shared intensions and collective responsibility. Blackwell, Boston
Grunwald A (1999) Verantwortungsbegriff und Verantwortungsethik. In: Grunwald A (ed) Rationale Technikfolgenbeurteilung. Konzeption und methodische Grundlagen. Springer, Berlin/Heidelberg/New York, pp 175–194
Hart HL (1949) The ascription of responsibility and rights. Proc Aristot Soc 44:171–194
Hillerbrandt R (2006) Technik, Ökologie, Ethik. Ein normativer Grundlagendiskurs über den Umgang mit Wissenschaft, Technik und Umwelt. Mentis, Paderborn
Jonas H (1979) Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation. Suhrkamp, Frankfurt a.M
Kamlah W (1972) Philosophische Anthropologie. Sprachkritische Grundlegung und Ethik. Bibliographisches Institut, Mannheim/Wien/Zürich
Knowles JH (1977) The responsibility of the individual. Daedalus. Winter 106(1):57-80

Krpic-Mocilar T (2003) Mitverantwortung für die eigene Gesundheit, vol 122, Studien zur Rechtswissenschaft. Kovac, Hamburg
Langanke M, Fischer T (2012) Gesundheitsmanagement liegt mir im Blut. Individualisierte Medizin und gesundheitliche Eigenverantwortung. In: Gadebusch-Bondio M, Siebenpfeiffer H (eds) Konzepte des Humanen. Ethische und kulturelle Herausforderungen. Alber, Freiburg, pp 139–172
Langanke M, Fischer T, Brothers KB (2012) Public health. It is running through my veins: personalized medicine and individual responsibility for health. In: Dabrock P, Braun M, Ried J (eds) Individualized medicine between hype and hope. Exploring ethical and societal challenges for healthcare. Lit, Münster, pp 165–188
Langanke M, Fischer T, Erdmann P, Brothers KB (2013) Gesundheitliche Eigenverantwortung im Kontext Individualisierter Medizin. Ethik Med 25:243–250
Lenk H, Maring M (eds) (1991) Technikverantwortung. Güterabwägung, Risikobewertung, Verhaltenskodices. Campus, Frankfurt a.M./New York
Lenk H, Maring M (2003) Natur, Umwelt, Ethik. Lit, Münster
Lewis KH, Rosenthal MB (2011) Individual responsibility or a policy solution – cap and trade for the U.S. diet? N Engl J Med 365:1561–1563
Maier B, Shibles WA (2010) The philosophy and practice of medicine and bioethics: a naturalistic-humanistic approach. Springer, Berlin/Heidelberg/New York
Marckmann G, Möhrle M, Blum A (2004) Gesundheitliche Eigenverantwortung. Möglichkeiten und Grenzen am Beispiel des Malignen Melanoms. Hautarzt 55:715–720
Marx B (ed) (2010) Widerfahrnis und Erkenntnis: Zur Wahrheit menschlicher Erfahrung. Evangelische Verlagsanstalt, Leipzig
Minkler M (1999) Personal responsibility for health? A review of the arguments and the evidence at century's end. Health Educ Behav 26:121–141
Nielsen ME, Andersen MM (2014) Should we hold the obese responsible? – some key issues. Camb Q Healthc Ethics 23:443–451
Ott K (1997) Ipso facto. Zur ethischen Begründung normativer Implikate wissenschaftlicher Praxis. Suhrkamp, Frankfurt a.M
Paul N (2010) Medizinische Prädiktion, Prävention und Gerechtigkeit: Anmerkungen zu ethischen Dimensionen eines biomedizinischen Ideals. Ethik Med 22:191–205
Pearson SD, Lieber SR (2009) Financial penalties for the unhealthy? Ethical guidelines for holding employees responsible for their health. Health Aff 28:845–852
Prainsack B, Buyx A (2011) Solidarity. Reflections on an emerging concept in bioethics. Nuffield Council on Bioethics, London
Prainsack B, Buyx A (2017) Solidarity in biomedicine and beyond. Cambridge University Press, Cambridge
Putnam H (1982) Why there isn't a ready-made world. Synthese 51:141–167
Resnik DB (2014) Genetics and personal responsibility for health. New Genet Soc 33:113–125
Rohr M, Schade D (2000) Selbstbestimmung und Eigenverantwortung im Gesundheitswesen – Ergebnisse des Workshops zu Forschungsbedarf im Bereich Medizin und Gesundheit. Arbeitsbericht Nr. 176, Akademie für Technikfolgenabschätzung in Baden-Württemberg, Stuttgart
Ropohl G (1993) Neue Wege, die Technik zu verantworten. In: Lenk H, Ropohl G (eds) Technik und Ethik. Reclam, Stuttgart, pp 149–176
Rudnick A (2004) An introductory course in philosophy of medicine. Med Humanit 30:54–56
Schicktanz S, Schweda M (2012) The diversity of responsibility: the value of explication and pluralization. Med Stud 3:131–145
Schmidt H (2008) Bonuses as incentives and rewards for health responsibility: a good thing? J Med Philos 33:198–220
Schmidt H (2009a) Just health responsibility. J Med Ethics 35:21–26
Schmidt H (2009b) Personal responsibility in the NHS Constitution and the social determinants of health approach: competitive or complementary? Health Econ Policy Law 4:129–138

Steinbrook R (2006) Imposing personal responsibility for health. NEJM 355:753–756
Veatch RM (1980) Voluntary risks to health. The ethical issues. JAMA 243(1):50–55
Venkatapuram S (2011) Health justice. An argument from the capabilities approach. Polity, Cambridge
Watkin D (1978) Personal responsibility: key to effective and cost-effective health. Fam Community Health 1:1–7
Weber M (1919) Politik als Beruf. In: Gesammelte Politische Schriften (herausgegeben von Johannes Winckelmann), 5th edn. J.C.B.Mohr, Tübingen 1988, pp 505–560
Werner MH (2006) Zur Eigenverantwortung gesetzlich Krankenversicherter. Beiträge zum Recht der sozialen Dienste und Einrichtungen. 61: 1–23
Werner MH (2011) Verantwortung. In: Düwell M, Hübenthal C, Werner MH (eds) Handbuch Ethik, 3rd edn. Metzler, Stuttgart/Weimar, pp 541–548
Werner MH (2013) Verantwortung. In: Grunwald A (ed) Handbuch Technikethik. Metzler, Stuttgart/Weimar, pp 38–43
Werner MH (2017) Verantwortung. In: Ott K, Dierks J, Voget-Kleschin L (eds) Handbuch Umweltethik. Metzler, Stuttgart/Weimar
Wikler DI (1978a) Persuasion and coercion for health: ethical issues in government efforts to change lifestyles. Milbank Mem Fund Q 56:303–333
Wikler DI (1978b) Coercive measures in health promotion: can they be justified? Health Educ Monogr 6:223–241
Wikler DI (2002) Personal and social responsibility for health. Eth Int Aff 16:47–55
Wiley LF, Berman ML, Blanke D (2013) Who's your nanny? Choice, paternalism and public health in the age of personal responsibility. J Law Med Ethics 41(Suppl 1):88–91
Wiley LF, Micah LB, Blanke D (2013) Who's your nanny?: choice, paternalism and public health in the age of personal responsibility. J Law Med Ethics 41:88–91
Yang YT, Nichols LM (2011) Obesity and health system reform: private vs. public responsibility. J Law Med Ethics 39:380–386
Yoder SD (2002) Individual responsibility for health: decision, not discovery. Hastings Cent Rep 32:22–31
Zimmerman MJ (2001) Responsibility. In: Becker LC, Becker CB (eds) Encyclopedia of ethics, vol 3, 2nd edn. Routledge, New York, pp 1486–1492

Part IV

Clinical Settings and Healthcare Personnel

Applying Medical Knowledge: Diagnosing Disease

39

William E. Stempsey

Contents

Abstract

The term "diagnosis" can refer to the name of a disease that afflicts a person or to the process of determining a diagnosis in the first sense. Complex philosophical, metaphysical, epistemological, normative, and logical issues permeate all aspects of diagnosis, beginning with the question of what is being diagnosed. The nature of disease is philosophically controverted. Both ontological and physiological conceptions of disease continue to influence thinking about disease and hence how to diagnose it. The question of whether the concept of disease is essentially value laden remains open. Diagnosis presupposes some classification of diseases, known as a nosology, in order to distinguish one disease from another. There are many possible ways to classify diseases, and nosologies can have different goals, e.g., providing a basis for rational treatment and prognosis, enabling statistical

W.E. Stempsey (✉)
Department of Philosophy, College of the Holy Cross, Worcester, MA, USA
e-mail: wstempse@holycross.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_31

reporting, fostering research, and for administrative aspects of health care. The major elements of the diagnostic process include the history of the illness, the physical examination, and various kinds of laboratory and clinical testing. Each of these elements requires interpretation and is influenced by philosophical presuppositions. Diagnostic reasoning makes use of probabilistic, causal, and deterministic models. It is fundamentally a process of hypothesis formation about possible diagnoses (differential diagnosis) and the systematic confirmation and ruling out of possibilities until one diagnosis is judged best to explain all the data. Diagnosing disease is important for a broad array of medical and social reasons.

Introduction

The term "diagnosis" can refer to the name of a disease or biomedical problem afflicting a person or to the process of determining the presence and nature of the disease. Although the process of diagnosis has been extensively studied and there is general agreement on basic aspects, the complexity of many real-life situations makes it difficult to formulate a universal description of the diagnostic process. One reason for this is that complex philosophical issues permeate all facets of diagnosing disease, including the question of just what disease itself is. This chapter addresses metaphysical, epistemological, normative, and logical issues in three aspects of diagnosing disease: the nature of disease, nosology or the classification of diseases, and the diagnostic process and logical models of diagnostic reasoning.

Disease

If diagnosis is a search for the presence and nature of disease, the nature of diagnosis will depend upon conceptions of disease and how diseases are classified. This section deals with the concept of disease and the following one with classification. The word "disease" can refer to the class of all diseases or to various subsets of that class, such as pneumonia, diabetes mellitus, arthritis, etc., and further subsets, such as bacterial pneumonia, viral pneumonia, rheumatoid arthritis, septic arthritis, etc. "Disease" can also be used to refer to a single instance of one of the above. Other related terms, such as illness and sickness, are often used in common discourse as synonyms but can also have precise meanings and relationships that vary in different theories of health and disease. This chapter focuses on disease as distinguished from related terms such as illness and sickness in the standard biomedical model. Disease refers to a set of biological phenomena that are said to be the cause of a person's experience of illness, which is feeling unwell. The social ramifications of disease and illness are described by the term "sickness."

The metaphysics of disease remains controversial. Historically, there have been two fundamental conceptions of the nature of disease qua disease: physiological and ontological (Temkin 1963, p. 631; Engelhardt 1975, pp. 125–141). The classic

physiological conception of disease is the humoral theory that goes back to the time of Hippocrates and dominated for centuries. It presupposes a biologically conceived teleology. When the body is functioning in accord with its nature, there is a proper balance of the four humors: blood, phlegm, black bile, and yellow bile. Disease is deviation from the normal healthy state, that is, an imbalance of the humors (Temkin 1973, p. 398). Even though the classical humoral theory has been long abandoned, the physiological theory of disease has remained influential in concepts such as homeostasis and the idea of disease as deviation from the state of normality called health.

While physiological conceptions take diseases to be deviations from some normal state, ontological conceptions take diseases to be things in themselves. This does not necessarily mean that diseases are concrete things. For example, Thomas Sydenham (1624–1689) held that diseases are observable clusters of signs and symptoms but that they cannot be localized to any particular organ in the body. Still, these clusters display regularity from individual to individual and were conceived as unchanging abstract objects similar to species of plants (Nordenfelt 1995, pp. 152–153). Other ontologists take disease to be actual physical entities. With the discovery of the association of bacteria, parasites, and the like, with certain clusters of symptoms, some ontologists identified these invading organisms as the disease. Another ontological view is that of Rudolf Virchow (1821–1902), the early champion of cellular pathology. Virchow at first repudiated the ontological conception of disease in favor of a physiological one. In 1847, he wrote that diseases were only physiological phenomena under altered conditions, but by 1895 he was calling himself a "thoroughgoing ontologist," regarding pathological cells within the body as the disease itself and not merely the cause of the disease (Virchow 1958, pp. 26, 192; Stempsey 2000, p. 72).

While the physiological and ontological conceptions of disease continue to influence certain aspects of contemporary diagnosis, current thinking is more focused on historical development of concepts of disease. Disease today is more likely to be seen as a process, influenced by current scientific thinking and cultural influences, emphasizing cause, the bearer of the disease, and the set of manifestations (Nordenfelt 1995, pp. 172–173). For example, since the completion of the mapping of the human genome, the concept of genetic disease has seen great emphasis.

One of the major issues philosophers of medicine have explored is the question of whether the concept of disease is essentially value laden. Two opposing positions are often referred to as naturalism and normativism. Naturalists, most prominently Christopher Boorse (1997) in his influential biostatistical theory, believe that the concept of disease is purely descriptive; disease is a value-free scientific concept. Normativists, who have proposed a diverse set of theories, take various kinds of values to be essential components of the concept of disease (Engelhardt 1975; Nordenfelt 1995; Stempsey 2000).

In short, contemporary thinking about the concept of disease can only be described as complex (Hofmann 2001). It is being reframed contextually, based on metaphysical, epistemological, and axiological commitments, to reflect the pathological processes that afflict human beings, bring them to the attention of medical professionals, and serve as warrants for treatment (Cutter 2003).

Nosology

If diagnosis is about explaining a set of symptoms (subjective experiences of feeling unwell) and signs (objectively observable phenomena) in terms of some particular disease, then the question of how one disease is differentiated from another is foundational for diagnosis. The classification of diseases is known as nosology. A fundamental philosophical question for nosology is whether disease classifications mirror independently existing diseases in the realist sense or whether classifications are constructed for various purposes and values. Whether the class of diseases constitutes a natural kind is still debated; Reznek (1987) has denied this. However, even if disease were a natural kind, it is clear from an examination of various existing nosological systems that disease classifications aim at much more than simply trying to map entities that exist in a realist sense.

There are several desirable characteristics for any nosology (Murphy 1997, pp. 122–126): (1) Disease categories should correspond to naturally occurring sets of characteristics seen in particular diseases. (2) A classification ought to be exhaustive, i.e., it should include all the conditions for which people seek medical help. (3) Categories should be disjoint. That is, no particular case should fall into more than one category. (4) The classification should be useful for understanding disease mechanisms, categorizing descriptive features, fostering effective treatment or management, and determining a prognosis and for purposes of such matters as administration, law, and education. (5) The classification should be as simple as possible and still achieve its goals. (6) The classification ought to be constructible, in the sense of allowing both exhaustiveness and disjointness. It is clear that these are ideal characteristics; they cannot all be achieved simultaneously. For instance, complete exhaustiveness and disjointness are in practice impossible to achieve. Most diseases can be classified in several ways, e.g., according to etiology, symptoms, or anatomical location. Exhaustive classifications will necessarily fail to be disjoint. Decisions about which ideals should take priority will have to be made according to usefulness for the particular intended purposes of the nosology.

Thomas Sydenham (1979), in the seventeenth century, and François Boissier de Sauvages (1768), in the eighteenth, constructed nosologies that sought explicitly to be exhaustive in scope. Assuming that diseases were ontological kinds, they relied on empirical observation over rationalistic systems, believing that this would provide a basis for rational treatment. René Laennec (1982), in the late eighteenth century, advocated an anatomico-clinical approach, showing that a nosology based purely on symptoms would not be exhaustive. He conceived of the human organism as consisting of three parts: solids, liquids, and the *principe vital*, an animating force. All diseases were classified as lesions in one of these parts. (Stempsey 2000, pp. 109–110).

Such nosologies are of interest today primarily for historical reasons, but they do reflect the ongoing interest in classifying diseases according to a set of ideals and also emphasizing the practical usefulness of a nosology for rational treatment. An examination of contemporary disease classifications shows a commitment to the same sorts of ideals but with a greatly expanded set of particular purposes.

The International Statistical Classification of Diseases and Related Health Problems (ICD) is one of a family of international classifications of health, disease, disability, and health interventions produced by the World Health Organization. The current, tenth revision (ICD-10), is intended to assist in the "systematic recording, analysis and interpretation and comparison of morbidity and mortality data" across time and place (WHO 2010, v.2, sec. 2.1). Thus, its primary purpose is to facilitate statistical reporting of morbidity and mortality. The classification is divided into 21 chapters. Chapters I–XVII cover local diseases, arranged by the main systems of the body. Chapter XVIII includes symptoms, signs, and abnormal clinical and laboratory findings that are not elsewhere classified. Chapter XIX covers injuries, poisoning, and other consequences of certain external causes. Chapter XX includes other external causes of morbidity and mortality. Chapter XXI classifies data explaining reasons for contact with health services by a person not currently sick, or circumstances in which a person is receiving care.

The Systematized Nomenclature of Medicine (SNOMED) is a comprehensive, multilingual collection of clinical health-care terminology. It is owned and maintained by the International Health Terminology Standards Development Organisation (IHTSDO), a not-for-profit association governed by its national members, twenty-seven countries as of April 2014, and headquartered in Denmark (IHTSDO 2014a). The current version, SNOMED CT (Clinical Terms) is a computer-based terminological system. Terminological systems provide terms denoting concepts and their relations from a specific domain and can be used to describe information in a structured and standardized way. SNOMED CT enables consistency in indexing, storing, retrieving, and aggregating clinical data across specialties and health-care venues. It enables computerizing medical records, thus providing consistency in the way data is stored, encoded, and used for clinical care and research (Cornet and deKeiser 2008). Thus, it serves primarily a clinical purpose: it brings consistency to the way patient data is stored.

SNOMED grew out of the Systematized Nomenclature of Pathology (SNOP), developed by the College of American Pathologists (CAP) in 1965 and later extended to other medical fields as SNOMED. In 2007, the intellectual property rights for SNOMED CT and previous versions of SNOMED and SNOP were transferred from the CAP to the IHTSDO (IHTSDO 2014b).

SNOP was intended to assist pathologists in standardizing terminology in the cataloguing of specimens. SNOP presumes an anatomico-clinical model of disease, in which all diseases can be described by some anatomical change. Although SNOP did not claim to be clinically oriented, it did provide an exhaustive classification of disease, if disease is understood in that particular anatomico-clinical model. Diseases are described with respect to four fields: topography (the part of the body affected), morphology (the structural changes produced in the disease), etiology (the etiologic agent responsible for the disease), and function (the manifestations of the disease). Within each field, terms are given a number up to four digits, the number of digits reflecting increasing specificity. For example, if a small bowel specimen shows ulceration (M4003) in the ileum (T65) with recovery of *Salmonella typhi* (E1361), and the patient shows clinical manifestations of disease (F9497), the specimen is

coded as T65-M4003-E1361-F9497 (Stempsey 2000, pp. 112–114). SNOMED in 1979 expanded the "SNOP concept" with three additional fields: a disease field was added to record discharge diagnoses for statistical reporting purposes; a procedure field to record administrative, diagnostic, and therapeutic and preventative procedures; and an occupation field to formally report a patient's work for purposes of specialties like industrial medicine (Stempsey 2000, pp. 114–115). Thus, even early versions of SNOMED have already moved beyond simple classification.

The current version, SNOMED CT, refines the primary purpose of this family of classifications with its emphasis on making health records accessible electronically and meaningful for clinical and administrative uses. Content is presented using three components: concepts, descriptions, and relationships. Concepts are arranged into hierarchies (clinical findings, procedures, body structures, organisms, physical objects, physical forces, events, social context, etc.) from the general to the more detailed; each concept has a unique numeric identifier. Descriptions link other appropriate terms to concepts. A concept can have several associated terms that describe the same clinical concept. Every description has a unique numeric description identifier. Relationships link concepts to other concepts. One example is the "is-a" relationship, which can be used to relate a concept to more general concepts. For instance, the concept "infective pneumonia" bears the "is-a" relationship to the more general concept "pneumonia," and both "bacterial pneumonia" and "viral pneumonia" bear the "is-a" relationship to the more general "infective pneumonia." SNOMED CT thus allows retrieval of information about many elements of disease and its clinical management with a high degree of complexity. This enables a wide range of clinical meanings to be captured in a record, without requiring the terminology to include a separate concept for every detailed combination of ideas that may be relevant to a particular case (IHTSDO 2014c).

The entire question of disease and nosology in psychiatry is more controverted and will only briefly be considered here. The fifth chapter of ICD-10 is devoted to mental disorders and serves as a worldwide standard, but the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) is perhaps more influential and has received more notoriety for its sometimes very significant revisions. The current, fifth major revision (DSM-5) was published in 2013. Successive revisions of the DSM have increasingly moved away from Freudian understandings of psychiatry to a more scientific, or at least evidence based, approach to understanding psychiatric disease and its classification. A major problem, however, is that the etiology of a great number of mental disorders is inadequately understood and so cannot serve as the basis for nosology as is increasingly the case with somatic disease. Furthermore, the obvious social dimensions of mental disorders and the difficulty in standardizing such dimensions create problems in trying to establish precise disease categories for mental disorders. Social norms change, and this raises the question of whether psychiatric diagnostic categories simply reflect current norms or are naturally occurring entities more like somatic disease entities. The inclusion and then removal of homosexuality as a mental disorder, for example, has led some to suggest that psychiatric diseases are simply reflections of current social mores and defined by committee fiat. The situation is far

more complex, however, and will not be further considered here. It is simply noted that the purpose of psychiatric nosology remains oriented toward the same practical goals as other types of nosologies (Jablensky 2012, pp. 77–94).

Current nosology thus is oriented toward very different goals than were the first attempts to categorize all diseases centuries ago. Current classifications are primarily reporting mechanisms. Although they still recognize the practical import of nosology as facilitating effective care of those who suffer, they make no claims at offering frameworks for choosing treatments. Rather, they serve primarily as means for categorizing disease for purposes of research that will foster public health and for administrative assistance in health-care management and payments.

Diagnosis

The process of diagnosis includes three basic elements: the history of the illness, the physical examination, and laboratory and other sorts of clinical tests. In a typical case, a physician "takes a history" from the patient; develops hypotheses about the diagnosis; performs a physical examination; generates a differential diagnosis, a list of possible diagnoses that fit the hypotheses; tests the hypotheses by laboratory and other sorts of clinical tests such as x-ray examinations; modifies the differential diagnosis; and repeats these various steps until a diagnosis is arrived at (LeBlond et al. 2009, pp. 11–12). This section first considers the three elements and then turns to the logic of diagnosis and the process of reasoning that physicians follow.

History of the Illness

"Taking the history" refers to the initial conversation of physician and patient in which the physician listens to the patient's relating of why he or she is seeking medical help and asks questions to help clarify the information provided by the patient. Some have recommended that the term "taking a history" is misleading and should be replaced because it puts the patient in a merely passive state (Lazare et al. 1995, p. 18), but "taking the history" is still most widely used.

Patient assessments, which include the history and other elements of diagnosis, can be comprehensive or focused. Comprehensive assessments are appropriate for new patients, whether in the hospital or in an outpatient setting. A comprehensive health history can provide fundamental and personalized knowledge about the patient and strengthen the clinician-patient relationship. Comprehensive initial assessments can also provide baselines for future assessments. Focused assessments are more appropriate for established patients, especially during routine or urgent care visits. They address particular concerns or symptoms, often restricted to a specific body system (Bickley and Szilagyi 2013, pp. 4–6).

The comprehensive adult health history seeks both subjective information, such as the patient's experience of pain, and objective information, such as age and family history. The history should include identifying data such as age, gender, occupation,

marital status, and an assessment of reliability, which may vary according to the patient's memory, trust of the physician, and other social factors. The initial focus is on the chief complaint or complaints, what it is that brought the patient to seek medical care. Several aspects of the patient's history are then sought. The history of the present illness includes the symptoms, how and when each symptom developed, and the patient's thoughts and feelings about the illness. The physician's questioning may seek to identify pertinent positives and negatives to aid hypothesis formation and to find out about the patient's medications, allergies, and history of smoking and alcohol use. The past history includes childhood illnesses; adult illnesses, whether medical, surgical, obstetric/gynecological, or psychiatric; and health maintenance practices such as immunizations, screening tests, and lifestyle issues. Family history includes age and health or cause of death of siblings, parents, and grandparents and the presence of any genetic diseases or other diseases that are disposed to run in families. Personal and social history includes educational level, family of origin, current household, and personal interests and lifestyle. These elements can provide important clues in the diagnostic process. The final part is the review of systems, in which the physician systematically asks about the presence or absence of common symptoms related to each major body system (Bickley and Szilagyi 2013, pp. 6–13).

The medical interview has three functions (Lazare et al. 1995, pp. 3–19). The first is determining and monitoring the nature of the patient's problem with the objective of enabling the clinician to establish a diagnosis, recommend further diagnostic procedures, suggest a course of treatments, and predict the nature of the illness. The medical interview enables a physician to generate multiple hypotheses during the course of the interview and elicit additional data to refute or confirm them. The history is estimated to contribute 60–80 % of data for diagnosis. The second function is focused on developing, maintaining, and concluding the therapeutic relationship. It helps to define the nature of the physician-patient relationship; communicate professional expertise, interest, respect, support, and empathy; and elicit the patient's perspective on the problem and the methods and goals of treatment. This may indirectly help the physician to glean effective diagnostic data from the interview. The third function is patient education and implementation of treatment plans with the objective of fostering consensus, patient satisfaction, cooperation, and improved treatment outcome. The importance of each function of the interview varies according to the nature of the interview, but the three functions are often interdependent. This functional analysis calls attention to the dynamics and complexity of the medical interview. The medical interview, or taking the history, has multiple purposes that go beyond mere diagnosis. The diagnostic process is ultimately carried out as a means for providing the most effective treatment for the patient and enabling the physician to offer a prognosis. Treatment involves providing drugs, surgery, and the like, but also developing a healing relationship between doctor and patient.

Kathryn Montgomery Hunter (1991, p. 5) sees all of medicine as fundamentally narrative, but most important are the opening stories that patients tell their physicians. She is representative of the school of thought that holds the nature of medicine to be better explained by the methodological analogue of literature than by the natural or social sciences. Even if one does not accept this narrative analysis for

all aspects of medicine, it does seem apt for certain activities, especially the medical interview. In this type of literary analysis, patients are seen as texts. While patients are ordinary readers of their texts, physicians are more like sophisticated interpreters. Physicians make sense of signs using a "diagnostic circle" very much like the hermeneutic circle, whereby parts of a text can be understood only with reference to the whole, and the whole can be understood only with reference to the parts (Hunter 1991, p. 9). Thus, with both patient and physician as readers, there is "one illness, two stories" (Hunter 1991, pp. 13–15). The patient's account of illness and the medical version of that account are fundamentally, irreducibly different narratives, and this difference is essential to the work of medical care. The "medical plot," the narrative organization of the case, is fundamentally shaped by the search for a diagnosis and an answer to the question of what is the best treatment for this particular patient (Hunter 1991, p. 65). The patient's story is itself given a sort of medical treatment, being retold in light of the physical examination and clinical tests; in the process, it becomes a fundamentally different story (Hunter 1991, pp. 128–130). The rewriting of the patient's story becomes part of the healing process (Hunter 1991, pp. 138–141). This sort of analysis reinforces the importance of seeing diagnosis in the context of a larger therapeutic relationship. It also highlights the importance that interpretation has in all aspects of diagnosis.

Physical Examination

The physical examination is the second major element that contributes data to the diagnostic process. A comprehensive presentation of the goals and processes of carrying out the physical examination can be found in well-established texts on physical diagnosis such as LeBlond et al. (2009) and Bickley and Szilagyi (2013); a detailed description of them will not be given here. The physical examination is often considered to be a purely scientific matter of observation and a source of objective data, opposed to the subjective data that patients report during the medical interview. However, there are several conceptual issues underlying the physical examination that influence how physical findings are perceived and interpreted. First, the observations made by clinicians during a physical examination are first of all perceptions of the examiner. As such, they are inherently subjective. As has been generally recognized by philosophers of science, all observations are fashioned by the underlying theories that are brought to the observations. To take one classic example, Koplik's spots are considered to be pathognomonic for measles. These are small, bluish-white specks on an irregular red background in the buccal mucosa and occur early in the course of measles. Whether particular bluish-white specks should be considered Koplik's spots is a judgment that depends on the subjective perceptions and judgments of the diagnostician. Such judgments already depend on conceptual commitments about the nosological entity of measles (Engelhardt 1981, pp. 305–306). Psychological factors, such as assumed probabilities and distortions due to expectations, are also known to influence perceptions of what is and what is not observed (Kahneman and Tversky 1982, pp. 144–146).

Diagnostic Testing

The third element of diagnosis consists of the many sorts of laboratory and other clinical tests carried out for the purpose of confirming or ruling out the hypotheses arrived at during the history and physical examination. In light of the ambiguity that can result from the perceptual issues inherent in the physical examination, laboratory and other clinical test results are often preferred and considered to be more objective. Many of the same conceptual issues, and even more complex ones, enter into the interpretation of test results, however.

Clinical tests can include laboratory testing of blood and body fluids such as urine, surgical biopsies of tissues, and clinical examinations such as x-ray studies, computerized tomography, echocardiography, sonography, and other types of procedures.

There are several purposes for which such testing might be done. First, laboratory and clinical tests can be a way to mitigate problems of interobserver variability in both the medical interview and the physical examination. Second, testing allows physicians to extend their medical examinations below the surface of the body. Third, quantitative assessment of levels of bodily constituents can give insight into the physiological processes occurring within the body. Fourth, testing such as bacterial and fungal culturing or toxicological screening can identify the etiological agents responsible for disease. Fifth, laboratory testing of body components can identify markers that might indicate risk for future development of disease. With the rapid development of biotechnology focused on the gene, this last purpose is assuming increasing importance (Stempsey 2000, pp. 149–150).

Two aspects of philosophical importance in laboratory and clinical testing are the choice of tests and the interpretation of tests. Choice of whether to do a test at all and choice of the particular tests to be done involve normative elements, which can be understood in terms of a complex cost-benefit analysis. First, the value of the test for confirming or excluding the particular diagnosis must be considered. Second, the consequences for the patient of including or excluding the diagnosis must be considered. Some diagnoses may be of little importance for the patient, but others, such as being HIV positive, bring serious social consequences. Reducing uncertainty or identifying risk may be important to some patients but not to others. Third, the risk of the diagnostic procedure itself to the patient has moral import. Tests bring discomfort and risk of harm to the patients. The importance of gaining information must be balanced against the potential harm of the test. Many routinely done tests expose patients to radiation, for example. Even a routine blood draw exposes a patient to a very small chance of bruising and infection. To take a more serious case, it is hard to see how one could justify exposing a patient to the dangers of a brain biopsy if the information that would be obtained could have no influence on decisions about treatment or prognosis. Fourth, testing incurs economic costs including wages to health-care personnel, cost of instruments and equipment, and possibly costs of hospitalization and loss of income to the patients. Fifth, if enough tests are performed, it is likely that conflicting results will be obtained. This is usually resolved by getting more tests, which might resolve the conflict but might also exacerbate it (Wulff 1976, pp. 109–110).

The second important philosophical aspect of laboratory and clinical testing is the interpretation of test results. As already mentioned, test results can conflict with one another. This problem is made worse by the easy availability of technology, such as automated machines that can produce many test results from one very small blood sample. Even when there is no conflict of data, results must still be interpreted for significance. Quantified data present particular problems. Precise numbers can suggest a degree of certainty that should not be presumed. In addition, establishing a range of values that is considered to be normal presents the difficulty of correlating population data and drawing inferences from that data for individual patients. These issues are further addressed in the next section.

Diagnostic Reasoning

The data obtained from the three elements of history, physical examination, and diagnostic testing must be gathered and interpreted to formulate a diagnosis, i.e., a name of a disease that best explains the data. This process is diagnostic reasoning. Diagnostic reasoning proceeds from effect to cause, the opposite logical direction used in explaining pathogenesis. Its process cannot be depicted by any one simple logical scheme for several reasons: because nosologies differ and change over time, because diagnosis has different end goals that vary in different clinical situations, and because many patients have multiple diseases rather than just one (Feinstein 1973a, pp. 212–232). For simplicity, "data" here refers to the evidence gleaned from the history, physical examination, and various clinical tests. The diagnostician must authenticate data, decide whether the data deviate from some designated state of normality, and consider the pertinence of the data for the goal of the particular diagnosis. Inferential reasoning then proceeds toward the goal of a diagnostic category and ends when the ultimate end goal, whether it is the ability to make a prognosis or to render a rational treatment, is reached (Feinstein 1973b, pp. 264–283).

It has been shown that physicians use many different strategies and make extensive use of heuristics in order to reach the end goal (Elstein et al. 1978, pp. 252–272; Tversky and Kahneman 1982, pp. 3–20). The process of diagnosis focuses on one diagnostic hypothesis or perhaps more than one. Hypotheses might be general or specific. Reasoning proceeds by progressive modification and refinement of the hypotheses. The arrived at diagnosis is then assessed for coherency, adequacy, and parsimony (Kassirer 1989, p. 894).

Psychologists have carried out extensive studies of how diagnosticians reason. Expert diagnosticians generate diagnostic hypotheses early. The hypotheses they consider are limited in number, rarely exceeding five. Physicians vary considerably in diagnostic effectiveness, depending on the nature of the problem at hand. Hence, diagnostic competence is not simply a characteristic of an individual diagnostician but is case dependent. Experience is also found to be a basic element of competence. Expert diagnosticians have knowledge of how findings relate to diseases or conditions, the relative frequency of the possible conditions in the population, and the

particular characteristics of those conditions that carry severe risk, even if their occurrence is low. Effective diagnosticians are able to retain all this information and use it correctly when needed. This capacity has been shown to be an outcome of repetitive practice (Elstein et al. 1978).

Although it is probably impossible to give an exact description of a precise reasoning process for every expert diagnostician in every case, three types of diagnostic reasoning are commonly used (Kassirer 1989, p. 894). First, probabilistic reasoning focuses on the probability of a particular diagnosis given the evidence. Assessment using only terms such as "likely," "common," and "rare" is problematic because such terms are vague and have no standard meaning. Probabilistic diagnostic reasoning turns to quantitative methods as more satisfactory. An ideal test would give an unequivocal answer that confirms or rules out a diagnostic hypothesis, but this is a rare occurrence except in cases where a disease is defined by a test result. Ascertaining the probability of a diagnosis given a positive or negative test result is covered by a rule formulated by Thomas Bayes in the eighteenth century. According to Bayes's theorem, the probability of a diagnosis in a particular patient depends on other probabilities: the prevalence of the diagnosis in the population from which the patient comes, the sensitivity of the test (the proportion of positive test results in people who have the disease), and the specificity of the test (the proportion of negative test results in people who do not have the disease). Thus, test results can be interpreted according to mathematical principles that are easily derived from the most fundamental laws of probability. This approach can be valuable for the diagnostic process, but its limitations must also be recognized. For example, suppose a test to detect a particular cancer has a false-positive rate of just 1 % (specificity = 0.99). Suppose, further, that the prevalence of the cancer in the adult population is known to be 100 per 100,000. Thus, of 100,000 people, 99,900 do not have cancer. If the test were administered to those without cancer, 1 % or 999 would have a false-positive test result (Bradley 1993, pp. 70–90). This alone shows the problem of relying on test results at face value and the need for understanding their statistical basis to allow proper interpretation. But there are also other limitations in this sort of probabilistic reasoning. The prevalence of a disease in a population is not always known; when this is the case, it must be subjectively estimated. In addition, many results cannot be described simply as positive or negative; continuous variables must be broken into discrete intervals to use in calculations. Bayesian calculation also depends on the assumption that diseases are mutually exclusive, which is problematic in many cases. In addition, certain diseases manifest themselves in stages and cannot be considered simply as present or absent (Kassirer 1989, p. 895).

The second type of diagnostic reasoning is causal reasoning, which is especially valuable for its explanatory power. Causal reasoning relies on the common sense notion of cause-and-effect relations between variables. In diagnostic reasoning, it focuses on describing anatomical, physiological, and biochemical mechanisms of the normally functioning human body, the body's pathophysiological behavior in disease, and idiosyncrasies of individual patients. Causal models (e.g., fluid-electrolyte equilibrium) are generated, usually relating stimuli and responses. The

process of testing, verifying, and falsifying hypothetical causal connections is a fundamental aspect of diagnosis. Causal reasoning can also be useful in setting the context for future data gathering in the diagnostic process. It can help in verifying a diagnosis and assessing its coherency. Major benefits of causal reasoning in diagnosis are its explanatory power and its ability to provide a rational basis for therapeutic interventions (Kassirer 1989, pp. 896–897).

The third type of reasoning is deterministic or categorical reasoning. Deterministic reasoning uses predominantly compiled knowledge that may arise from probabilistic or causal associations between clinical findings. It requires the identification of rules that describe routine practices. The rules can have many purposes, such as describing therapeutic approaches or making prognoses. They might also recommend further diagnostic tests given certain already ascertained data. The rules are in the form of conditionals: If x obtains, then do y; if x does not obtain, then do z. They are represented by branching algorithms, ordered sets of instructions in a flow chart. The flow chart contains a diagram of graphic symbols for each act of reasoning. Two main types of "boxes," often referred to as "nodes," are used to indicate logical activities. A decision box contains a statement of a question to be answered; an execution box contains a statement of a procedure to be performed. A decision box is followed by a branching pathway of at least two possibilities; the reasoning pathway takes a direction indicated by the answer to the question. Arrows are used to indicate the exits and pathways leading from one decision or execution box to the next (Feinstein 1974, pp. 6–7). Typically, each nonterminal node requires unequivocal answers, which then serve as the matter of the branches leaving that node. The terminal nodes represent precise outcomes, answering to the questions the algorithm was designed to answer.

Algorithms are particularly useful in relatively straightforward cases where the logic of the diagnostic process can be precisely defined. Another advantage is that with a well-defined and explicit procedure, it is difficult to omit important questions or tests. Deterministic reasoning, however, depends on the quality of the data that serve as input and does not deal effectively with uncertainty. It may also yield bad answers if the algorithm is applied in a context sufficiently different from the one for which it was designed. Finally, the need to formulate all the rules necessary for even moderately complex diagnostic tasks is a challenge. In complex cases, the branching algorithm can become unwieldy (Kassirer 1989, pp. 897–898).

Each of these three approaches has benefits and limitations, but the limitations can sometimes be ameliorated by the concurrent use of other approaches. Hence, the three approaches are complementary. Probabilistic models can be useful for triggering hypotheses but are dependent on knowing the prevalence of disease in the population from which the patient comes. Causal models, on the other hand, are specific to disease entities and independent of the patient population. They are dependent on fundamental knowledge of physiological function and dysfunction. Once a hypothesis proposes a particular cause, causal reasoning is useful for verification of the cause and for explaining the observations. Causal reasoning can also identify circumstances in which the assumption of independence between diseases required by probabilistic models does not hold. When a knowledge base is built from these models,

deterministic models may be constructed to aid in future diagnosis and even serve as bases for computer-assisted diagnosis (Kassirer 1989, p. 898).

Diagnostic Goals and Context

From a semantic standpoint, there are several different kinds of diagnoses. Nosological diagnosis purports to identify a disease or diseases from which a patient suffers. This, however, cannot completely describe diagnosis. It would require that every aspect of the diagnosis is the name of some disease and this does not reflect actual medical practice. Some diagnoses, for instance, are abnormality diagnoses; they include disease but also other disorders, injuries, wounds, lesions, defects, deformities, disabilities, etc. Other diagnoses are causal diagnoses; they give accounts or explanations of the data obtained in the diagnostic process. The category of diagnosis itself may not have clear boundaries, but may be "fuzzy" (Sadegh-Zadeh 2012, pp. 328–335).

The diagnostic context includes the patient, the physician, the physician's practice, the hospital, the patient's family, medical knowledge, and other factors; it produces a diagnosis as one of its outputs. Although a diagnosis is commonly purported to be a statement of some truth, it is perhaps better described as a performative utterance, a speech act, which generates truth and triggers individual, group, and even organizational behavior. A diagnosis imposes a social status on a person. It can exempt people from normal obligations, provide special financial compensation, and cause people who have committed crimes to be found non-culpable by reason of insanity. Thus, diagnosis is also essentially a social act. Diagnosis is, in this sense, a social construct (Sadegh-Zadeh 2012, pp. 335–339). While it may be the case that a diagnosis is constructed from facts, taken in a realist sense, the process of diagnosis depends on conceptual commitments and value judgments at every stage (Stempsey 2000). Diagnosing disease is important not only as a basis for effective treatment, but for a much broader array of medical and social reasons.

Definitions of Key Terms

Disease	(1) a set of biological phenomena that are said to be the cause of a person's experience of illness, which is feeling unwell; (2) the class of all diseases or various subsets of that class; and (3) a single instance of (2)
Diagnosis	(1) the name of a disease that afflicts a person; (2) the process of determining (1)
Differential diagnosis	(1) a set of diagnostic hypotheses that fit the data obtained from the diagnostic process; (2) the process of formulating (1)
Nosology	Classification of diseases

Probabilistic model of diagnostic reasoning	Ascertaining the probability of a diagnosis given particular data using standard mathematical models of probability such as Bayes's theorem
Causal model of diagnostic reasoning	The process of testing, verifying, and falsifying hypothetical cause-effect relationships that explain anatomical, physiological, and biochemical mechanisms of the body's pathophysiological behavior in disease
Deterministic model of diagnostic reasoning	Formulating rules that describe routine diagnostic practices based on compiled knowledge that may arise from probabilistic or causal associations. Rules are typically represented by branching algorithms in a flow chart containing decision points about possible ways to proceed given some determined answer

Summary Points

- "Diagnosis" can refer to the name of a disease or the process of determining a disease present in an individual.
- The concept of disease is philosophically controverted, and it influences judgments about diagnosis.
- Diagnosis presupposes nosology, which can take many forms depending on the goal or goals judged to be most important.
- The history of a patient's illness, the physical examination, and various kinds of laboratory and clinical tests all provide data for diagnosis.
- The process of diagnosis requires interpretation of all elements that disclose data and depends on conceptual and value commitments.
- Diagnostic reasoning includes various strategies and uses probabilistic, causal, and deterministic models in a complementary way.
- Diagnosis is essentially a social act and carries important social implications.

References

Bickley LS, Szilagyi PG (2013) Bates' guide to physical examination and history taking, 11th edn. Wolters Kluwer/Lippincott William and Wilkins, Philadelphia

Boorse C (1997) A rebuttal on health. In: Humber JM, Almeder RF (eds) What is disease? Humana Press, Totowa, pp 3–143

Bradley GW (1993) Disease, diagnosis and decisions. Wiley, Chichester

Cornet R, deKeiser N (2008) Forty years of SNOMED: a literature review. BMC Med Inform Decis Mak 8(suppl 1):S2. Available via http://www.biomedcentral.com/1472-6947/8/S1/S2. Accessed 24 June 2014

Cutter MAG (2003) Reframing disease contextually. Kluwer Academic, Dordrecht

de la Croix B, de Sauvages F (1768) Nosologia methodica sistens morborum classes juxta Sydenhami mentem et botanicorum ordinem. Fratrum de Tournes, Amsterdam

Elstein AS, Shulman L, Sprafka SA (1978) Medical problem solving: an analysis of clinical reasoning. Harvard University Press, Cambridge
Engelhardt HT Jr (1975) The concepts of health and disease. In: Engelhardt HT, Spicker SF (eds) Evaluation and explanation in the biomedical sciences. Reidel, Dordrecht, pp 125–141
Engelhardt HT Jr (1981) Clinical judgment. Metamedicine 2:301–317
Feinstein AR (1973a) An analysis of diagnostic reasoning. I. The domains and disorders of clinical macrobiology. Yale J Biol Med 46:212–232
Feinstein AR (1973b) An analysis of diagnostic reasoning. II. The strategy of intermediate decisions. Yale J Biol Med 46:264–283
Feinstein AR (1974) An analysis of diagnostic reasoning. III. The construction of clinical algorithms. Yale J Biol Med 47:5–32
Hofmann B (2001) Complexity of the concept of disease as shown through rival theoretical frameworks. Theor Med Bioeth 22:211–236
Hunter KM (1991) Doctor's stories: the narrative structure of medical knowledge. Princeton University Press, Princeton
IHTSDO (2014a) Welcome to IHTSDO. Available via http://www.ihtsdo.org. Accessed 25 June 2014
IHTSDO (2014b) History of SNOMED CT. Available via http://www.ihtsdo.org/snomed-ct/history0. Accessed 25 June 2014
IHTSDO (2014c) Introducing SNOMED CT, the global clinical terminology. Available via http://ihtsdo.org/fileadmin/user_upload/doc/download/doc_StarterGuide_Current-en-US_INT_20140222.pdf. Accessed 25 June 2014
Jablensky A (2012) The nosological entity in psychiatry: a historical illusion or a moving target? In: Kendler KS, Parnas J (eds) Philosophical issues in psychiatry II: nosology. Oxford University Press, Oxford, pp 77–94
Kahneman D, Tversky A (1982) Variants of uncertainty. Cognition 11:143–157
Kassirer JP (1989) Diagnostic reasoning. Ann Intern Med 110:893–900
Laennec R (1982) Traite d'anatomie patholigique. In: Boulle L, Grmek MD, Lupovici C, Samion-Contet J (eds) Laennec: catalogue des manuscrits scientifiques. Masson et Fondation Singer-Polignac, Paris, ms 2186 (III)
Lazare A, Putnam SM, Lipkin M (1995) Three functions of the medical interview. In: Lipkin M, Putnam SM, Lazare A (eds) The medical interview: clinical care, education, and research. Springer, New York, pp 3–19
LeBlond RF, Brown DD, DeGowin RL (2009) DeGowin's diagnostic examination, 9th edn. McGraw Hill, New York
Murphy E (1997) The logic of medicine, 2nd edn. Johns Hopkins University Press, Baltimore
Nordenfelt L (1995) On the nature of health: an action-theoretic approach, 2nd rev and enlarged ed. Kluwer, Dordrecht
Reznek L (1987) The nature of disease. Routledge & Kegan Paul, London
Sadegh-Zadeh K (2012) Handbook of analytic philosophy of medicine. Springer, Dordrecht
Stempsey WE (2000) Disease and diagnosis: value-dependent realism. Kluwer, Dordrecht
Sydenham T (1979) The works of Thomas Sydenham, MD, from the Latin edition of Dr. Greenhill, with a life of the author by Latham, RG, facsimile of the Sydenham Society's two-volume edition of 1848–1850, Classics of Medicine Library, Birmingham
Temkin O (1963) The scientific approach to disease: specific entity and individual sickness. In: Crombie AC (ed) Scientific change: historical studies in the intellectual, social and technical conditions for scientific discovery and technical invention, from antiquity to the present. Basic Books, New York, pp 629–647
Temkin O (1973) Health and disease. In: Wiener P (ed) Dictionary of the history of ideas: studies of selected pivotal ideas, vol 2. Scribner's, New York, pp 395–407

Tversky A, Kahneman D (1982) Judgment under uncertainty. In: Kahneman D, Slovic P, Tversky A (eds) Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge, pp 3–20
Virchow RLK (1958) Disease, life, and man: selected essays. Stanford University Press, Stanford
WHO (2010) International statistical classification of disease and related health problems, 10th rev, 2010 ed, 3 vols. World Health Organization, Geneva
Wulff HR (1976) Rational diagnosis and treatment. Blackwell Scientific, Oxford

Technology and Dehumanization of Medicine

40

Keekok Lee

Contents

Abstract

As the subject matter raised by the title is extremely large, this chapter can only focus on a few aspects for some detailed examination. The nature of technology in general would have to be looked at in order to set the scene for a later discussion of medical technology. Technology cannot be understood except as part of the philosophy of Modernism which involves the ontological *volte-face* of holding that all organisms, including the human organism are machines. This means that Modern Medicine not merely treats patients as machine but also uses machines to treat patients. Machines are intended to be cost-effective in the long run by increasing productivity; as such they necessarily replace human labor. In which medical contexts, then, would replacing human beings by machines constitute the highest level of dehumanization?

K. Lee (✉)
Faculty of Humanities, University of Manchester, Manchester, UK
e-mail: keekok.lee@manchester.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_69

Introduction

This essay would address the following aspects of the subject set out in its title:

1. The nature of technology; the history of technology in terms of different types of technologies, their respective relationships to basic/theoretical science; the values embedded in technology; the economics of technology.
2. The implications of above for medicine (which in this context includes aspects of healthcare), especially, in terms of (a) the doctor-patient relationship, (b) the nurse-patient relationship, (c) the relationship between (a) and (b) on the one hand and medical technology of the near future, and (d) the de-skilling, in one crucial aspect, of health professionals (doctors and nurses) with the advent of high-tech. The de-humanization of medicine would be explored within these contexts.
3. The essay would show that 2 is but the outcome of the logic of Modernity, since its inception in Western Europe from the seventeenth century – standing behind the Scientific Revolution, which ushered in Modernity, is a philosophical revolution involving the profound ontological *volte-face* of transforming the universe (including organisms and, therefore, human beings in it) to become machine. This constitutes the radical change to the artifactual mode in the perception and understanding of Nature and ourselves under Modernity.

Relationship Between Technology and Basic Science at the Empirical Level

What is technology? Put very simply and simplistically, technology is nothing but the tools which we humans use to enable us to accomplish certain ends, what we cannot achieve relying only on our four limbs as well as our sensory organs, such as our eyes. In the long history of human-kind, our early ancestors deployed what is called "found" technology or "prototechnology" (Ihde 1993, p. 48), that is to say, whatever object they happened to come across which could do the job they had in mind, such as, the fallen branch of a tree which could be used to reach for ripe fruit or nuts high up a tree, as a walking stick to help those limping with a bruised foot. In a similar manner, a rock with a sharp edge – an adze – was used to cut up meat or scrape clean an animal's hide. With the minimum amount of tinkering, the object could be made to become an instant tool. This type of tool use is certainly also found in other primates, such as chimpanzees.

"Found" technology is, however, not what pre-occupies the scholars of technology in general who are more interested in the history which followed that very early phase. For instance, those who study European technological civilisation have suggested dividing it up into various phases. Mumford (1946) proposes a threefold division (whose edges are meant to be overlapping) in terms of the type of energy and characteristic materials used. The eotechnic phase is a water-wind-and-wood complex; the paleotechnic phase is a steam-coal-and-iron complex; the neotechnic

phase is an electricity-and-alloy (as well as synthetic compounds) complex. The first, for him, stretches roughly from 1000 AD to 1750, the second, from 1750 to 1850s, and the third, from the 1850s to the present. Mumford's classification is heuristically enlightening in general but, perhaps, less helpful from the standpoint of this essay which is concerned with medical technology. So a different division is proposed, not based so much on the conjoint variables of energy and material, but on whether the technology is craft or science based. In the case of the latter, it shows that what is significant is the relationship between the technology and the kind of science it is based on. The more basic the theoretical discovery the more powerful, in general, is the technology generated – for instance, technology in medicine and agriculture based on Mendelian genetics (at the level of chromosomes) is less powerful than biotechnology based on DNA genetics and molecular biology – see Lee (2005). The suggested classification in the context of European technological history is as follows (however, bearing in mind that the boundaries between them are not meant to be neat and tidy, but overlapping):

Phase I: Relatively autonomous craft-based technology.
A: Roughly equivalent to Mumford's eotechnic phase.
B: Roughly equivalent to Mumford's paleotechnic phase.
Phase II: Science-theory-led technology.
A: Roughly equivalent to Mumford's neotechnic phase, but ending by the 1940s.
B: From the 1940s to the present.

Note that this division fails to superimpose neatly upon that which obtains in the history of science itself. There, the radical cleavage is between pre-modern science (up to the seventeenth century) and the rise of modern science (from the seventeenth century onwards). Phase IA falls clearly into the pre-modern scientific era, but Phase IB (roughly up to 1830s) falls clearly into the modern scientific period. In other words, the major cleavage has been drawn between the kind of technology which is theory led and inspired, in contrast to that which is relatively autonomous of basic scientific theories and discoveries themselves. Although Phase IB, in terms of temporal location, coincided with the rise of modern science, the technology it represented was, nevertheless, by and large, not a spin-off of theoretical advances.

On the contrary, during IB, it often happened that technology inspired theoretical research rather than that theoretical advances led the way to new technologies. For instance, this relationship of technology preceding theory is true in the case of the invention of the steam engine, which first appeared in the form of the steam pump, as a response to the demands of the coal mining industry to mine seams at deeper levels where flooding occurred. It later made railway transportation possible as the steam locomotive, and replaced sailing ships on the high seas in the form of the steamer. Attempts to improve its efficiency eventually led to the establishment of the abstract, fundamental science of thermodynamics by Sadi Carnot, a French army officer and engineer, and worked on later by famous scientists like Joule, Kelvin, Clausius and Boltzmann, Atkins 1984, p. 7 writes:

> The aims adopted and the attitudes struck by Carnot and by Boltzmann epitomize thermodynamics. Carnot traveled toward thermodynamics from the direction of the engine, then the symbol of industrialized society: his aim was to improve its efficiency. Boltzmann traveled to thermodynamics from the atom, the symbol of emerging scientific fundamentalism: his aim was to increase our comprehension of the world at the deepest levels then conceived. Thermodynamics still has both aspects, and reflects complementary aims, attitudes, and applications. It grew out of the coarse machinery: yet it has been refined to an instrument of great delicacy. It spans the whole range of human enterprise, covering the organization and deployment of both resources and ideas about the nature of change in the world around us. Few contributions to human understanding are richer than this child of the steam engine and the atom.

Even more remarkably, during IB, technological discoveries, which formed the very basis of the Industrial Revolution (at least in Britain), were made by people who knew no science, had no formal education and, indeed, in some cases, could not even read or write. The most famous of these apprentices and craft-based mechanics is George Stephenson. Later in life, when he became famous and rich, he was only partially successful in overcoming his illiteracy. What is now called the Davy Lamp – the safety lamp for miners, which first appeared in 1815 – was also an invention by Stephenson. But because of his humble background, illiteracy and ignorance of physics and chemistry, Humphrey Davy – Fellow and later President of the Royal Society on whom a baronetcy was eventually conferred – could not credit Stephenson as a fellow inventor. See Davies (1980, pp. 19–32).

Phase I, A and B, in spite of differences between them, share the essential similarity of being craft-based and relatively autonomous of explicit scientific/theoretical input. In other words, both IA and IB displayed a split between science and technology – either science was pursued relatively autonomously of technology or that technology led the way to scientific theorizing. The causal direction the other way round, of theory inducing technology, by and large, did not occur until much later under Phase II when the major technological innovations are theory led or induced. With regard to Phase IIA, on the theoretical side, by 1850, many of the fundamental scientific discoveries had already been made. Regarding electromagnetism, Faraday, in 1831, found that a conductor cutting the lines of force of a magnet created a difference in potential. This, together with the work done by Volta, Galvani, Oersted, Ohm, Ampere and Henry, provided the theoretical foundation for the conversion and distribution of energy as well as for such significant inventions like the electric cell, the storage cell, the dynamo, the motor, the electric lamp. During the last quarter of the nineteenth century, these were spectacularly translated into industrial terms in the form of the electric power station, the telephone, the radio telegraph. Augmenting these were the phonograph, the moving picture, the steam turbine, the airplane. That was on the physics front. On the chemistry front, equivalently spectacular developments followed theoretical advances. Mumford (1946, pp. 217–218) again has aptly written:

> In (this) phase, the main initiative comes, not from the ingenious inventor, but from the scientist who establishes the general law: the invention is a derivative product. It was Henry who in essentials invented the telegraph, not Morse; it was Faraday who invented the dynamo, not Siemens; it was Oersted who invented the electric motor, not Jacobi; it was

> Clerk-Maxwell and Hertz who invented the radio telegraph, not Marconi and De Forest. The translation of the scientific knowledge into practical instruments was a mere incident in the process of invention. While distinguished individual inventors like Edison, Baekeland and Sperry remained, the new inventive genius worked on the materials provided by science.

In other words, it was only roughly from 1850 onwards that modern society began to reap the material benefits promised by modern science, its method, its philosophy and its ideological goal of controlling Nature. That promise took more than two centuries to materialise when the paths of pure (theoretical) science and technology no longer diverged acting, by and large, independently of each other, but began to be harnessed to work as joint forces. However, at least on one level of understanding, the team may be said to be led by pure science, the senior partner, whilst technology follows. (Yet at a deeper level, this may be an over-simplification – for qualifications, see what follows.) In Phase I when each was relatively autonomous, technology, sometimes, led the way to theoretical advance – witness the relationship between the steam engine and the fundamental science of thermodynamics. However, under the new settlement, technology has lost that causal initiative and now becomes, much more so than before, the executive arm, so to speak, of pure science. Bear in mind what has already been observed, namely, that as technology becomes basic science led, it becomes more and more powerful as shown in the very history of medical technology recounted below.

As far as medical technology is concerned the phase called "found" technology in the history of (Western) modern medicine is neither here nor there. However, the phase of craft-based technology is relevant to its history. One instance which springs immediately to mind is the stethoscope whose life began when René Laennec (in 1819) invented it by rolling up a piece of paper, putting one end on the chest of the patient, the other end to his ear. He resorted to such a device because the patient in question happened to be an obese lady, making it difficult for him to listen to her lungs without such a make-shift medium. The scalpel in early surgery could be another instance, as it was basically a knife. However, do not forget that the crucial distinction between Phase I and Phase II is not the dates but that the former is craft-based technology and the latter is (basic/theoretical) science–induced. Just one example will be cited to illustrate this point, namely, the first kidney dialysis machine which appeared was a Heath Robinson contraption. The Dutch doctor, Willem J Kolff built an "artificial kidney" for his dying patient, made out of wooden drums, some cellophane tubing and laundry tubs. In this make-shift fashion, he was successful in draining the blood from the patient, removing impurities from it and then pumping the clean blood back into the patient. The date was as late as 1945, but the technology was purely craft-based – see Healthtechnologies timeline (2014).

Medical technology, since the seventeenth century, first came under Phase I, A and B. An example with some basic science input is the thermometer, an instrument with a history of several centuries; however, the modern user-friendly version was not available until Fahrenheit in 1724 constructed the Fahrenheit scale, with the freezing point of water at the lower end and the boiling point at the higher, and then manufactured a tool using such a scale and mercury (a material with a high

coefficient of expansion) to measure the fluctuations in temperature of the human body. The compound microscope, based on the discovery of the lens, was made by two Dutch spectacle makers, Zacharias Jansen and his father as early as 1590. However, it existed more as a novelty rather than as a serious tool for scientific research. It was left to van Leeuwenhoek, a Dutch draper turned scientist, to perfect it to advance biological knowledge, being the first to see and describe bacteria, yeast plants, the circulation of blood corpuscles in capillaries. Eventually, it enabled medical research to usher in the age of bacteriology with Robert Koch's discovery, first, of the anthrax bacillus in 1876, and even more importantly of the tubercle bacillus in 1882, and even later, the age of antibiotics (post-1945). Antibiotics became available to the general public when these were produced industrially after the Second World War made possible via the work of Howard Florey and Ernst Chase, who isolated the bacteria-killing substance found in the mould a decade after Alexander Fleming accidentally chanced to come upon some on one of the glass plates in his laboratory in 1928, which he had at an earlier date coated with staphylococcus bacteria as part of his research. Furthermore, Florey got an American drug company to mass produce the penicillin just in time to treat all the cases of bacterial infections amongst the Western troops on D-Day (6 June 1944). In 1945, Fleming, Chain, and Florey were awarded the Nobel Prize in medicine – see Lee (2012).

However, disease is not only caused by certain bacteria but also by certain viruses, with which the compound microscope cannot cope. Further progress was only made when the electron microscope was invented in the late 1930s. In other words, the history of the microscope covers both Phase I as well as Phase IIA, as the invention of the compound microscope is primarily a craft-based technological product, using the lens and grinding it, whereas the electron microscope could not have been invented without the discovery of basic science, that is, of quantum physics as pioneered by Bohr (1885–1962), Einstein (1979–1955), and others. More than the electron microscope, the invention of the X-ray machine and other later even more high-tech machines (under Phase IIB) illustrate excellently the indispensable role played by basic scientific discoveries such as radium and radiation through the pioneering work of Marie and Pierre Curie (1897–1904). In 1895, the physicist Wilhelm C. Roentgen discovered a form of electromagnetic radiation which could pass through the body, leaving on a photographic plate an image of the bones or organs, thereby enabling the doctor to see the human interior for the first time in medical history, trailing in its wake a whole suite of high-tech diagnostic tools, such as the electrocardiograph (1903) developed by the Dutch physician and physiologist Wilhelm Einthoven which involved a "string" galvanometer suspended in a magnetic field, measuring small changes in electrical potential as the heart contracts and relaxes. By strapping the device to the arms and left leg of the patient, Einthoven could record the heart's wave patterns – the string, by moving, obstructed a beam of light whose shadow was then recorded on paper or a photographic plate. For the invention of this machine, he was awarded the Nobel Prize in medicine in 1924. The CAT scan (computerised axial tomography) invented in 1972 goes beyond X-rays combining them with a computer to create very detailed images of the inside of the body, with the X-ray tube rotating around the body and the computer

producing an image of the scan. Unlike standard X-rays, a CAT scan can show up structures of blood vessels, tumours as well as bones – see "CAT scan (2014)." For this invention, the British engineer, Godfrey Hounsfield and the South African-born physicist Allan Cormack were awarded the Nobel Peace in medicine in 1979. On the other hand, magnetic resonance imaging (MRI) relies on magnetic fields and radio waves to produce also very detailed images of the inside of the body, including the brain and spinal cord as well as bones/joints, heart and blood vessels, soft tissues such as breasts, internal organs including the womb or prostate gland; indeed it is capable of doing a whole-body scan. As this diagnostic tool does not use X-rays, it eliminates the fear of radiation and also makes it possible to scan people with certain types of medical implants, such as a pacemaker operated by a battery. The first such equipment entered medical service in 1981. Paul Lauterbur and Peter Mansfield were awarded the Nobel Prize in medicine in 2003 for this diagnostic tool, although not without provoking a controversy. (See "MRI" (2014), Dreizen (2004)).

Technology and Economics

Technology, as tool, is meant to help us gain better control of Nature, to realise our own ends and projects. However, executing our goals and intentions requires the use of resources, whether these are taken directly from Nature (such as wood, titanium, fossil-fuel/solar/wind energy) or indirectly derived from Nature (such as plastic, brass or bronze). This means that technology and economics necessarily cross paths, as economics in general is concerned with the efficient allocation of resources, which in turn is linked with the notion of productivity. Productivity is generally defined as the amount of output per unit of input; furthermore, it is also regarded as a basic yardstick to measure the health of an economy. "It can be said without exaggeration that in the long run probably nothing is as important for economic welfare as the rate of productivity growth" (Baumol et al. 1989; see also Field 2008). However, for the purpose of this essay, the remit of the concept of productivity may be made much narrower, confining it only to that more familiar aspect which involves labor productivity in the economic system. Labor productivity simply means output divided by the number of workers, or the number of hours worked. Take the following hypothetical, though historically based, example: in the eighteenth century, in England, when weaving was a cottage industry, a worker, working 8 h a day, was able to weave, say, a foot of cloth. In the nineteenth century, a weaving machine, first powered by water and later by steam with one worker, working the machine 8 h a day, could produce a 100 f. of cloth. Labor productivity gain (or economic growth) would then be a 100 %. Machines replacing human labor historically were, and still, are a major means of increasing productivity (or growth), although sometimes, the increase could be obtained through a change in the technique of production (the software side of production, so to speak) rather than directly in replacing humans by machines (the hardware side of production), such as in the famous example cited by Adam Smith about the division of labor in the manufacture of pins. If one worker were to manufacture a pin from the beginning of the process to the end, then that

worker would produce, say, one pin a day, whereas if "(o)ne man draws out the wire, another straights it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations. . . to whiten the pin is another . . .. in this manner, into about eighteen distinct operations . . .." If these operations were shared between ten persons, between them, they could produce "upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made twenty, perhaps not one pin in day. . ." (Smith 1776, Book I, Chap. 1).

In similar spirit, Henry Ford combined the innovations above when he is said to have pioneered mass production in the motor car industry through automation – this meant that machines made large quantities of the parts needed which were then assembled together to make up the car as fast as the parts were produced by the machines – see "The evolution of mass production" (2014).

However, labor productivity as a concept is Janus-faced – on the one hand, it increases productivity in general, but on the other, it necessarily renders the workers it displaces at least, temporarily, if not permanently, out of work, their skills having been rendered superfluous. Historically, this trend has not been worrying, for the simple reason that another sector of the economy would open up to offer opportunities for employment – for instance, as labor productivity in agriculture improved, making farm workers redundant, many of these workers would become factory operatives as the manufacturing sector began to grow; when labor productivity in turn occurred in manufacture, many displaced (or younger) workers turned to the growing service sector of the economy. However, digital technology of late has (even ignoring the impact of robotic technology which will be looked at a little later) greatly improved labor productivity in all sectors of the economy including service sectors such as banking and retailing. Is there yet another sector of the economy waiting to absorb workers thus displaced by digital technology? As none so far has appeared on the horizon, this leads some to postulate that the near future will not be like the past, and some economists, such as Erik Brynjolfsson (a professor at the MIT Sloan School of management, as cited by Rotman 2013), to say: "It's one of the dirty secrets of economics: technology progress does grow the economy and create wealth, but there is no economic law that says everyone will benefit." (See also Brynjolfsson and McAfee 2011; Rifkind 2005). Hence, there is both gain and loss involved; in other words, although some would gain, many may well lose as the race against machines intensifies especially in the near and further future.

Technology, Economics, and Medicine

In today's society, medicine is an important part of any economy and may be considered to be an industry, whether the medicine practised is primarily state or privately funded. As such, the laws of economics involving the notion of labor

productivity would apply to it as relentlessly as they would apply to any other industry. What exactly then is the impact of economics upon the practice and theory of medicine? This section will look at the former.

By and large, in a state-funded health service, the onus on those operating it is to reduce the cost to the public purse; in a privately funded service, it is to return as much profit as possible to the shareholders of the corporation involved. (The issue is further complicated by the requirements of internal accountancy which demands each part of the health service to operate in the black). Inevitably, the health service has no choice but to opt for labor productivity, as labor costs are a standing item of expenditure, whereas if a machine could replace labor, although the initial investment (what is called fixed capital) may be great, it is a one-off investment of capital, such that in the longer term, gain rather than loss will show up in the accounting spread sheets.

Before this aspect of the impact on the practice of medicine will be explored in greater detail, one must straightaway point out that one should distinguish between machines and machines. For instance, the justification for some machines in medicine lies, it is said, primarily, to improve success rate in diagnosis and treatment. The CAT, the MRI scans, for instance, would fall into this category; so would a machine such as the laparoscope which permits keyhole surgery which is considered as less invasive than the older method, thereby permitting less pain and bleeding post operation, reduced less scarring, a shorter hospital stay, and a faster recovery time. (See "Laparoscopy" (2014) for a brief description of the instrument and other accompanying devices, under the specific conditions of their use for instance in removing a damaged or diseased organ in the patient.) The *raison d'être* of such machines is to improve the quality of the medical intervention, not to render the surgical/nursing team redundant (See Ballantyne (2002)). However, although this is a crucial matter not to be overlooked, this does not mean that all forms of machinery involved in medicine would be so positive in their impact upon the medical team/patient relationship.

This is because, as already shown in the preceding section, machines are the standard method of procuring labor productivity. In other words, the more fixed capital per worker is used, the more productive the worker will be (other things being equal) – for example, if one nurse sitting in the ward office could monitor on screen the data coming in from one or more machines attached to a dozen or more patients, indicating the condition of each patient lying in the adjacent ward, then the hospital would have gained on the cost of hiring, say, only one nurse to monitor the progress of 12 or more patients, every month, every year of the life span of the said machine (s). Furthermore, if the machines are designed with the capability of flashing up warning visual and sound signals about the condition of a patient as it is about to turn critical, then this would also enable the nurse on duty to sit in the office doing other administrative tasks while keeping an eye and an ear open regarding the monitoring system. This high-tech form of nursing would in the end alter the very nature of nursing itself – nursing would be less about looking after patients in an intimate, personal manner but more about occupying a monitoring role of their purely medical conditions, as an adjunct to doctors. In other words, the original Florence

Nightingale model of nursing which was about providing comfort, succour, and compassion to the suffering would attenuate, if not be totally superseded.

The above sort of consideration leads to the crux of this essay, namely, the link between high-tech in particular and dehumanization of medicine in the practice of medicine. Low-tech or craft-based technology, as previously shown, is more labor-intensive while high-tech is *ex hypothesi* less labor but more capital intensive. This means that in the context of running a hospital, a patient admitted to a modern hospital is more likely to encounter fewer human-contact/interaction moments, if not actually fewer human beings than if admitted to a "backward", less well-equipped establishment. Imagine the following: on arrival, the main door is operated automatically (quite unlike the scenario at a five-star hotel where a commissionaire stands at the porch, rushing forward to open the door of the car or taxi bearing the customer to its portals, with a bell-boy or two following immediately to take care of the luggage). In other words, if the front door of a hospital (whether automated or not) had a porter standing by to help open the door, carry the case to the reception, to say a cheerful hello and so on, the patient would feel more welcome than if such a worker had been dispensed with on the ground of saving cost. Take another scenario: instead of a human being (in the role of a nurse or administrator) taking the history of the patient and the illness, the patient in some establishments would be given a specially designed small computer and told to tick the right boxes to the various matters as presented by the electronic questionnaire. Only the literally illiterate or the computer-illiterate would be exempt from this impersonal mode of communication, the rest would have to struggle as best they can to make sense of it. The data inputted could then be said to be standardized and objective but the downside is that it denies the patient yet another occasion of making contact with a fellow human being who is in the position of an expert to help them negotiate their way through the process and procedure of seeking medical attention for their predicament.

If one were to peep into the near future of care of both the sick and/or the elderly (in advanced economies), the following scenario emerges in which robotics appears to play an increasingly prominent part. Japan plays the lead role in this evolution – see Dethlefs and Martin (2006). In terms of demographics, Japanese society is fast becoming one with an increasingly large elderly population; in terms of industrial manufacturing, it has pioneered the use of robots. So obviously, the robotic solution to the matter of caring for the sick but especially the elderly is an obvious option. In 2013, the Japanese government allocated 2.39bn yen to develop such robots. The really sophisticated ones are expected to replace human beings even to the extent of offering companionship; one such belonging to an older generation of machines is Paro who is regarded as a friend by the residents in a Japanese care home – see Hudson (2013), Kelly (2013). Furthermore, there is one version in the shape of a baby seal which can respond to its name; its interactions with the user make it seem as if it is alive, moving its head and legs, and it can learn to respond in ways the user prefers. The user can stroke it, it "feels" being stroked via its tactile sensor – see "Paro therapeutic robot" (2014). This version is acknowledged to have a wholesome psychological effect on the patients, enabling them to relax, by providing stimulation and motivation in very much the same way as a sympathetic fellow human carer can

do. In other words, such a robot appears to be able to replace a human being even in terms of providing companionship and fellowship. Technically, scientists/engineers could aim to combine such "psychological" capabilities with the physical ones of lifting the patient from one position to another, moving the patient from one location to another, helping the patient to dress and undress, to keep proper personal hygiene, to remind and jolly her to take the right pills at the right time, fetching and carrying, heating up a prepared meal in a microwave, vacuuming the floor, alerting the hospital when the elderly person takes a turn for the worst, (the elderly could live and be cared for in her own home looked after by if not one, then two, or a suite of robots). This futuristic scenario would then raise the question: if the robots are equivalent in function and performance to a human, then surely this kind of development in elderly care would not amount to dehumanization of medicine in the straightforward understanding of the term? However, it would be beyond the remit of this essay to further address this set of issues.

As for the doctor–patient relationship, one must distinguish between two different things: (a) the pressures on the time of the doctor (let us say) given the rising numbers of patients passing through the surgery means that no more than a few minutes could be given to each patient; (b) high-tech medical diagnosis which is the order today. Both contexts may make the patient feel the impact of the so-called dehumanization of medicine (the impact is said to be less in the private health system); however, here, the second context will only be considered. In days of yore, the doctor personally put the stethoscope against the patient's chest (today, while the stethoscope may still hang around the neck of the doctor, it is no longer used as a serious diagnostic tool but more perhaps as a trade icon), would feel the pulse, would palpate an organ or two, would look even at the condition of the tongue, the complexion, and so on, apart from asking questions about the onset and subsequent development of the illness – at the end of the consultation, the doctor would give the patient a diagnosis. However, today, diagnosis depends on the results of tests, involving blood, urine samples, tissue samples in biopsies and so on (see Pillinger (2014), Green (2005), Rull (2012)). Samples once taken from the patient are forwarded to specialist labs for analysis by experts such as a hematologist who would not have met and examined the patient; the doctor (s) would not (or would not dare to) pronounce until the results of all these tests become available and deciphered, no matter how sure she is in her mind about the condition of the patient. Impersonal tests mediate between the patient and the doctor; as a result, the patient may feel that her medical fate, for better or for worse, is sealed not so much by the doctor(s) but by machines and experts which carry out the analysis of these tests in a distant laboratory, who necessarily are faceless and nameless and have no immediate knowledge of their personal suffering and pain.

Another dimension of the dehumanization of medicine in the context of the doctor–patient relationship is even more radical as it involves in principle (though not in practice yet) of rendering the individual doctor superfluous. This is machine diagnosis; in one sense, it is already part of today's medical culture. The project had long been on the horizon but today with information technology in the ascendant, it looks as if a medical diagnostic software has successfully been designed and created.

It is called Isabel. The story began in 1999 when a young girl called Isabel was struck down by chickenpox; however, the doctor(s) in charge had overlooked two rare but well-known complications of chickenpox, namely toxic shock syndrome and necrotizing fasciitis. As a result, the raging effects of the latter are still with the patient, even today. Her father, Jason Maude, started that same year Isabel Healthcare, establishing a Web-based checklist system aimed at helping doctors uncertain about their diagnosis. It is not marketed as a replacement of doctors but as a back-up tool in case of uncertainties, as well as a teaching tool in medical education. In other words, this mode of presentation, though user-friendly to the medical profession, nevertheless, logically implies that (at least in the majority of, if not necessarily, in all cases), Isabel is more reliable than the averagely competent doctor the patient may encounter in the average surgery and average hospital – see Nash (2010) and Hafner (2012). The target market of Isabel is, therefore, the medical profession rather than the ordinary individuals who may prefer Isabel to diagnose their conditions and who can afford to buy a license for its use. However, today, there is a poor man's equivalent of Isabel; people use the internet to access information about their conditions and to self-diagnose in the light of such knowledge – see Kluwer (2014), which shows that individuals who use this mode believes that "collectively," the information yielded via the internet is greater than that held by any one individual practitioner of medicine. It appears that their line of reasoning is no different from that of those professionals who use Isabel. In turn, this raises the question: does machine diagnosis invariably dehumanize medicine or can it, under certain appropriate circumstances, empower the individual to redress to an extent the imbalance which has traditionally existed between the all-knowing professional and the, by and large, ignorant patient?

Division of Labor, Ontological *Volte Face* and Dehumanization of Medicine

Adam Smith's account of the division of labor is not confined merely to economics but has been extrapolated under Modernity to apply in the intellectual domain. In medicine, part of the tremendous growth in knowledge must be laid at the door of specialization which is the intellectual equivalent of the division of labor in manufacturing. Smith, while applauding the benefits such a technique undoubtedly would bring to economic growth, was also well aware of its downside, namely, to use the terminology of this essay, to dehumanize the worker. To quote him again, Smith 1776, Book V, Chap. I:

> The man whose whole life is spent in performing a few simple operations, of which the effects are perhaps always the same, or very nearly the same, has no occasion to exert his understanding or to exercise his invention in finding out expedients for removing difficulties which never occur. He naturally loses, therefore, the habit of such exertion, and generally becomes as stupid and ignorant as it is possible for a human creature to become. . . . in every improved and civilized society this is the state into which the labouring poor, that is, the great body of the people, must necessarily fall, unless government takes some pains to prevent it.

The downside of the division of labor in the manufacturing sector upon the laboring poor, as spelt out above by Smith, of course, simply does not obtain in the same fashion in the knowledge sector of the economy, (such as in medicine) as in the lower rungs of the manufacturing sector. However, the spirit of Smith's critique may be said to obtain, all the same, as intellectual specialisms, in principle, while permitting the expert to be an authority of his specialism, nevertheless, has the unfortunate effect of excluding him from knowledge in related, neighboring domains of knowledge, thereby forcibly making him ignorant about such fields. In other words, the human body is divided into parts, the study of each part falling into the domain of its own particular specialism – the hematologist is the expert on blood, the brain surgeon on the brain, the orthopedic surgeon on bones and fractures, the psychiatrist on the mental aspects of the individual, and so on. The hematologist has neither knowledge nor a professional view about the brain and vice versa – in the language of trade unions, everyone respects work boundaries. This fragmentation of knowledge means that often a patient might have to be passed along from one specialist to another, undergoing one test and another, before the patient would finally, with luck, arrive at the door of the right specialist for a proper diagnosis – this, indeed, is one argument for the relevance of machine diagnosis as performed by a sophisticated software such as Isabel, as Isabel does not have to respect such work and knowledge boundaries.

Furthermore, this proliferation of specialisms necessarily entails fragmentation of the human being, such that wholeness of the person is lost (Reiser 1978). In analyzing modern medicine or biomedicine today, division of labor and fragmentation must be understood at, at least, two levels, namely, the epistemological and the ontological as well as the relationship between them. Epistemological fragmentation has already been briefly referred to just above. But closely entwined with that is the fragmentation involved when the human being, since the beginning of modern science/medicine, was no longer regarded as organism but as machine – this is the ontological *volte face* which underpins the Scientific Revolution itself – see Lee (2005, 2012).

A machine is a human artifact, made up of parts, specifically designed, constructed, and put together in order to help its creator to achieve a certain goal. A car is paradigmatically such a machine. It is the ontological contrast of organism, as organisms (in the history of their evolution in Nature) are simply the end results of Natural Selection involving a long and complicated process of the interaction between the organism with its genetic inheritance and the environment. A machine is peculiarly unproblematic both from the epistemological and ontological points of view – as it is a human creation, we humans, necessarily know precisely how to construct and deconstruct it in a straightforward manner. When we dismantle a watch into its parts, we are dismantling a whole into its parts, such that it is obvious, a whole is no more than the sum of its parts; we may then put the parts together again, and nonmysteriously, the whole appears again in front of our very eyes. If the universe and everything it contains is nothing but machine, then the universe no longer poses mysteries for us humans, which, we, over time and with assiduity, could not unravel and deconstruct at will. In this way, we, moderns, can leave the

obscurantist philosophy of Aristotle about organisms behind, as organisms have now been revealed to be *au fond* nothing but machines.

When this kind of world-view is then applied in the domain of medicine, illness is perceived to be mal-functioning of one or sometimes more than one of its parts, in the same way that when a watch today stops working, we diagnose that its battery has run out, we then open it up, we remove the exhausted battery and put in a new one, and we can immediately see that the watch starts to tick again. When a scan or two, reveal that it is the kidneys which are diseased and therefore not functioning, we then do a kidney transplant, and the patient, to all intents and purposes, starts to live a normal life again. In theory, one can conceive of a patient surviving with all the major organs being transplanted organs, although in practice, as far as one knows such a feat of re-engineering has not yet been accomplished. Bones could be replaced, such as in the case of hip replacements. The research programme in medicine behind which stands the ontological view that the human organism is machine, remains, today, a fruitful one, holding out further promises of success, especially with the help of more recent technologies such as IT, nanotechnology, biotechnology, and others. Success is sweet for all parties concerned, practitioners and clients alike. However, it remains fair to observe that the recipients, grateful though they undoubtedly are, know that the rest of their lives is dependent on drugs, some of which can have some disturbing side effects.

Conclusion

The subject matter raised by the title of this essay is immensely large and complicated; as one cannot do justice, here, to all aspects, one has been highly selective and focussed on only a few for limited discussion. From that discussion, several points appear to have emerged:

1. The trend of machines replacing humans in medicine and the health system will continue as such a cost-saving imperative is axiomatic to the dominant model of economics and accountancy prevailing today.
2. Machine diagnosis of illness is already a part of medical culture, although it has not so far entirely diminished nor rendered superfluous the role of doctors in the actual practice of medicine. However, in principle, there appear no inherent difficulties in developing in this direction. As the cost of training a doctor is high, the logic of economics and accounting may well point to the day when fewer doctors worldwide might be considered to be required to keep up with the same level of health care.
3. The ordinary person in the street and potential/actual patient appear not to be too concerned with 2 as far as machine diagnosis is concerned provided the diagnosis yielded is as good if not better than that provided by the average doctor in the average surgery or hospital. In other words, although machine diagnosis in the abstract may appear to constitute a threat to the individual in depersonalized

context, nevertheless, in practice, it may not be perceived to be so threatening, as the patient is primarily interested in a correct diagnosis and as quickly as possible.

4. The ordinary person may turn out not to be too concerned, if at all, with the existential threat involved in the ontological *volte face* that human beings are nothing but machines, as long as they can survive with the aid of a medicine based on such a *volte face*, and be able to lead an existence with sufficient quality of life to it.
5. However, the ordinary person does not or may not find congenial the dehumanizing effects of machines displacing humans in the larger context (not the narrower/restrictive one of diagnosis) of the doctor–patient relationship and the nurse–patient relationship, especially in the latter domain, when patients expect a human-to-human relationship where care and compassion may obtain which can console, comfort, and ameliorate the suffering of the sick. In other words, the patient is not simply a diseased organ, a fractured leg, a dicey heart, a ropey kidney, a peptic ulcer, but a person with emotions, feeling pain and so on, not a malfunctioning machine whose defective parts could be technologically replaced or repaired.
6. The trend in the care of the elderly appears to be pointing in the direction of robotics. The sick or the elderly are psychological/social beings who happen not to be well and/or frail. If robots are to replace human carers, then it is imperative that robots become humanized, as Paro, the baby seal appears to demonstrate. However, this recognition is telling as it is nothing but the recognition that technology/machine is in principle dehumanizing in the context of medicine and health system.

References

Atkins FW (1984) The second law. Scientific American Books, New York

Ballantyne GH (2002) Review of early clinical results of telerobotic surgery. http://cmaps.cmappers.net/rid=1HZ2RWKZY-1Y1GHF0-G08/ResearchPaper6.pdf

Baumol WJ, Blackman SAB, Wolff EN (1989) Productivity and American leadership: the long view. MIT Press, Cambridge, MA

Brynjolfsson E, McAfee A (2011) Race against the machine: how the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming the employment and the economy. The MIT Center for Digital Business, Cambridge, MA

CAT scan (2014) http://www.nhs.uk/conditions/ct-scan/pages/introduction.aspx. Accessed 21 Aug 2014

Davies H (1980) George Stephenson. Hamlyn Paperbacks, Middlesex

Dethlefs N, Martin B (2006) Japanese technology policy for aged care. Sci Public Policy 33 (1):47–57. https://www.uow.edu.au/~bmartin/pubs/06spp.html

Dreizen P (2004) The Nobel Prize for MRI: a wonderful discovery and a sad controversy. Lancet 363(9402):78. Accessed 21 Aug 2014

Evolution of mass production (2014) Ford Motor Company. DIALOG. http://www.ford.co.uk/experience-ford/heritage/evolutionof mass production

Field A (2008) Productivity. The Concise Encyclopedia of Economics. Available via DIALOG. http://www.econlib.org/library/Enc/Productivity.html#abouttheauthor. Accessed 21 Sept 2014

Green R (2005) Blood tests. http://www.netdoctor.co.uk/liver_kidney_urinary_system/examinations/bloodsamples.htm
Hafner K (2012) For second opinion, consult a computer? http://www.nytimes.com/2012/12/04/health/quest-to-eliminate-diagnostic-lapses.html?pagewanted=all&_r=0
High technologies timeline (2014) Greatest engineering achievements of the 20th century. National Academy of Engineering. DIALOG. http://www.greatachievements.org/?id=3824. Accessed 21 Aug 2014
Hudson A (2013) A robot is my friend: can machines care for elderly? BBC News. http://www.bbc.co.uk/news/technology-24949081
Ihde D (1993) Philosophy of technology: an introduction. Paragon House, New York
Kelly H (2013). Robots: the future of elder care? http://whatsnext.blogs.cnn.com/2013/07/19/robots-the-future-of-elder-care/
Kluwer W (2014) Wolters Kluwer Health Q1 Poll: Self-Diagnosis. http://www.raredisease.org.uk/documents/RDUK-Family-Report.pdf
Laparoscopy (keyhole surgery) (2014) http://www.nhs.uk/Conditions/laparoscopy/Pages/introduction.aspx
Lee K (2005) Philosophy and revolutions in genetics: deep science and deep technology, 2nd edn. Palgrave Macmillan, Basingstoke
Lee K (2012) The philosophical foundations of modern medicine. Palgrave Macmillan, Basingstoke
MRI (2014) http://inventors.about.com/od/mstartinventions/a/MRI.htm. Accessed 21 Aug 2014
Mumford L (1946) Technics and civilization. George Routledge & Sons, London
Nash DB (2010) Isabel, a new diagnostic aid for the 21st century. Pharm Ther 35(12):651. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3008375/
Paro therapeutic robot (2014) http://www.parorobots.com/
Pillinger J (2014) Urine test. http://www.netdoctor.co.uk/health_advice/examinations/urinesample.htm
Reiser SJ (1978) Medicine and the reign of technology. Cambridge University Press, Cambridge
Rifkind J (2005) The end of work. http://www.foet.org/press/interviews/Spiegel-%20August%203%202005.pdf
Rotman R (2013) Destroying jobs. http://www.technologyreview.com/featuredstory/515926/how-technology-is-destroying-jobs/
Rull G (2012) Biopsy. http://www.patient.co.uk/health/biopsy
Smith A (1776/1976) Of the Division of Labour. In: Campbell RH, Skinner AS, Todd WB (eds) An inquiry into the nature and causes of the wealth of nations. Clarendon Press, Oxford

Professionalism in Health Care

41

Andrew Edgar

Contents

Abstract

This chapter will explore professionalism historically, from the work of Gregory and Percival in the eighteenth century to contemporary "new professionalism." The chapter will identify how the core traditional values of professionalism, in particular commitments to an other-regarding social ethic and to maintaining high levels of scientifically informed expertise, alongside the defense of professional self-regulation, have been articulated and challenged. Classic accounts of professionalism are found in the work of Durkheim, Tawney, and Parsons. Critics have argued professionalism is in practice self-serving, particularly insofar as a professional ethic has justified the autonomous self-regulation of the profession. Over the last 30 years, responses to the perceived crisis of professionalism – due to the loss of broad public trust in the professions, changes in the nature of professional expertise, and increased demands for external regulation – have precipitated a series of more or less radical responses.

A. Edgar (✉)
Cardiff University, Cardiff, UK
e-mail: edgar@cardiff.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_30

New professionalism has now begun to question the desirability of professional autonomy and self-regulation and to articulate a professionalism committed to public engagement and the acceptance of external regulation.

Introduction

"Professionalism" may be understood as the set of competences or virtues that a practitioner is expected to manifest insofar as they are a member of a profession. Such virtues would typically be thought to include an altruistic concern for the best interests of the patient or client over and above those of the professional him- or herself and a commitment to maintaining high levels of training, expertise, and competency in the exercise of professional skills. Broadly, "professionalism" is then the quality of being a good professional, where goodness is understood both, morally, in terms of the professional's relationship to their clients and to a wider public and in terms of the sustaining of appropriate expertise and technical competence. As such, understandings of professionalism will have important consequences for the education and training of the professional and for the regulation of professional practice, being articulated in codes of conduct.

Within this broad definition, there is considerable scope for debating its substantial content. The precise interpretation of what professionalism entails will vary from profession to profession and even within a profession such as medicine, between such subdisciplines as general practice, nursing, and psychiatry. The substantial understanding of professionalism changes as professions face diverse pressures, both from without, for example, as legal and political environments change, the assertiveness of clients develops, or as markets intrude upon professional practices, and from within – as professional practices themselves develop and diversify. This chapter will therefore explore diverse interpretations of professionalism by using the framework of a historical review of the development of the medical profession, recognizing how this history is itself entwined with reflection, by both practitioners and academics, on the nature of the profession and processes of professionalization.

Eighteenth-Century Origins of Professionalism

In eighteenth-century Britain, medicine, alongside the law and the church, may be seen to have already established itself as professions. As such, the physician had a relatively high social status in a highly stratified society. Crucially, to be recognized as a practitioner of a profession raised one above the mere status of trade or craft. The professional was a member of "polite society" and thus the social equal of their middle-class clientele. As Leake characterizes the situation, the eighteenth-century physician "disdained work [and] condescended to help sick people," taking no fee, although a guinea would be left for them "as they withdrew" (1970, p. 68).

Thus, in contrast to trade, which is pursued instrumentally in order to earn one's living, pursuing a profession implied some notion of vocation or calling.

A professional was distinguished from a mere craft or trade by both the more theoretical nature of the professional's expertise and by the requirement for the professional to exercise judgment in applying that expertise. While craft or trade might presuppose considerable manual skill, professional practice required theory and phronesis. Thus, physicians received a university education (preeminently then in Leiden or Edinburgh) embracing chemistry, surgery, anatomy, medical theory, and clinical practice. The eighteenth-century medical education also incorporated the pedagogical technique of "walking the wards" (Lindemann 2008). That is to say that the physician learned their profession, not merely from theory nor like the craft worker through application of learned rules of practice, but rather through exposure to diverse examples of real patients. Such exposure developed a subtle and contextually sensitive clinical judgment. In contrast, apothecaries, surgeons, and barber surgeons, as trades, continued to be trained through apprenticeships, within the remnants of the medieval guild system (Lindemann 2008).

The complexity of professional knowledge and practice, even in the eighteenth century, is already such that the lay person is not readily able to assess the success or efficacy of professional practice. While the efficacy of a trade or craft worker can typically be assessed by anyone – for the potter's cup will hold liquid, the carpenter's chair will be strong and comfortable, and the wheelwright's spokes are sturdy enough for the dirt road – that a patient is not cured may not be evidence of the failure or incompetence of the physician, and the accuracy of a diagnosis is not readily judged by any but other physicians (see Edgar 2011). The patient is thus required to trust the physician in a way that they do not trust the potter, carpenter, or wheelwright.

While the distinction between professions and trades begins to articulate the epistemological grounding of professionalism, it is in the social status of the profession, and in the social relationship of the professional to their client, that the moral dimension of professionalism emerges. Even though the eighteenth-century professional is already aspiring to a degree of knowledge that is largely incomprehensible to their lay clientele, the physician does not stand above their patient as a man of science. The scientific authority of the professional is not yet so universally accepted as to command, on its own, the attention and obedience of the client. Indeed, to assert scientific superiority would be to violate the etiquette and mores of "polite" middle-class society. Rather, in acquiring middle-class status, the professional is placed in a precarious equality with their clientele (see Porter 1997, pp. 255–258 and 281–287). It is precisely this social equality, and not science, that secures a relationship of trust between the physician and the patient, and thus the patient's obedience to the physician's instructions.

The first overt guidance on medical ethics in the modern literature, Thomas Percival's *Medical Ethics*, was published in 1803 (Percival 1849). This may be understood in some part as a response to this issue of patient trust. The Hippocratic Oath had served as a basis for medical ethics and the moral self-understanding of physicians since the fifth century BCE. Percival is concerned with the situation of

the modern physician. He thus reinterprets the spirit of the Hippocratic Oath in terms of the medical profession as a modern guild, where the profession is in a compact with the general public. "Every man who enters into a fraternity engages by a tacit compact not only to submit to the laws, but to promote the honour and interest, of the association, so far as they are consistent with morality and the general good of mankind" (Percival 1849, part I §22). As such, an ethos of self-regulation is already placed at the core of professionalism, with the individual practitioner under an obligation to maintain the public reputation of the profession as a whole. *Medical Ethics*, in consequence, outlines the knowledge and awareness that the physician should have that goes beyond mere biomedical training. While issues such as the law (including the physician's responsibilities with respect to dueling) and religious sensibilities are discussed at some length, Percival's primary focus is on matters of etiquette, in the physician's relationship both to patients and to fellow practitioners. (The book's original motivation was to clarify the relationship between physicians, surgeons, and apothecaries in his own Manchester hospital – where the latter were struggling for professional recognition.) It is thus in the observance of professional etiquette that the patient's trust is secured, rather than in the assertion of scientific competence. The concern with etiquette does indeed allow Percival to begin to articulate issues of genuine ethical concern, such as patient confidentiality, albeit that critics suggest that he fails effectively to separate the two and thus fails to recognize the greater importance of ethics over etiquette (see Leake 1927, pp. 2–3).

John Gregory's *Lectures on the Duties and Qualifications of a Physician* was delivered, while he was chair of physic at Edinburgh between 1766 and 1773 (Gregory 1817). Gregory is more overtly concerned than is Percival with medical malpractice and incompetence. While celebrating the breadth of knowledge, and indeed genius, required of the physician, he recognizes the uncertainty of the judgment of the individual physician. He argues that there is "no established authority to which [physicians] can refer in doubtful cases. Every physician must rest on his own judgment, which appeals for its rectitude to nature and experience alone" (Gregory 1817, p. 17). In response to this, Gregory stresses the urgent need to establish objective criteria for the assessment of medical practice and the mechanisms of scrutiny that will enforce compliance with them. The problem lies, in part, precisely in the exclusion of the lay person from any understanding of medical science. "The science of medicine alone is kept so carefully concealed from the world, and the art must necessarily be practised in so private a manner, as renders it difficult for the public to form a just estimate of a physician's knowledge from the success of his practice" (Gregory 1817, p. 210). Gregory's solution to this problem is not merely to strengthen the rigorous scientific basis of medicine, thereby anticipating modern appeals to the importance of epidemiology and evidence-based medicine, but also to ensure that physicians are open about their mistakes and failings with the public and crucially to educate that lay public in medicine, so that they may fairly judge medical practice (see Boyd 2005). Part of the ethics of the physician thus lies in their obligation to communicate clearly, honestly, and effectively beyond the narrow limits of the profession itself (see Gregory 1817, pp. 187–188).

In summary, the eighteenth century sees not merely the modern profession taking shape but also the outline of disputes over the nature of professionalism that is still current. Percival's ethics begins to articulate the ideology of professional self-regulation. While Percival focuses on the personal relationship between the physician and patient as guarantor of trust, Gregory begins to pose critical questions about the actual self-serving nature of such autonomy and the potential that the ideology of professionalism might have to conceal malpractice. The trust that the profession requires from the public is thus grounded, neither in ethics nor in an authoritarian appeal to arcane knowledge, but rather in public engagement and dialogue.

Civic Professionalism

The British Medical Association (BMA) was founded in 1832 (initially as the Provincial Medical and Surgical Association) and the General Medical Council (GMC), which has responsibility for maintaining a register of doctors and for education and training, in 1858. In 1847 the American Medical Association (AMA) was founded. These institutions may be seen to consolidate the paternalism and autonomy of the profession; precisely insofar the professional association takes over, from the state, the legal responsibility for regulating practitioners. In the USA, this regulation was, from the first, grounded in the adoption by the AMA of a code of ethics. This code was based upon Percival's *Medical Ethics* and covered the duties of the physician to the patient, to other professionals, and to the public (wherein the physician has a role as a "good citizen," using their medical expertise to advise on matters of public health) and also the duties of patients and the public to physicians (see Baker 1995, pp. 75–87). The GMC, while responsible for disciplining doctors found guilty of improper conduct, actively resisted the adoption of a code of ethics, turning instead to jurisprudence, and as such an approach grounded in common law. The limits of acceptable practice, in the British context, thus came to be established through case law (see Crowther 1995).

The AMA exemplifies the profession's moral responsibility for safeguarding the interests of patients and a wider public when, in the early twentieth century, it critically investigates and transforms medical education. Crucially, the early regulation of the medical profession, and thus the emerging definition of professionalism, rests not merely on the moral behavior (or even etiquette) of the practitioner but also upon the scientific rigor of their expertise. Percival had railed against quackery (1849, ch. 2 §21), as had Gregory (1817, p. 124). The professional medical bodies thus sought to undermine the claim of "irregulars," such as homeopaths, to the title of doctor or physician, as well as ensuring that those legitimately claiming the title were properly educated. In the second half of the nineteenth century, the AMA worked to consolidate medical education as grounded in laboratory work, clinical instruction, and an extensive program of lectures, so that medical training would cease to be a mere apprenticeship (Baker 1995, pp. 14–15). In the face of continuing evidence of widespread quackery and poorly or even uneducated doctors, the AMA created the Council for Medical Education in

1904 with the objectives of establishing minimal standards for acceptance into medical schools and determining medical school curricula. Abraham Flexner's survey of medical school, *Medical Education in the United States and Canada*, was published in 1910. This led to the eventual closure of proprietary schools that existed merely to provide profits to their owners (frequently awarding "honorary" medical degrees) and to the integration of medical schools into universities. In addition, a medical curriculum based upon 2 years of scientific study and 2 years in teaching hospitals (reflecting the practice of "walking the wards") was established and state licensing of physicians negotiated. The regulatory bodies of the medical profession may thus be seen to be protecting patients from failures in the medical market. Lacking the necessary expertise to judge between good and bad physicians or indeed physicians and quacks, the professional body steps in, paternalistically, to control the market in physicians.

The philosophical articulation of this development of civic professionalization – specifically a view that sees the profession as occupying a crucial place in civil society and thus inculcating a sense of public service into the practitioner – may be found in work of sociologist Emile Durkheim (1992) and historian R. H. Tawney (1921). Both present professionalism as a solution to social ills arising from the advance of liberal capitalism. For Durkheim, this is the anomie, the loss of moral values and meaning, brought about by an advanced division of labor. For Tawney, it is the mistaken sense that the acquisition of property is a good in itself, rather than something to be judged by the benefit that it offers to society as a whole.

Durkheim's lectures on "Professional Ethics," originally delivered between 1890 and 1900, develop themes already explored in his *The Division of Labour* (1984). *The Division of Labour* argues that while the advanced division of labor characteristic of modern industrial societies was necessary and advantageous to economic prosperity, it had the disadvantage of fragmenting social solidarity, leaving the individual member of society increasingly isolated, with little or no sense of communal belonging or moral orientation. Individual self-interest trumps collective morality. The modern state is seen as too large and bureaucratically impersonal a body to instill communal identity in its citizens. Occupational groupings, akin to Roman and medieval guilds, are thus proposed as a check against both the impersonal and distant state and the individualism and self-interest of the market. While the economic exchanges and the practice of business encourage the individual to think only of what is in their own interest, professions offer to their practitioners the pursuit of purposes – social functions – that have merit beyond individual gain. As such, membership of an occupation restores a sense of purpose and collective identity in the face of anomie. While each occupation has a particular social function and thus, Durkheim argues, its own morality, membership of the occupational group will not merely inculcate a sense of internal group solidarity, but will have ramifications for the practitioner's relationships with those outside the group. The virtues learned within the occupational group blossom in a general sense of social solidarity.

Durkheim here articulates an ideal of professionalism whereby the profession stands as a bulwark against the alienating encroachment of both market and state. The professional pursues their occupation not for pecuniary reward, but rather from

a motivation, or indeed a calling, to serve society. The implication is that professional bodies, such as the AMA or BMA, should constitute themselves as modern guilds. However, it may be noted that Durkheim is not strictly theorizing "professions" in the English sense of the term. The French term refers to any occupation and not merely to the *professions liberals* (Freidson 2001, p. 53). Durkheim's call is for a fundamental reorganization of all occupational groups.

Tawney takes the profession, in the English sense of the term and as exemplified in medicine and law, as a model for occupational organization. All occupations benefit from professionalization. The profession is not merely an aggregate of workers all pursuing the same occupation nor even a trade union, protecting its members' economic interests, but rather it "is a body of men who carry on their work in accordance with rules designed to enforce certain standards both for the better protection of its members and for the better service of the public" (1921, p. 106). The profession meets a social function, but as such gives its members a sense of purpose beyond mere pecuniary gain. To practice a profession is to know that one is responsible to some "higher authority" (p. 14). Members within a profession compete with each other, not for financial or material reward, but rather for honor and reputation. So the professional will not perform certain acts (such as the sale of patent medicines (p. 109)), harmful to the client or wider public, no matter the potential financial reward. Public and professional service is thus put before personal interest.

The model of professionalism that emerges from Durkheim and Tawney may also be seen in the work of pragmatists and reformers such as John Dewey, Jane Addams, and Herbert Croly. The civic professional becomes a new type of hero, providing a model of selfless civic service to ordinary people. The professions use their scientific knowledge in improving social life, but not for the amoral and pecuniary ends encouraged by capitalist markets (see Light 2010, pp. 273–274). The model receives a more complex articulation in Talcott Parsons' structural-functionalist sociology. Parsons is specifically concerned with the medical profession, taking it as paradigmatic of modern professionalism (1951, pp. 288–323). His analysis, while highly theoretical, is grounded both in empirical study and in a significant degree of personal sympathy with medicine as a vocation. (He originally had some intention to follow the example of his brother and to train as a physician.) At the core of his analysis lies the assumption that professions fulfill core functions for a modern society. All societies, within Parsonian theory, require certain functions to be fulfilled in order to stabilize and reproduce themselves (such as the production of the means of subsistence, the education or socialization of children, and the maintenance of political and legal order). Illness and disease pose fundamental problems both to the individual and to society as a whole. The ill person cannot continue their everyday activities and, as such, cannot fulfill the functions (or "social roles") that society expects of them. Illness is thus a form of social deviance which the medical profession controls and corrects (Parsons 1951, pp. 288–289). The professionalism that is expressed in scientific expertise and certain standards of public and private morality facilitates the realization of this function.

The bold outline of Parsons' structural functionalism omits much of the subtlety of his analysis and not least his sensitivity to illness as a culturally embedded and interpreted phenomenon, emotionally affecting and meaningful to both the patient and the physician. The patient is vulnerable and frequently fearful (Parsons 1951, p. 300). Medical examination and treatment requires intrusions into aspects of the patient's life that are usually kept private, including the eliciting of personal information and intimate physical contact, both of which contemporary Western culture treats as potentially embarrassing or compromising (p. 309). The patient's lack of knowledge of medical science and the uncertainties of their diagnosis and potential recovery may lead to seemingly irrational behavior and magical beliefs (of which the physician may have to be tolerant, if such beliefs aid recovery) (p. 315). Equally physicians themselves are potentially affected by the emotional strain of treating certain patients (and Parsons gives the example of a surgeon's relationship to a 9-year-old (p. 308)) and must protect themselves from this exposure. Further, while rigorously trained, the physician is still confronted by the frustrations of uncertainty in diagnosis, treatment, and prognosis and thus in the limitations of medical science (p. 302). Finally, appreciative of the insights of psychoanalysis and other forms of psychotherapy, Parsons highlights the complex dependencies and vulnerabilities that lie in the relationship between patient and physician (see pp. 304–305).

Society responds to the need to develop an effective means to cope with illness, in the face of these cultural and emotional pressures, through the institution of the sick role and the physician role. Social roles may be understood as patterns of behavior, bound up with certain obligations and privileges. In adopting the "sick role" of the patient is permitted to relinquish many of the activities and duties they normally must pursue, such as work, but with the reciprocal obligation to strive for recovery. The complementary role of the physician entails an obligation to do their best to aid the patient and not to exploit the patient's vulnerability. In return the physician enjoys significant social status and prestige. The physician thereby exemplifies a set of qualities that characterize "the "professional" pattern in our society, namely, achievement, universalism, functional specificity, affective neutrality and collectivity-orientation, in that order" (p. 305). It is precisely the practice of these professional qualities that allows the physician to fulfill their social function. In explicating what Parsons means by these terms, his ideal type of the professional will become clear.

The achievement orientation of the professional entails the grounding of their practice in a rigorous scientific knowledge base. To claim that professionalism is universalistic, rather than particularistic, is in part to acknowledge the universality of legitimate scientific inquiry. Parsons notes the initial rejection of Pasteur's discoveries by the medical profession, for he was a mere chemist, as a clear violation of such universalism (1951, p. 306). The recruitment and registration of professionals is similarly universalistic in being meritocratic, thus avoiding the particularism of, say, nepotism. Parsons notes that particularistic forms of recruitment may strengthen group solidarity but will weaken social solidarity, as in-groups are set against each other (p. 306). While this may be taken as a comment on and

reinforcement of Durkheim's account of solidarity, it also highlights the need for the profession to be oriented to the needs of society as a whole and not to its particular interests. Universalism does not however entail that professionals are generalists or "wise men" (pp. 292 and 306). This may be seen in the legitimation of the obligations attendant on the sick role. The patient must follow the instructions of the physician. Yet the physician lacks any formal sanction by which they can enforce compliance. The specificity of their expertise is thus significant precisely in that it legitimates the physician's claims upon the patient. The physician acquires not a generalized authority over the patient but an authority to ask specific things of the patient, in the interests of their health (p. 307). A specific expertise is thus part of the grounding of the patient's trust in the physician, not least insofar as it clearly articulates the nature and degree of the physician's legitimate intrusion across emotional and symbolically sensitive boundaries of personal privacy and decorum. The affective neutrality of the professional develops upon this, as the professional distances their practice from their personal feeling, treating patients irrespective of personal preferences, likes or dislikes, or moral judgments. In addition, this entails that the trust of the patient is further secured in that the profession is overtly working for the patient's interests and not their own (p. 308).

It may be noted that Parsons explicitly places "collectivity-orientation," which is to say the sense of civic duty of the professional, suppressing their self-interest in favor of patient and public interests, last. This orientation differentiates the professional from the commercial entrepreneur. The social role of entrepreneur positively sanctions self-interested behavior. In medicine this differentiation from commercialism is, Parsons suggests, of fundamental importance. In other professions, such as law and engineering, the relationship of the professional to commercial activity may be significantly more ambiguous (Parsons 1939, p. 458). Not so medicine. Parsons observes that US physicians at the time of writing were prohibited from various forms of commercial activity, such as advertising. Similarly the physician cannot refuse a patient on the grounds that they are a poor credit risk. The implication is that these are, as much as anything, symbolic legal prohibitions, expressing something fundamental about the nature of professionalism (Parsons 1951, p. 312). The functional importance of an overt "collectivity-orientation" lies in the vulnerability of the patient. Ignorant of their own condition and of its effective treatment and given the severity of the consequences of a mistake in choosing appropriate medical care, the typical advice offered to a consumer of "caveat emptor" cannot apply. The physician is in a dominant position and potentially able to exploit the patient. This potential must be suppressed in order to secure the relationship of trust between physician and patient (pp. 311–312).

The collectivity-orientation suggests the civic ethic with which Durkheim and Tawney characterize professionalism. However, while they tended to see this precisely as an ethic and thus as a moral culture inherent to professionalism, Parsons is more skeptical. While he entertains the possibility that people of a generally altruistic motivation are attracted to medicine and repelled by business, it is more important that the institutional structures of the medical profession negatively sanction self-interested behavior and reward altruistic behavior (Parsons

1951, p. 318, 1939, pp. 465–466). Thus, an institutionally well-ordered profession, which socializes students appropriately and, perhaps more importantly, that refuses to reward dishonest, self-serving, and otherwise unprofessional behavior, forces its members to behave as if they are altruistic, regardless of their personal and psychological motivation.

In summary, the work of Durkheim, Tawney, and Parsons may be seen to articulate the medical profession's self-understanding in a "golden age" of professionalism. Trust is sustained between the professional and their client, grounded in the professional's commitment to both sustaining the scientific expertise that informs their practice and maintaining high standards of other-regarding moral behavior. Professionals are responsible enough to regulate themselves and thus ensure that standards of expertise and morality are maintained. Parsons' model of the professional is presented as an ideal type and as such a heuristic to guide sociological research. This highlights a certain ambiguity in the literature as to whether the type is a normative ideal to which the professions ought to aspire or an empirical description of current professional practice. Parsons observes that: "It is true that medical associations do have committees on ethics and disciplinary procedures. But it is exceedingly rare for cases to be brought into that formal disciplinary procedure" (1951, p. 316). In part this suggests that individual professionals are self-regulating and that the institutional sanctions work effectively. More subtly and problematically, however, Parsons notes that the strict enforcement of professional standards by formal disciplinary committees would introduce significant strains and conflicts into the profession. Professionals would, for example, be required to testify against each other, and to expose their failings to the public. The implication of Parsons' argument appears to be that it is better – or more functional – for professional misdemeanors to be dealt with informally and out of public view than to risk public trust through formal disclosure. It is here that skepticism over the role of professions, and indeed the ethos of professionalism, begins to emerge.

Professional Dominance Theory and Deprofessionalization

In Act 1 of his 1906 play, *The Doctor's Dilemma*, George Bernard Shaw's character Sir Patrick Cullen – an aging doctor and teacher, "not yet quite at the end of his tether" – remarks that: "All professions are conspiracies against the laity." Shaw himself remarks in the substantial preface to the play that the "medical profession [is] a conspiracy to hide its own shortcomings" (Shaw 1909). Shaw offers a vigorous attack on the medical profession, questioning its overt altruism as being little more than a convenient veil that prevents the public from recognizing a multitude of professional shortcomings – including the performance of unnecessary operations for the sake of the fee. Significantly Shaw is not launching an attack on the personal morality of physicians but rather upon institutional factors, such as low pay, that encourage, or indeed necessitate, immortal practices. Shaw thereby anticipates the core of the critical approach to professionalism that was to emerge

in the 1970s, when the Parsonian assumption that the professional is collectivity-orientated comes to be questioned by both sociologists and historians. This period also begins to mark the decline in public trust in the professions, and academic research may be seen both to stimulate and to give voice to that growing distrust.

Eliot Freidson's *Profession of Medicine* (1970) argues that professionalization is a fundamentally political process. Freidson's professional dominance perspective argues that professionalization is the political process through which a high degree of autonomy is secured for the profession. Autonomy allows the profession to regulate itself, identify and discipline malpractice, and determine the most appropriate form that the provision of its services should take. While the Parsonian perspective suggests that the legitimacy of this autonomy is largely self-evident, given the functional value of the profession to contemporary society, for Freidson "[a] profession attains and maintains its position by virtue of the protection and patronage of some elite segment of society which has been persuaded that there is some special value in its work" (Freidson 1970, p. 72). The profession must make its case, not least to the state and the general public. The trappings of professionalism, including the appeal to expert knowledge and ethics, serve to make this case. Insofar as the complexity of the knowledge base is comprehensible only to the trained physician, no one outside the profession is in a position to criticize its practice. More subtly, if the codes of conduct that govern professional practice are more than just etiquette – more than the maintenance of a polite and financially lucrative relationship to the patient, alongside stable relationships within the profession – being rather genuine ethics, grounded in an ethos of public service, then again external legal regulation of the profession is unnecessary. The skeptic, in making their critical argument against professional autonomy, will question both of these assumptions, reducing professionalism to a rhetoric.

For the skeptic, professional autonomy in medicine is not seen as the necessary facilitation of benevolent and paternalistic action toward patients and the general public but rather as a means of securing market dominance over health-care provision. Autonomy works ultimately in the self-interest of the profession, not the general public, by securing the profession a more or less monopolistic position within the health-care market (Berlant 1975; Bledstein 1976; Larson 1978). Autonomy thus gives the profession an economic advantage in that it can exclude competing occupations from entering the health-care market, thereby increasing the power of the profession to determine fees (Elston 1991). In the nineteenth century, the AMA is seen to work, actively, to exclude alternative or "irregular" practitioners such as homeopaths, from the market (Starr 1982). It may be noted, as a problem with such an argument, that much contemporary sociology of medicine tends, unlike Parsons, to affect an agnosticism toward issues of medical efficacy. If homeopathy and other alternative medicines are genuinely ineffective, and potentially harmful if prescribed for serious conditions, as contemporary evidence-based medicine strongly suggests, then the appeal to medical expertise made by the AMA is more than simply a matter of monopolizing a market. It has a genuine ethical dimension in protecting the public from harm. A stronger case is made by critics of professionalism with respect to the relation between medical subdisciplines:

disciplines such as nursing may be denied full professional status as the medical division of labor is controlled by physicians (Freidson 1970, pp. 57–63). It can be also argued that medical autonomy has restricted the forms of effective medicine being made available. Autonomous and self-regulating professionals come to be suspected of providing the services they want to provide, rather than those that patients need. As gatekeepers to medical services, they are criticized as being nonaccountable. Profitable curative approaches, for example, have been promoted over and above preventative medicine (see Light 2010, p. 276). More subtly, a monopolistic medical profession can have undue influence in determining public understandings of health and health care. The sick role itself is thereby revealed as a site of political negotiation, as the legitimacy of the state of a particular condition, such as chronic fatigue syndrome or degrees of mental health, as illness is contested and constructed.

Freidson summarizes the problems of professionalization by claiming that: "While the profession's autonomy seems to have facilitated the improvement of scientific knowledge about disease and its treatment, it seems to have impeded the improvement of the social modes of applying that knowledge" (1970, p. 371). Freidson acknowledges that, historically, the medical profession did need protection from "the urgent ignorance of its clientele [and] the mischief of low-class competitors" (ibid), but that autonomy has now led to a destructive degree of complacency in the profession as it isolates itself from external criticism. Crucially, it loses sight of the patient's perspective. Most fundamentally this is expressed in a failure to identify poor and negligent practice. In the UK, in the 1970s, evidence of medical malpractice and poor standards was becoming more public (RCGP 1974) and yet was receiving no official response from the GMC and other professional bodies. The Merrison Inquiry into the regulation of the medical profession, commissioned by the Secretary of State for Social Services in 1972, made no reference to the evidence that had been presented to it concerning poor standards of practice (Secretary of State for Social Services 1975). Self-regulation was failing to protect the patient or public. Freidson argues that this failure is rooted in a reluctance by professions to criticize each other, thereby mirroring Parsons' analysis. Mistakes are regarded as inevitable in complex practices, and while self-criticism may be encouraged, the open criticism of others violates a requirement for mutual charity. It may after all be the critic's turn to be the subject of criticism next (Freidson 1970, p. 179). This suggests that the institutions to which Parsons appealed in order to secure the altruistic behavior of professions are actually fundamentally flawed and work against the interests of patients and public.

For some (Light 2010, p. 272), an irony of Freidson's argument lies in its publication occurring just as the dominance of the medical profession is in decline. While, as Freidson readily admits, the medical profession's autonomy was never absolute, it has been argued that changes within the provision of medicine and the organization of the medical subdisciplines as well as changes in wider society have begun to erode the power that the medical profession exercised during its "golden age" (roughly 1945–1965). At the extreme, this has been argued to constitute a

deprofessionalization of medicine (Reed and Evans 1987). Critical changes include the intrusion of profit-making organizations into the medical market in the 1980s (e.g., in the USA an increase in for-profit health care and more stringent financial management of health-care provision and the introduction of quasi-markets into the organization and ethos of the UK National Health Service), whereby decisions on the provision of health care are shifted away from the physician. Internal changes within medicine have compounded or been entwined with this shift. These include the successful struggles of nursing and other "professions allied to medicine," such as physiotherapy, to assert their own professional identity. For Freidson (1994), the rise of the professional manager, and thus the development of more finely structured hierarchies within medicine, similarly undermines the physicians' autonomy, as decision making over policy and even prescribing shifts to the managerial profession. Perhaps more fundamentally, the development of evidence-based medicine (and organizations such as the UK's National Institute for Health and Care Excellence (NICE) that offer guidance to governments and the medical profession on the efficacy of treatments), alongside the use of clinical governance and even clinical targets, has begun to challenge traditional notions of clinical judgment. The autonomy, not merely of the profession but more specifically of the individual practitioner, is compromised insofar as compliance with nationally agreed clinical guidelines for treatment reduces the need for phronesis, and the erstwhile professional becomes a mere technician.

There has been a significant loss of patient trust in the profession, aligned with an increasingly consumerist attitude on behalf of patients, thus leading to increased litigation and external investigation and regulation. The model of a passive and ignorant patient used by Parsons has been challenged. In part this is due to a general rise in levels of public education, so that the physician is more likely to be confronted by patients who are as well-educated and as articulate as themselves. Patient groups, supporting either sufferers of particular conditions or defending patients' rights in general, have demanded a voice for the patient in negotiating their treatment. Significantly, Parsons argued that a function of the medical profession was to keep patients isolated, in order, given the deviant nature of illness, to inhibit the formation of groups of deviants (1951, pp. 320–321). Such groups now play an important role in checking abuses of professional power. Talbot argues that the loss of public trust is as much rooted in a crisis over the behavior of physicians toward patients, including an increasing business orientation and loss of effective communication skills, as in a direct experience of poor practice (Talbot 2011, p. 127). This would suggest either a decline in the physician's standards of professionalism (expressed in less demonstrable respect for patients) or higher expectations of professional behavior from the public. With respect to a wider public, a number of well-publicized instances of malpractice in the UK and elsewhere, since the 1990s, including those of individual practitioners such as Harold Shipman, Richard Neale, and Rodney Ledward or institutional failings at Bristol Royal Infirmary, Alder Hey in Liverpool, and most recently the Mid-Staffordshire NHS Trust, have served to underline a public perception of the failings of professional self-regulation.

In summary, professional dominance perspectives share sociology's earlier assumption that the medical profession does have autonomy and considerable power. They differ from early perspectives in questioning the supposed altruism and public service ethos of the profession. Professionalism, expressed in a commitment to public service and the maintenance of a scientifically grounded expertise, is challenged as a mere rhetoric or ideology that conceals, consciously or otherwise, the self-serving nature of the profession. Growing public awareness of failures in self-regulation, alongside structural and cultural changes in the provision of health care, has led to the establishment of mechanisms of external regulation. Debates that began in the eighteenth century, with the concerns of Gregory over standards of professional practice and Percival's defense of a medical ethic, thus reemerge and pose a new challenge to develop a viable contemporary conception of professionalism.

New Professionalism

Hafferty (2006) suggests that initial responses, in the 1980s, to the perceived crisis in professionalism were restricted to a largely polemical defense of professionalism in the face of increasing commercialism. In the early 1990s, renewed attempts to define and operationalize "professionalism" arise, integrating notions of professionalism as a competence into physician training and accountability. This in turn leads to attempts to measure professionalism (see Arnold et al. 1998; Epstein and Hundert 2002). Understandings of professionalism are nonetheless diverse. Indeed, it may be argued that the values espoused by professionals continued, well beyond this period, to be highly ambiguous and contested (Pattison and Pill 2004). A continuum of responses may nonetheless be identified, running from a conservative reassertion of the traditional values of professionalism to a more reform-oriented "new professionalism," variously rethinking the demands of civic professionalism, embracing the need for external regulation and guidance, and thus abandoning professional autonomy as a defining characteristic of professionalism.

Conservative responses to the crisis presuppose the continuing need for professional autonomy. This is argued for, somewhat surprisingly, by Freidson in his final work, *Professionalism: The Third Logic* (2001). Reconsidering the golden age of professionalism and thus something closely akin to the Parsonian ideal type of a profession, Freidson places the logic of professional practice between that of the market, on one side, and bureaucratic planning on the other. He thus, in effect, restates familiar arguments that defend professionalism in the face of both the corrupting influence of commercialism (which would place self-interest above public service) and state regulation (which would undermine professional judgment). At the core of his argument lies a reassertion of the specialist and complex nature of professional expertise, characterized as it is by uncertainty and contingency. The nonprofessional, be this either the consumer in the marketplace or the civil servant, is then assumed to be unable to understand and judge good

professional practice. Monopoly and self-regulation are thus essential if professions are to continue to serve the public. While Freidson's argument has been rigorously challenged (see Hafferty et al. 2003), not least in that it fails to take account of the very institutional failures documented in his earlier work, similar restatements of traditional arguments abound.

Cruess et al. (2004), for example, articulates the familiar view of professionals using expertise to the good of society – governed by codes of ethics that commit them to traditional values of "competence, integrity and morality, altruism, and the promotion of public good" – in return for status and financial rewards, as a social contract. Swick (2000) similarly offers a normative definition of "professionalism" in terms of nine behaviors, focusing around the need to subordinate person interests to those of the patient, thus to act ethically and to demonstrate humanistic values such as compassion and integrity; professionals thereby respect the social contract between the profession and the public; they are committed to maintain high levels of technical excellence and to reflect upon their practice and be accountable to peers. Such behaviors would, Swick argues, restore public trust in the profession.

The conservative arguments perhaps do little other than repackage old ideals of professionalism. The debate does, nonetheless, lead to strategies for reinforcing and policing the ethical grounding of professionalism and thus strategies to make professional autonomy workable. Two prominent examples may be briefly reviewed. The American Board of Internal Medicine (ABIM), through its Medical Professionalism Project, developed the "Physician Charter" (ABIM Foundation et al. 2002). This widely adopted code of conduct articulates professional ethics by promoting "physician responsibilities" that include commitments to professional competence and scientific knowledge, honesty, and confidentiality. Further, patient autonomy and social justice are included in the governing principles of the charter. This is indicative of an awareness of the rise of the patient's rights movement and the implications that a social contract has in the context of debates over the rationing and prioritization of health care. Corresponding responsibilities thus include commitment to the quality of care, access to care, and just distribution of finite resources. Crucially the charter makes explicit issues that challenge or undermine professionalism. These include the abuse of power, arrogance, greed, misrepresentation, lack of conscientiousness, and conflicts of interest (ABIM 1995). At the very least, this suggests that professionalism may be as much recognized in its absence as in its presence but also entails responsibilities to managing conflicts of interest and a (perhaps somewhat vaguer) commitment to professional responsibilities.

The Accreditation Council for Graduate Medical Education (ACGME) takes the debate a step further by including "professionalism" as one of the six core competences that a physician requires and through which training and revalidation can be oriented. Professionalism requires the demonstration of respect, compassion, and integrity; responsiveness to patient needs superseding self-interest; accountability to patients, society, and the profession; excellence and ongoing professional development; adherence to ethical principles; sensitivity and responsiveness to diverse

patient population; and respect for patient privacy and autonomy (Swing 2007). "Professionalism" by this operationalization is complemented by competences in "medical knowledge" and "practice-based learning and improvement" – and thus the traditional commitments to maintaining scientific expertise – and by a competence in patient care, which includes effective communication, caring, and respectful behavior (ibid). The increased recognition of the importance of communication skills suggests that etiquette toward the patient, if not adopted cynically, has become an important aspect of professional conduct.

A number of concerns may be expressed with these traditional responses and in particular with the integration of professionalism into education and validation. Firstly, it may be noted that the treatment of professionalism as a measurable competence sees resistance from the likes of Wear and Aultman (2006), not least in their argument that professionalism may be reduced to that which is measurable, expressive of a fear that the phronetic, contextual, and reflexive competences of the professional may be marginalized in favor of a largely mechanic rule following. Secondly, research into the experiences and attitudes of medical students suggests that traditional values such as altruism no longer resonate with them. The superseding of personal interest disrupts a desirable work-life balance and may be seen to leave students vulnerable to exploitation by their teachers and physicians vulnerable to exploitation by patients (Hafferty 2002). This unease with selflessness may be even more strongly felt in traditionally subordinate medical subdisciplines and particularly in nursing. Finally, where traditional approaches see respect for codes of conduct as an integral part of professionalism, it may be asked whether an ethics, not dissimilar in its outline to that advocated by Durkheim, has the power to check the abuses and failings of professional autonomy (see Freidson 2001, p. 215). Traditional approaches to professionalism tend thereby to take for granted the assumption that professions can autonomously regulate their own practitioners. This assertion of a code of conduct may, nonetheless, be interpreted critically as a political move, consciously or unconsciously, to safeguard a self-interested autonomy, rather than a genuine response to the problem.

Hafferty's review of the more reformist "new professionalism" suggests two issues (2006, p. 198f). One rests upon the individual disciplines of the practitioner him- or herself; the other on the practitioner's place in wider society. The former develops the idea of the reflective practitioner (Schön 1983) and as such grounds the integration of professionalism into education, specifically in practice-based learning. For Epstein, critical self-reflection facilitates professionalism, in that it "enables physicians to listen attentively to patients' distress, recognize their own errors, refine their technical skills, make evidence-based decisions, and clarify their values so that they can act with compassion, technical competence, presence, and insight." Indeed, the lack of such reflection is blamed "for some deviations from professionalism and errors in judgment and technique" (1999, p. 833). In part, this approach may be seen to reassert the autonomy and indeed phronesis of the individual practitioner, thereby defending them against further regulation. Epstein's

approach also suggests that professional reflection is a largely tacit competence, to be learned not through the following of explicit rules but rather through the examples and guidance provided by mentors (ibid). As such, it may be seen to reproduce the distinction that has held since the eighteenth century between a craft, as the mere mechanical application of a rule of practice, and a profession, requiring a phronetic capacity to understand the relevance of theoretical knowledge contextually and particularistically.

Epstein's approach may renew the ethical commitment of the professional but continues to be potentially self-serving if it lacks rigorous enforcement. The rethinking of civic professionalism by the likes of Sullivan (1999, 2004), Mechanic (2000), and Frankford and Konrad (1998) begins to respond to this challenge. The altruism that is fundamental to traditional notions of professionalism is reinterpreted as civic engagement and thus as a professional commitment to civic equality and social justice. Echoing the calls of Gregory in the eighteenth century, the patient and public are brought into dialogue with the professional within a "body politic." Trust is restored in the profession, not simply through the institution of more rigorous or inventive codes of conduct but rather through the profession leading discussion with the public as to what the nature and role of the profession should be (Sullivan 2000). This leads Frankford and others to reject the traditional commitment to medical autonomy (Frankford et al. 2000). The rise of evidence-based medicine and the quality-of-care movement, not least insofar as the "patient experience," is placed centrally to any judgment of good treatment and confronts the profession with external standards for assessing and regulating their practice. More precisely, the responsibility for such assessment may not lie most effectively with the profession's own regulative bodies. NICE and the Care Quality Commission in England play such a role. New professionalism may then be understood as embracing external regulation and the imposition of clinical guidelines. Such acceptance need not, as some fear, reduce the practitioner to a mere technical expert. Rather, as Light defends "accountability-based" professionalism, there is a shift from a variable quality of care, due to its individual determination by the autonomous practitioner, to the use of "guidelines, protocols, and care pathways" to ensure outcomes grounded in clinical research. A new clinical research elite sets evidence-based standards, which would for Light emphasize primary care, prevention, and the management of illness over and above the curative interventions that have been the tradition sources of professional prestige. This would in turn require teamwork and cooperation between the medical subdisciplines (including managers and specialists in evidence-based medicine and epidemiology), disrupting the traditional hierarchy that allows physicians to control and delegate (Light 2010, p. 279).

Within the UK, Irvine's conception of "patient-centered care" offers a model of such new professionalism. Patient-centered care recognizes that the modern age is one of patient autonomy, not professional autonomy. Patients have better advocacy from charitable and support groups and are given a clear voice in legal actions

against inadequate care but are also increasingly expert in their own conditions – so throwing into question the model of the patient found in say Parsons and later Freidson. For Irvine, the patient is the final arbiter of what is right for them, and:

> [T]hey equate professionalism with consistently good doctoring. For them 'good doctors' are up to date, competent, respectful, courteous, kind, empathetic and honest; people who will listen to them, relate to them, do their best to find out promptly what is wrong with them, prescribe the right treatment and care for them in a manner which makes them feel that their interests come first. Patients want their doctors to be good team players when teamwork is needed. (Irvine 2014, p. 7)

Crucially, this model is not presented as mere personal motivation and ethic but as something that must be regulated and enforced. Irvine sees this in the development of the GMC in the 1990s and beyond, as its code "Good Medical Practice" was instituted, not merely as advice but as the framework within which education, regulation, and all importantly the ongoing revalidation of the physician's fitness to practice proceed.

In summary, new professionalism has responded to the crisis in public trust by rethinking codes of conduct and thus the image of what "good doctoring" is, integrating such codes rigorously into education and validation but more radically questioning the traditional value of autonomy. Professionalism thereby ceases to be a mere ethic and comes to embrace the acceptance of regulation and cooperation with other professional disciplines; professional altruism is transformed into the acceptance that the patient lies at the center of health-care provision and has a fundamental right to consistent, high-quality care.

Definition of Key Terms

Civic professionalism	Approach to professionalism that focuses on the professional's participation in civil society.
Deprofessionalization	Thesis that professions are losing their distinctive status, due to the loss of autonomy and increased routinization of the application of expertise.
Dominance theory	A sociological theory developed by Eliot Freidson and others, critically analyzing the power exercised by professions, as a dominant position working in the profession's own interests.
New professionalism	Response to the perceived crisis in professional, due to a loss of public trust and increase regulation. New professionalism challenges traditional professional characteristics and in particular professional autonomy.
Professionalism	Set of values traditionally associated with professional performance, focusing on an altruistic and other-regarding ethic, alongside commitment to maintaining professional expertise.

Summary Points

- Professionalism is traditionally characterized as entailing an other-regarding ethic and a commitment to the maintenance of high standards of scientifically based expertise.
- The "golden age" of professionalism, between approximately 1945 and 1965, sees high levels of public trust in the professions, allowing profession to be self-regulating and autonomous.
- The self-regulation of professions is challenged in the 1970s, as it is increasingly recognized that self-regulation serves the interests of professionals and fails to deliver high standards of care to patients and society as a whole.
- A crisis of professionalism occurs in the 1970s and 1980s, as professional self-regulation is seen to fail patients and the general public and as public trust in professions declines.
- Developments such as the emergence of new professions within medicine, evidence-based medicine, and a renewed emphasis on patient-centered care and patient rights undermine traditional defenses of professional autonomy.
- New professionalism responds to the crisis in professionalism, most fundamentally by rejecting the ideal of professional autonomy, in favor of evidence-based external regulation and the imposition of practice guidelines.

References

ABIM Foundation. American Board of Internal Medicine, ACP-ASIM Foundation. American College of Physicians-American Society of Internal Medicine, European Federation of Internal Medicine (2002) Medical professionalism in the new millennium: a physician charter. Ann Intern Med 136(3):243–246

American Board of Internal Medicine (1995) Project professionalism. American Board of Internal Medicine, Philadelphia

Arnold EL, Blank LL, Race KE, Cipparrone N (1998) Can professionalism be measured? The development of a scale for use in the medical environment. Acad Med 73:1119–1121

Baker R (1995) The codification of medical morality: historical and philosophical studies of the formalization of Western medical morality in the eighteenth and nineteenth centuries volume two: Anglo-American medical ethics and medical jurisprudence in the nineteenth century. Kluwer, Dordrecht

Berlant JL (1975) The profession and monopoly. University of Californian Press, Berkeley

Bledstein B (1976) The culture of professionalism. W. W. Norton, New York

Boyd KM (2005) Medical ethics: principles, persons, and perspectives: from controversy to conversation. J Med Ethics 31:481–486

Crowther MA (1995) Forensic medicine and medical ethics in nineteenth-century Britain. In: Baker R (ed) The codification of medical morality. Kluwer, Dordrecht, pp 173–190

Cruess SR, Johnston S, Cruess RL (2004) Profession: a working definition for medical educators. Teach Learn Med 16:74–76

Durkheim E (1984) Division of labour in society. Macmillan, London

Durkheim E (1992) Professional ethics and civic morals. Routledge, London

Edgar A (2011) Professional values, aesthetic values, and the ends of trade. Med Health Care Philos 14(2):195–201

Elston MA (1991) The politics of professional power. In: Gabe J, Calnan M, Bury M (eds) The politics of professional power. Routledge, London, pp 58–98
Epstein RM (1999) Mindful practice. JAMA 282:833–839
Epstein RM, Hundert EM (2002) Defining and assessing professional competence. JAMA 287:227–235
Flexner A (1910) Medical education in the United States and Canada: a report to the Carnegie foundation for the advancement of teaching, vol 4, Bulletin. The Carnegie Foundation for the Advancement of Teaching, New York
Frankford DM, Konrad TR (1998) Responsive medical professionalism: integrating education, practice, and community in a market-driven era. Acad Med 73:138–145
Frankford DM, Patterson M, Konrad TR (2000) Transforming practice organizations to foster lifelong learning and commitment to medical professionalism. Acad Med 75:708–717
Freidson E (1970) Profession of medicine: a study of the sociology of applied knowledge. Harper & Row, New York
Freidson E (1994) Professionalism reborn: theory, prophecy and policy. University of Chicago Press, Chicago
Freidson E (2001) Professionalism: the third logic. University of Chicago Press, Chicago
Gregory J (1817) Lectures on the duties and qualifications of a physician. Carey and Son, Philadelphia
Hafferty FW (2002) What medical students know about professionalism. Mt Sinai J Med 69:385–397
Hafferty FW (2006) Definitions of professionalism: a search for meaning and identity. Clin Orthop Relat Res 449:183–204
Hafferty FW, Havighurst CC, Relman AS, Freidson E (2003) Review symposium on Eliot Freidson's professionalism: the third logic. J Health Polit Policy Law 28(1):133–172
Irvine D (2014) Patients, their doctors and the politics of medical professionalism: address to members of the American Osler Society. Available from http://www.pickereurope.org/individual-clinician-feedback-is-vital/. Accessed 14 Aug 2014
Larson MS (1977) The rise of professionalism: a sociological analysis. University of Californian Press, London
Leake CD (1927) Percival's medical ethics. Williams and Wilkins, Baltimore
Leake CD (1970) Percival's medical ethics: promise and problems. Ann N Y Acad Sci 169:388–396
Light DW (2010) Health-care professions, markets, and countervailing powers. In: Bird CE, Peter C, Fremont AM, Stefan T (eds) Handbook of medical sociology, 6th edn. Vanderbilt University Press, Nashville
Lindemann M (2008) Medicine, medical practice, and public health. In: Wilson PH (ed) A companion to eighteenth-century Europe. Blackwell Reference Online, Oxford. http://www.blackwellreference.com/subscriber/tocnode?id=g9781405139472_chunk_g978140513947212. Accessed 1 Aug 2014
Mechanic D (2000) Managed care and the imperative for a new professional ethic: a plan to address the growing misfit between traditional medical professionalism and emerging health care structures. Health Aff 19:100–111
Parsons T (1939) The professions and social structure. Soc Forces 17:457–467
Parsons T (1951) The social system. Routledge, London
Pattison S, Pill R (eds) (2004) Values in professional practice: lessons for health, social care and other professionals. Radcliffe, Oxford
Percival J (1849) Medical ethics, 2nd edn. Shrimpton, Oxford
Porter R (1997) The greatest benefit of mankind: a medical history of humanity from antiquity to the present. Harper Collins, London
RCGP (1977) Evidence to the Royal Commission on the NHS, Journal of the RCGP 27:197–206
Reed RR, Evans D (1987) The Deprofessionalization of Medicine: Causes, Effects and Responsibiltieis. JAMA 258:3279–3282

Schön D (1983) The reflective practitioner, how professionals think in act. Basic Books, New York
Secretary of State for Social Services (1975) Report of the Committee of Inquiry into the regulation of the medical profession. Chairman Dr A.W. Merrison. Cmnd 6018. Stationery Office, London
Shaw GB (1909) Preface to 'The Doctor's Dilemma. Available http://www.gutenberg.org/files/5069/5069-h/5069-h.htm. Accessed 14 Aug 2014
Starr P (1982) The social transformation of American medicine. Basic Books, New York
Sullivan WM (1999) What is left of professionalism after managed care? Hastings Cent Rep 29:7–13
Sullivan WM (2000) Medicine under threat: professionalism and professional identity. Can Med Assoc J 162:673–675
Sullivan WM (2004) Can professionalism still be a viable ethic? Good Soc 31:15–20
Swick HM (2000) Toward a normative definition of medical professionalism. Acad Med 75:612–616
Swing SR (2007) The ACGME outcome project: retrospective and prospective. Med Teach 29:648–654
Talbot JA (2011) Medical professionalism in the new century: accomplishments and challenges in the future for an American medical school. In: Bhugra D, Malik A (eds) Professionalism in mental healthcare: experts, expertise and expectations. Cambridge University Press, Cambridge, pp 126–138
Tawney RH (1921) The acquisitive society. Bell and Sons, London
Wear D, Aultman JM (2006) Introduction. In: Delese W, Aultman JM (eds) Professionalism in medicine: critical perspectives. Springer, New York, pp. vii–xi

Skilled Know-How, Virtuosity, and Expertise in Clinical Practice

42

Hillel D. Braude

Contents

Abstract

The terms "skilled know-how," "virtuosity," and "expertise" all denote forms of technical mastery. Applied to medicine they refer to different aspects of medical expertise. Yet, what exactly is medical expertise? Is it a kind of cognition, or action, or a combination of both? Additionally, what aspects of clinical practice does technical expertise refer to? In this chapter technical expertise in clinical practice is analyzed in terms of three identified components: cognition, motoric action, and interpersonal relations. Furthermore, the three components of technical mastery are related to Aristotle's concept of practical wisdom, or *phronesis*. In his *Nicomachean Ethics*, Aristotle differentiated between two different forms of human action: *techné* and *phronesis*. In broad terms *techné* refers to an action that results in the production of external objects, while *phronesis* refers to an action that has its end in itself. This distinction provides the architectonic keystone of this analysis of expertise in clinical practice. This approach presents an alternative to the predominant cognitive conception

H.D. Braude (✉)
The Mifne Center, Rosh Pinna, Israel
e-mail: hillel.braude@mcgill.ca

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_68

of technical expertise in clinical practice. A full understanding of technical expertise, skilled know-how, and virtuosity is not possible without highlighting the important role of intentionality in action and other forms of pre-reflective knowing in clinical practice.

Introduction

The terms "skilled know-how," "virtuosity," and "expertise" all denote forms of technical mastery. Applied to medicine they refer to different aspects of medical expertise. Yet, what exactly is medical expertise? (Despite the subtle differences of meaning, the terms skilled know-how, virtuosity, and expertise will be referred to more or less synonymously under the rubric of medical expertise.) Is it a kind of cognition, or action, or a combination of both? Additionally, what aspects of clinical practice does technical expertise refer to? Geoffrey Norman et al. note that, "Expertise in medicine requires mastery of a diversity of knowledge and skills – motor, cognitive, and interpersonal . . ." (2006, 339). These three elements pertain to each stage of clinical practice, i.e., patient diagnosis, evaluation of possible therapies, and deciding the best course of action in the particular circumstance (Pellegrino and Thomasma 1981). The development of this chapter will provide a discursive analysis of technical expertise in clinical practice in terms of these three components of motoric action, cognition, and interpersonal relations. Furthermore, these three components of technical mastery will be related to Aristotle's concept of practical wisdom, or *phronesis*. In his *Nicomachean Ethics*, Aristotle (1925) differentiated between two different forms of human action: *techné* and *phronesis*. In broad terms *techné* refers to an action that results in the production of external objects, while *phronesis* refers to an action that has its end in itself. This distinction provides the architectonic keystone for this analysis of expertise in clinical practice. Additionally, where it is helpful for elucidation, this analytic review of expertise in clinical practice will draw on evidence from the neurosciences. This approach presents a critique of the predominant conception of technical expertise primarily in terms of cognition, which arguably has obstructed the development of comprehensive literature around skilled know-how in clinical practice beyond a primary level. A full understanding of technical expertise, skilled know-how, and virtuosity is not possible without highlighting the important role of intentionality in action and other forms of pre-reflective knowing in clinical practice.

Clinical Practice as *Phronesis* or *Techné*

A number of authors have argued that clinical reasoning is best understood as a form of Aristotelian *phronesis* (see, e.g., Pellegrino and Thomasma 1981; Gatens Robinson 1986; Widdershoven-Heerding 1987; Beresford 1996; McGee 1996; Montgomery 2000; Braude 2012b). In the *Nicomachean Ethics*, Aristotle (1925)

lists *phronesis* as one of the intellectual virtues alongside philosophic wisdom, or *sophia*, and understanding, or *nous* (1103a6). Aristotle notes that practical reasoning requires means of verification that are appropriate for the subject matter at hand: "For it is the mark of an educated man to look for precision in each class of things just so far as the nature of the subject admits; it is evidently equally foolish to accept probable reasoning from a mathematician and to demand from a rhetorician scientific proofs" (1094b). In this statement, Aristotle presents the radical idea that certain kinds of practical knowing are justified primarily through calculating the ends of their actions in the real world, not in terms of mathematical or other kinds of theoretical abstraction.

Clinical practice is an exemplary form of *phronesis* because of its perennial concern for what is temporally in flux. Aristotle (1925) notes that a key aspect of *phronesis* is its association with what is variable:

> No one deliberates about things that are invariable, nor about things that it is impossible for him to do. Therefore, since scientific knowledge involves demonstration, but there is no demonstration of things whose first principles are variable (for all such things might actually be otherwise), and since it is impossible to deliberate about things that are of necessity, practical wisdom cannot be scientific knowledge nor art; not science because that which can be done is capable of being otherwise, not art because action and making are different kinds of thing. The remaining alternative, then, is that it is a true and reasoned state of capacity to act with regard to the things that are good or bad for man. (1140a)

Arguably clinical medicine is the exemplary science of the contingent. All clinicians, not only expert ones, necessarily are faced with individual variability in their everyday practice. Similar diseases present differently due to biological variability and the influence of patient subjectivity on the experience of illness. Human biology presents a multileveled complexity than can be accounted for purely by the physical sciences (Schaffner 1994). The epistemological attempt to define clinical reasoning needs to take account of this inherent variability in the human condition, best accounted for through the Aristotelian virtue of *phronesis*.

Is it not a category error, however, to describe clinical practice as a form of practical wisdom, and not rather a kind of technical expertise? In the Hippocratic tradition, medicine was considered to be a form of craft (*techné*) (Edelstein 1967). Techniques of all sorts characterize modern clinical practice, from auscultation to conducting an autopsy on a cadaver. Is not this sense of technical mastery associated with craftsmanship more suitable to describe clinical practice, especially in relation to skilled know-how, virtuosity, and expertise? Aristotle (1925) distinguished *phronesis* from *techné* in one essential sense: the end of *phronesis* is the action itself (1140b5–7). On the other hand, *techné* is defined by being an action that has an end other than itself (1140b6–7). Furthermore, *phronesis* and *techné* can also be understood in terms of their relation to two further categories of human activity, *poiesis* and *praxis* (Heidegger 1997. See also the discussion on *phronesis* informed by this distinction by Hans Georg Gadamer 1975; and Robert Bernasconi 1989). Thus, *phronesis* can be distinguished further as a form of *praxis* and *techné* as a form of *poiesis*. *Praxis* is a form of practical activity that is intended to further

human well-being or the good and is not associated with any particular end product external to the act. *Poiesis* refers to any human activity that results in a product external to the human activity itself. The act of making associated with crafts (*techné*) is associated, therefore, with *poiesis*.

Aristotle (1925) conceived of the end of medicine being health (1094a). If curing disease or health is considered to be separate from medical action, then indeed, medicine should be considered a form of *techné*. This is the position taken by Barbara Hofmann (2002), in her insightful critique of medicine as practical wisdom. If, on the other hand, clinical practice is concerned with intermediary steps in order to achieve personal well-being, then it is akin to *phronesis*. As argued, however, medicine through clinical reasoning is unusual in sharing aspects of *techné* and *phronesis* (Braude 2012a). Or rather, clinical practice modeled on *phronesis* explicitly incorporates aspects of *techné* without contradiction. As has been observed, "*Phronesis* demonstrates this dual quality by being concerned with both technical issues and intermediary steps towards the end of action" (Braude 2016). Yet, what differentiates a specific act in the world as being one of *phronesis* or *techné* depends on the constitutive variables concerning the agent, the intention, and the outcome. This point will become more apparent through the analysis of medical expertise in this chapter.

Aristotle uses the metaphor of medicine as exemplary for the dual nature of *phronesis*. Thus, in the *Nicomachean Ethics*, Aristotle argues that medicine is a form of practical wisdom whose end is the health of the individual. "Medicine," Aristotle writes, "does not govern health, but is for the sake of health" (1925, 1145a11). Medicine is an exemplary form of practical wisdom, since in determining what is best for a particular individual, technical know-how is necessary but not sufficient. "Medicine," as Kathryn Montgomery writes, "is neither a science nor a technical skill (although it puts both to use) but the ability to work out how general rules – scientific principles, clinical guidelines – apply to one particular patient" (2006, 5). Medicine as a form of practical wisdom lies close to the determination of *poiesis* in requiring technical expertise to achieve the ends of clinical action. At the same time, this knowledge can never be divorced from the self-knowing associated with *phronesis*. Clinical practice shares with *phronesis* an inherent structure of moral agency (Gallagher 2007). As philosopher Stephen Toulmin elegantly states, "Once brought to the bedside, so to say, applied ethics and clinical medicine use just the same Aristotelian kinds of "practical reasoning," and a correct choice of therapeutic procedure in medicine is the right treatment to pursue, not just as a matter of medical technique but for ethical reasons also" (Toulmin 1982).

Cognitive Expertise

Cognition is the first component of medical expertise mentioned by Norman et al. (2006). Cognition refers to all mental processes related to knowledge, including but not limited to memory, attention, perception, representational schemas, consciousness, and language. Norman et al. also note that, "much of what

we call medical expertise is really closer to medical diagnostic expertise . . ." (2006, 340). In other words, clinical expertise is often conflated with processes associated with clinical reasoning, which in turn is most often assessed in terms of cognition. It is worth pondering the reasons for this cognitive emphasis in the literature around medical expertise, and how this cognitive bias influences the understanding of clinical expertise. For the moment, however, this analysis will focus on clinical reasoning itself as a form of "skilled know-how" conceptualized in terms of cognition. (For a fuller analysis of clinical reasoning and cognitive knowing, see Braude, 2016).

The influential hypothetico-deductive model conceives of clinical reasoning as a form of cognition applied to evaluating and managing a patient's medical problem (Barrows and Tamblyn 1980). According to this model, a clinician creates working hypotheses inferred from the patient's presenting symptoms. The clinician then refines these hypotheses through a gradual process of elimination, until the most compelling hypothesis is chosen, which best fits with the primary clinical diagnosis (Barrows and Feltovich 1987). In its cognitive approach to clinical reasoning, the hypothetico-deductive model only accounts for explicit analytic inferences. Its intellectual and methodological basis is derived from the psychological heuristics and biases approach developed a number of decades previously by Daniel Kahneman and Amos Tversky (1974). (For a review of the application of the discipline of cognitive psychology to evaluate clinical decision-making, see Elstein 2000, 2009).

Physician Pat Croskerry, among others, has championed applying the tools of cognitive science in order to examine our inherent cognitive biases involved in clinical reasoning. Croskerry cites more than 40 different kinds of cognitive and affective biases that together may contribute to impaired clinical judgment (Croskerry 2008; Croskerry et al. 2008). The kinds of cognitive impediments affecting effective clinical decision include those relating to the structure of cognition, psychological ego defenses of the decision-maker, and the neurophysiological state of the decision-maker at the moment of decision. Despite their different nuances and variations, these are all cognitive; however, their influence and impact are to a great extent unknown to the decision-maker at the time of the decision. While skeptical toward the inherent biases, Croskerry's cognitive program is ultimately amelioristic in providing a means of improving cognitive decision-making. Thus, Croskerry conceives that these largely unconscious cognitive biases can be made explicit through processes of self-reflection, introspection, and metacognition.

Metacognition refers to the processes involved in thinking about one's own thinking. It may also refer to the activity of monitoring and controlling one's own cognitive activity (Proust 2013). Applied to clinical reasoning, metacognition provides a cognitive mechanism enabling a clinician to validate or reject a diagnostic or therapeutic decision (Marcum 2012). Faced with a set of clinical symptoms that do not fit a recognized picture, or a patient reacting adversely to a specific treatment, the clinician is forced to reflect more explicitly on the clinical presentation and her diagnostic reasoning. The cognitive processes whereby diagnostic reasoning may be improved through metacognition facilitate the development of clinical expertise. The master clinician is differentiated from the average clinician

by his/her conscious ability to reflect on inherent cognitive biases and change them during the normal course of medical practice.

While the cognitivist approach to clinical reasoning attempts to render implicit biases into explicit knowing, it only mediates processes of clinical reasoning that may be rendered into explicit inferences. The cognitivist approach does not attempt to account for tacit knowing as an essential element of diagnostic reasoning. The model of tacit knowing posited by philosopher Michael Polanyi (1962, 1966) includes nonanalytic or non-inferential forms of knowing. Tacit knowing refers to knowledge that functions at the periphery of attention and makes explicit knowledge possible. Stephen Henry (2010) argues that there are two features of tacit knowing that are especially relevant to clinical medicine. Firstly, explicit knowledge could not exist without the prior existence of a "tacit background." For example, Henry observes that tacit knowing occurs when a physician who is explicitly listening to a patient's story is simultaneously aware, but in a qualitatively different way, of the patient's tone of voice, facial expression, and choice of words. Moreover, these tacit particulars are crucial to informing the physicians' processes of clinical judgment. Secondly, the mechanism of how "tacit particulars give rise to explicit knowledge cannot be fully captured in formal models or discrete steps; the relationship is ultimately inarticulable" (Henry 2010, 293). This theory of tacit knowing fits in well with theories of clinical reasoning based on gestalt perception, which provides evidence that perception of a given object exhibits intrinsic qualities that cannot be completely reduced to its constitutive sensible components (Cervellin et al. 2014). Theories of tacit knowing help explain the clinical reliance on pattern recognition, a critical component of diagnostic expertise. In summary, models of clinical reasoning based in terms of tacit knowing presents an alternative model of clinical rationality. A complete model of clinical reasoning needs to include both cognitive and tacit forms of knowing (Marcum 2012).

Tacit knowing is non-inferential. Yet, since it improves with experience, it too can be considered a form of technical skill or mastery. This intuitive skill is privileged in (the mathematician) Stuart and (philosopher) Hubert Dreyfus' influential schema of adult skill acquisition, which has also been applied as a model for the development of medical expertise (Dreyfus and Dreyfus 1988; Dreyfus 2004). They conceive five levels of developing expertise from that of novice to advanced beginner, to competence, to proficiency, to expertise. The development from novice to real expert is characterized by the movement from rule-based decision-making to situational discriminations, whereby the expert demonstrates flexibility, wisdom, and improvisational ability. For Dreyfus and Dreyfus, learning how to play the game of chess provides the exemplary form of human skill acquisition and cognitive mastery. For example, in reference to stage five, that of proficiency, they write:

> The proficient chess player, who is classed a master, can recognize almost immediately a large repertoire of types of positions. He or she then deliberates to determine the move that will best achieve his or her goal. One may know, for example, that he or she should attack, but he or she must calculate how best to do so. (Dreyfus 2004, 179)

Increased proficiency includes the ability for gestalt pattern recognition, as well as the technical ability to devise instrumental means to achieve one's aims. Stage five – that of expertise – is similar to the stage of proficiency, but even that much more refined. The real expert possesses a "vast repertoire of situational discriminations" as well as the ability to see "immediately how to achieve this goal." The expert possesses the ability to distinguish between apparently similar situations requiring different responses.

This model is similar to others whereby developing expertise passes through a cognitive rule based, to an autonomous phase (Fitts and Posner 1967). In the autonomous phase, knowledge is characterized by being implicit and non-verbalizable (Masters 1992). What most defines this conception of expertise is the ability to rely on immediate intuition. As Stuart Dreyfus describes, "The brain of the expert gradually decomposes … situations into subclasses, each of which requires a specific response. This allows the immediate intuitive situational response that is characteristic of expertise" (Dreyfus 2004, 180).

The sense of expertise as immediately intuitive is synonymous with Hubert Dreyfus' understanding of Aristotelian *phronesis*, informed by the interpretation provided by the German philosopher Martin Heidegger. This understanding of *phronesis* has two main aspects. Firstly, *phronesis* is akin to a form of "pure perceiving." Secondly, arising from this *phronesis* is completely nonconceptual. In Heidegger's terms, *phronesis* "no longer falls within the domain of the logos" (1997, 112). This intuitive certainty is derived from the combination of great training and skill, together with an embodied immersion in the life-world in which expertise functions.

Dreyfus and Dreyfus's conception of expertise has been proliferative, generating further analysis and secondary discussion, particularly in its application to medicine (see, e.g., Benner 1984; Thornton 2010). The correctness of their schema is not examined here. Rather, for the present purposes it is important to consider its application in relation to the two types of diagnostic expertise so far discussed. For the most part, the cognitive model of clinical reasoning has been privileged as more rational than that of tacit reasoning, because it fits in with the cognitive science model of rationality (Stanovich 2011). Yet, clinicians themselves continue to argue for the validity of gestalt recognition and intuition in their day-to-day clinical practice (see, e.g., Woolley and Kostopolou 2013). Interestingly Hubert Dreyfus (2006) suggests that the cognitive model occurs prior to and becomes superseded by intuitive expertise that is essentially nonconceptual. However, as others have noted, when a breakdown in diagnostic reflection occurs, the clinician necessarily reverts back to more explicit cognitive reflection (Marcum 2012). Thus, there is a dialectic relation between cognitive and intuitive processes of diagnostic reasoning.

Whether *phronesis* is absolutely nonconceptual is a matter of philosophical speculation that cannot be settled here. As previously argued, *phronesis* does have an intuitive component, more akin to pre-reflective consciousness, than explicit conception (Braude 2012b, 2013). Additionally, *phronesis* has an integrative function that can unite the two main forms of clinical reasoning. *Phronesis* is a

particularly apt and useful model for clinical reasoning because it allows for the possibility of linking and integrating different cognitive processes according to the context and need of the moment. Different forms of cognition, such as affect, emotion, executive attention, rational cognition, and intuition may all constitute components of practical wisdom. Additionally, *phronesis* links the moral, ontological, and epistemological components of clinical medicine into a single framework. As has been observed, "Modelling clinical reasoning on *phronesis* succeeds in providing a means of integrating the different cognitive components of clinical reasoning, while maintaining respect for the gestalt of clinical reasoning as a particular form of conscious experience" (Braude 2012a, 947–948). Analyzing medical expertise needs to take account of these different levels of clinical reasoning. However, there still remains to be addressed the technical component of medical expertise, albeit in relation to the conception of clinical practice as a form of *phronesis*. This will be addressed in the second element of medical expertise, i.e., motoric activity and its relation to intentionality.

Motoric Expertise

In discussing technical expertise in clinical practice, it is obvious that medicine encompasses so many different kinds of skills and situations that it is impossible to define a single kind of "virtuosity." The skills necessary for a psychiatrist are obviously very different from that of family physician, and in turn very different from a cardiac surgeon. However, clinical practice as a form of *phronesis* unites very different kinds of clinical action. The nature of a clinical action then may be evaluated in terms of the relation between technical action and practical wisdom of its constitutive parts. While it is true that developing expert cognition, for example, in terms of diagnostic reasoning, is obviously a critical skill for clinical practice, technical skill refers arguably pre-eminently to motoric action. The virtuosity of a skilled hand surgeon is more explicitly embodied than the psychiatrist skilled in taking an expert history. However, as mentioned, the literature on technical mastery emphasizes processes of cognition at the expense of motoric action. While diagnostic reasoning is more cognitive than motoric, it is not more mental. Thus, motoric action can similarly be mapped in terms of cortical and subcortical function and the neural networks between them, as well as cortical structures directly involved in action planning (Jeannerod and Frak 1999). The question that then arises is why has motoric action not been adequately taken account of in the literature on technical expertise? Secondly, is motoric action fundamentally different from mental processes of diagnostic processing in terms of inherent conceptuality? In short, is motoric action closer to Aristotelian *phronesis* or *techné*? In order to avoid the emphasis on conceptual processes, Tim Thornton (2010) discusses anesthesia as exemplary for clinical expertise, with its emphasis on both manual and mental dexterity. All forms of surgical practice requiring technical skill fit this picture of technical expertise. Yet, does manual dexterity constitute an essentially different kind of clinical expertise to diagnostic reasoning?

This discussion on motoric activity will focus initially on one component of manual dexterity that is, perhaps, metonymic for clinical dexterity more generally, i.e., visuospatial ability. In his bestselling volume "How Doctors Think," Jerome Groopman (2007) quotes the expert cardiologist James Lock describing the visuospatial skills necessary to insert a cardiac catheter through a child's blood vessels and then into his heart:

> The catheter appears as a thin white line on a flat monitor screen next to the table. It can be difficult in such a two-dimensional projection to know the catheter's position. "The combination of how your hand moves and what the image looks like will tell you whether the catheter is pointed toward you or away. I can tell where it is even if my hand is off the catheter. Knowing in which direction you are going shouldn't be something you need to think about." . . . "You need to process what you see very quickly and act on the information in a split second," Lock said, "because the heart is beating. It's not like you can stop the child's heart and ponder. Once you are inside of a kid's heart with a catheter, you have an enormous amount you have to accomplish, and there is a great deal of risk if what you do is not done quickly and well." (Groopman 2007, 141)

Groopman observes that this visuospatial ability may be more paramount in determining surgical skill than nimbleness of hands, or straightforward manual dexterity. Additionally, like other kinds of technical skill, it is acquired through a combination of formal learning and repeated practice (Norman et al. 2006). This aspect of motoric skill makes it analogous to other kinds of cognitive expertise. Additionally, what is perhaps most emblematic of motoric skill, i.e., its automaticity makes it exemplary for Dreyfus' model of expertise. As William James first noted, (1890), well-practiced tasks can be performed with little effort or cognitive control, as opposed to novice performances of the same task. In terms of visuospatial ability, researchers have demonstrated that subjects "can adjust their movements in response to a change in the location of a visual target of which they are perceptually unaware" (Haggard and Johnson 2003, 76). This kind of automatic motoric skill, bypassing higher cortical control, fits in well with the Dreyfus' five stage schema of technical expertise. What they refer to as intuition is synonymous with automaticity of skilled motoric action. Indeed, the notion that intuition is a form of practical wisdom that is inherently nonconceptual seems to fit in best with the model of motoric action.

Is the motoric component of a clinical action best assessed in terms of *phronesis* or *techné*? Motoric action at first glance appears closer to technical expertise considered in terms of crafthood, or *techné*, than other forms of clinical practice, since it produces an externally visible result. However, differentiating between *phronesis* and *techné* in motoric action in clinical practice requires a closer analysis in terms of phenomenology of action. In this regard, an action can be defined as a "movement of the body, resulting from specific mental preparation, and aimed at some goal that the agent desires to achieve" (Haggard and Johnson 2003, 73). Action in the clinical context is no different. A clinician will initiate a specific motoric action resulting from a process of cognitive reflection in order to achieve a therapeutic effect. This might include, for example, taking a blood sample for a

pathological investigation, writing a prescription on a pad of paper, or inserting a scalpel in the skin to drain an abscess. Technical expertise or skill can focus on each of these three elements, however, the middle component – movement of the body – is strictly speaking the only truly motoric component of action.

Haggard and Johnson's (2003) analysis of the phenomenology of action highlights the following paradox: a phenomenology of action demonstrates that we have minimal conscious experience of many of our motoric activities, especially those which are automatic, such as breathing and walking. On the other hand, we are able to report in considerable detail the processes of preparation and execution of our actions. Thus, we are able to provide a richer phenomenological description of the mental processes reflecting on an action prior to its occurrence, and the results of a specific action, rather than the action itself, which consists of unconscious micro-components. This fact helps explain the relative lack of literature on motoric expertise in clinical practice, as opposed to processes of diagnostic reasoning. It also helps to differentiate motoric action in terms of the relation between *phronesis* and *techné*.

In her incisive book on *Intention*, philosopher Elizabeth Anscombe (1963) puts forward the thesis that an agent's knowledge of what he or she is doing is not characteristically based on observation. For Anscombe, practical action is characterized by a kind of intentionality that is not observable in terms either of the physical preparation of action nor of its measurable outcomes. This does not mean that these two aspects, prior to and post action, are not associated with practical action. Rather, they are not what is essential about practical action. As Anscombe states, "That what one knows as intentional is only the intention, or possibly also the bodily movement; and that the rest is known by observation to be the result, which was also willed in the intention." (1963, 51–52). In other words, it is the embodied action itself that is associated with practical knowledge. It is this, as Haggard and Johnson emphasize, that most resists direct observation.

Anscombe's conception that practical knowledge is non-observational is motivated by the idea that practical knowledge is "the cause of what it understands," rather than being derived from "objects known" (1963, 87–8). Without going into a detailed comparison, Anscombe's insight restates the central understanding that the end of *phronesis* is the action itself, in contradistinction to *techné*, which produces a result external to the action. For Anscombe, what characterizes a practical action is that it is motivated by an intention that is non-observable. An action might outwardly be the same, but is differentiated by its motivating intentionality. This insight differentiates between an action that might be associated with *techné*, such as a sculptor hammering a statue, and an act of practical knowledge, such as an orthopedic surgeon doing a similar action while inserting a hip replacement. Another key difference between these two kinds of action is that practical knowledge is self-reflexive, whereas *techné* is not necessarily so. In other words, an act resulting from practical knowledge will necessarily impact self-referentially on the doer, as well as result in an external product of the action. This introduces an ethical dimension into practical knowledge – although this was not explicitly addressed by Anscombe in her work on *Intention* – that is especially pertinent for clinical action

involving a vulnerable other – the individual patient. Thus, when a physician performs a clinical act, it is self-reflexive, even when it is primarily other-directed. A clinical act should always include self-knowledge on the part of the clinician. For this reason, Eric Cassell cites approvingly the two essential habits of mind essential for clinicians posited by the great Canadian physician Sir William Osler (1848–1919), i.e., imperturbability and equanimity. "Imperturbability means coolness and presence of mind under all circumstances, calmness amid storm, clearness of judgments in moments of grave peril and impassiveness." (Osler 1905, 3ff). Equanimity is, "an evenness of mind or temper. The ability not to be disturbed or upset by the foolishness around you, the temper or fits of emotion or agitation of others (your patients above all)" (Cassell 2015, 231).

As stressed in this chapter and elsewhere (Braude 2012a, 2013), this kind of self-awareness is not simply the ability for metacognition, but the ability to become self-aware of the presence and importance of pre-reflective states of consciousness. This includes the kind of non-observable kinesthetic awareness that Anscombe associates with *phronesis*. Touching themes very close to the ones developed in this chapter, philosopher Shaun Gallagher observes in an important essay on "Moral Agency, Self-Wisdom and Practical Agency" that the phenomenological conception of intentional action is always accompanied by a pre-reflective self-consciousness. Gallagher continues that, "the person with *phronesis* knows what they are doing on an implicit level which is best expressed not by reflective or theoretically abstract propositions, but by descriptions on the highest pragmatic level of discourse . . ." (2007, 217).

In summary, clinical action is a kind of *techné* in being associated with an external outcome, e.g., draining an abscess. At the same time, it is always a form of *phronesis* impacting on the doer. This insight central to medical practice is exemplified in the Doctrine of Double Effect, which gives moral legitimacy to an equivocal action provided that the good and not harmful effect is intended, even though the unintended harmful effect may be foreseen (Mangan 1949). In medicine, the application of the Doctrine of Double Effect allows physicians to administer adequate palliative care to patients, even if it may lead to their death. It also provides support for the moral argument against physician-assisted suicide and euthanasia (Sulmasy and Pellegrino 1999). Intentionality is important because of the effect of an act on the self of the physician and may not be assessed purely in terms of the actual outcome.

Interpersonal Expertise

The third and final component of medical expertise referred to by Norman et al. (2006) is that of interpersonal relations. This component relates both to the quality of a particular medical action and its purpose or end. As such, the intersubjective context of a particular action might not explicitly inform the processes of diagnostic reflection or the therapeutic action directed toward an individual patient,

but nevertheless provides an overarching structure that influences every component of clinical practice. A clinician may be particularly skilled in personally relating to patients and in conveying information. At the same time, this medical action is laden with moral responsibility. Clinical practice reduced to its most simple equation as the intersubjective relationship between an individual physician and patient for the latter's physical and mental well-being necessarily introduces an ethical dimension into the discussion about medical expertise. This ethical dimension undergirds the question whether technical expertise in medicine can ever be "value-free" or is always embedded with moral values arising from the clinical encounter. (The relation between technical mastery and ethics is exemplified in the word "virtuosity." With roots in Post-Classical Latin and Middle French and referring to an exceptional performative ability, especially in the realm of musical instrumentation, virtuosity highlights the intimate relation between exceptional technical skill and moral virtue.) That a clinical action always needs to be related to the well-being of the individual patient firmly grounds clinical reasoning as a form of *phronesis*. Like the other two components of clinical expertise, cognition and motoric activity discussed in the earlier section of this chapter, focusing on interpersonal relations, forces one to brush against the issue of *phronesis* and its relation to *techné*.

Knowledge of the physician is intersubjective. In other words, the knowledge is never purely objective but is possessed by one subjectivity about another. This requires a shift in the traditional medical focus from objective disease to subjective categories, such as personal goals and function. Thus, Eric Cassell, observes that:

> Clinicians and clinical medicine require an alternative definition of sickness that does not diminish the importance of pathophysiology and the effects of disease but encompasses the impact of sickness on the patient's life and the impress of the patient on the sickness . . . . The goal of the clinician and clinical medicine is to restore the sick person to function so that goals and purposes can be achieved and well-being restored. (2015, 22)

Determining what constitutes the correct clinical goals and purposes is an inherently intersubjective process that requires the empathic ability of the clinician to feel and understand something of one patient's experience of pain, vulnerability, and suffering and express appropriate concern (Braude 2016). Here too *phronesis* is necessary to provide a clinician with the means to achieve the appropriate balance between clinical distance and empathic concern. Phenomenologists consider empathy to be a *sui generis* form of intentionality directed at other experiencing subjects (Zahavi and Overgaard 2012). Moreover, according to the phenomenological account provided by Edith Stein (1989), empathy is direct, unmediated, and non-inferential. Another key phenomenological aspect of empathy is the fact that empathy is always both self- and other-relating. In other words, intersubjectivity in clinical practice is always as much about the subjectivity of the clinician, as it is about the patient. A good clinician needs constantly to be assessing his/her motives/biases in relation to that of his patient. Yet, the direction of the intersubjective

reflection should always ultimately be directed toward the well-being of the patient, even during a moment of personal self-reflection.

Medical empathy is a cognitive and affective phenomenon that mediates the internal aspects of clinical reasoning processes, with outer worldly directed motoric action. As with other aspects of clinical practice that incorporate tacit, intuitive dimensions, it is questionable how empathic expertise can be formally taught and developed. Nonetheless, it is certain that technological advances in brain imaging techniques, together with second- and third-person observational methods will increasingly inform the social neuroscientific understanding into the nature of medical empathy (Schilbach et al. 2013). For example, studies on the empathy of pain primarily demonstrate significant action in regions involved in the affective aspects of the pain-processing network. These include the anterior cingulate cortex, the anterior insula, the cerebellum, and the brainstem (Singer et al. 2004). A number of functional MRI and MEG investigations of participants observing facial expression and stimuli depicting trauma to body parts have reported significant signal change in both the affective dimension of pain as well as the somatosensory cortex and posterior insula involved in the sensory discrimination of pain (Decety and Svetlova 2012). Jean Decety argues that physicians are able to downregulate their own pain response to observing the pain of others through managing and controlling their own higher cortical response (Decety Forthcoming). Management of their "negative arousal" enables physicians to liberate their cognitive resources necessary for effective therapeutic action and empathic concern. Through "feeling less," or having less affectivity, physicians are able to provide more effective care. This brief description of medical empathy presents just a hint of the relevance of social neuroscience for the development of interpersonal expertise in clinical practice. To translate into clinically relevant information, these second- and third-person-based neuroscience studies will need to be combined with first-person experience. Translated into formal techniques, processes of introspection, i.e., the perception of internal physical or physiological states (Wiens 2005), can help bring to conscious awareness the cognitive and affective bases of empathy and other physiological manifestations of intersubjectivity. It is predictable that affective introspection will become as valuable a tool for self-reflection on clinical reasoning as the use of metacognition to evaluate the cognitive foundations of clinical reasoning. Introspection is also an important component of narrative competence, defined as "the set of skills required to absorb, interpret, and be moved by the stories one hears or reads" in order to achieve "the genuine intersubjective contact required for an effective therapeutic alliance" (Charon 2004, 862–863).

In summary, as with the other two components of clinical practice, clinical empathy can be assessed in terms of both *phronesis* and *techné*. Considering clinical empathy as a teachable ability that can be objectified through neuroscientific techniques situates it on the side of *techné*. In this way, clinicians will be able to ratchet up or down their empathic concerns, based on the clinical needs of the moment. Considering empathy as an embodied phenomenon existing between two

corporeal beings, that is, both self-other related, as well as being direct, unmediated, and non-inferential, empathy is situated firmly on the side of *phronesis*. Thus, empathy is akin with the intersubjectivity that Gallagher claims "is endogenous to the embodied practices that constitute practical knowledge" (2007, 206).

Conclusion

This chapter has provided an analysis of skilled know-how, virtuosity, and expertise in clinical practice. Specifically, the three identified components of clinical expertise, i.e., cognitive, motoric, and interpersonal, have been related to Aristotelian *phronesis*, in particular the tension that arises in each of these three elements between *phronesis* and *techné*. As emphasized, whether technical expertise is closer to *phronesis* or *techné* is determined through evaluating the ends of a specific action. If a clinical action has an end in itself that is self-referential, involving the moral virtue of the person doing the action, then the clinical action is most likely a form of practical wisdom. If the action results in a product external to the action then it is associated on the spectrum of *techné*. Expert clinical reasoning is associated with *phronesis* in not being able to be simply reduced to a form of cognition. Similarly, a motoric action can be considered *phronetic* in relation to the intentionality possessed by the actor. Finally, empathy exemplifies the interpersonal dimension of clinical practice, also associated with *phronesis*.

While *techné* and *phronesis* have been distinguished throughout this chapter, the purpose was not to establish an irresolvable dichotomy between these two fundamental forms of human action. Rather, this analysis suggests that the relation between *techné* and *phronesis* in clinical practice is fluid. A clinical action that may be considered at one moment as a form of *techné* may be considered a form of *phronesis* in another context, particularly if performed with another intention. Moreover, the sense of moral agency that is most associated with *phronesis* may also pertain to a lesser extent in more purely technical actions. As Gallagher observes, the "secondary contextualization of action in pragmatic and social settings . . . is necessary for both the development of expertise and the acquisition of *phronesis*" (2007, 210). Particularly, in medicine, it is not possible to ultimately separate purely technical actions from their moral ends (Braude 2012b).

Finally, in considering clinical expertise equally in terms of cognition, motoric activity, and interpersonal relations, this analysis has moved away from the theoretical conception privileging cognition in clinical reasoning. Elucidating pre-reflective categories such as affect and intentionality plays an important role in determining the nature of clinical expertise. Reconsidering cognition as an "essentially unitary phenomenon" (Cosmelli and Ibáñez 2008, 235) implies the need to reconsider cognition in cognitive science, as well as in the uniquely particular context of clinical practice. Arguably, the emphasis on cognition in the literature around medical expertise has prevented the full understanding of its noncognitive and nonconceptual dimensions, especially in terms of motor activity and intentionality. A key motivation behind this

chapter review of skilled know-how, virtuosity, and expertise in clinical practice has been to address this "cognitive" deficit.

Definition of Key Terms

Action	A movement of the body, resulting from specific mental preparation and aimed at some goal that the agent desires to achieve.
Cognition	All mental processes related to knowledge, including but not limited to memory, attention, perception, representational schemas, consciousness, and language.
Doctrine of double effect	Ethical principle that gives moral legitimacy to an equivocal action provided that the good and not harmful effect is intended, even though the unintended harmful effect may be foreseen.
Metacognition	The processes involved in thinking about and monitoring one's own cognitive activity.
Phenomenological introspection	Philosophical method to become self-aware of prereflective states of consciousness.
Praxis	Practical activity that is intended to further human well-being or the good and is not associated with any particular end product external to the act.
Poiesis	Any human activity that results in a product external to the human activity itself.
Phronesis	The moral capability to evaluate the means and ends of a particular action. In clinical reasoning phronesis affords the means of linking and integrating different cognitive processes.
Tacit Knowing	Knowledge that functions at the periphery of attention and makes explicit knowledge possible.
Techné	An action associated with craftsmanship, resulting in the production of external objects.

Summary Points

- This chapter analyzes skilled know-how, virtuosity, and expertise in clinical practice in terms of its three components, i.e., cognitive, motoric, and interpersonal expertise.
- These three identified components of clinical expertise are related to Aristotle's conception of practical wisdom, phronesis, and techné.

- Whether technical expertise is closer to phronesis or techné is determined through evaluating the ends of a specific action.
- If a clinical action has an end in itself that is self-referential, involving the moral virtue of the person doing the action, then the clinical action is most likely a form of practical wisdom.
- Analyzing clinical expertise equally in terms of cognition, motoric activity, and interpersonal relations makes a break from the theoretical conception of clinical expertise primarily in terms of cognition.

References

Anscombe GEM (1963) Intention. Cornell University Press, New York
Aristotle (1925) Nicomachean ethics (trans: Ross WD). Clarendon Press, Oxford
Barrows HS, Feltovich PJ (1987) The clinical reasoning process. Med Educ 21:86–91
Barrows HS, Tamblyn RM (1980) Problem based learning: an approach to medical education. Springer, New York
Benner P (1984) From novice to expert: excellence and power in clinical nursing practice. Addison-Wesley, Reading
Beresford EB (1996) Can phronesis save the life of medical ethics? Theor Med 17:209–224
Bernasconi R (1989) Heidegger's destruction of phronesis. South J Philos XXVIII(S1):127–147
Braude HD (2016) Clinical reasoning and knowing. In: Marcum J (ed) The Bloomsbury companion to philosophy of medicine. Bloomsbury Academic Press, London
Braude HD (2016) The affective limits of medical empathy. In: Bustan S (ed) Fundamental transdisciplinary questions on suffering and pain. Springer, New York
Braude HD (2012a) Conciliating cognition and consciousness: the perceptual foundations of clinical reasoning. J Eval Clin Pract 18:945–950
Braude HD (2012b) Intuition in medicine: a philosophical defense of clinical reasoning. The University of Chicago Press, Chicago
Braude HD (2013) Human all too human reasoning: comparing clinical and phenomenological intuition. J Med Philos 38(2):173–189
Cassell EJ (2015) The nature of clinical medicine: the return of the clinician. Oxford University Press, Oxford
Cervellin G, Borghi L, Lipps G (2014) Do clinicians decide relying primarily on Bayesian principles or on Gestalt perception? Some pearls and pitfalls of Gestalt perception in medicine. Intern Emerg Med 9:513–519
Charon R (2004) Narrative and medicine. N Engl J Med 350(9):862–864
Cosmelli D, Ibáněž A (2008) Human cognition in context: on the biologic, cognitive and social reconsideration of meaning as making sense of action. Integr Psychol Behav Sci 42:233–244
Croskerry P (2008) Cognitive and affective dispositions to respond. In: Croskerry P, Cosby K, Schenkel S, Wears R (eds) Patient safety in emergency medicine. Lippincott, Williams and Wilkins, Philadelphia, pp 219–227
Croskerry P, Abbass A, Wu A (2008) How doctors feel: affective issues in patient safety. Lancet 372:1205–1206
Decety J (Forthcoming) Empathy in medical practice: challenges and opportunities. In: Bustan S (ed) Fundamental transdisciplinary questions on suffering and pain. Springer, New York
Decety J, Svetlova M (2012) Putting together phylogenetic and ontogenetic perspectives in empathy. Dev Cogn Neurosci 2:1–24
Dreyfus SE (2004) The five stage model of skill-acquisition. Bull Sci Technol Soc 24(3):177–181
Dreyfus HL (2006) Overcoming the myth of the mental. Topoi 25:43–49

Dreyfus HL, Dreyfus SE (1988) Mind over machine: the power of human intuition and expertise in the era of the computer, 2nd edn. Free Press, New York
Edelstein L (1967) Ancient medicine. The Johns Hopkins University Press, Baltimore/London
Elstein AS (2000) Clinical problem solving and decision psychology. Acad Med 75(10): S134–S136
Elstein AS (2009) Thinking about diagnostic thinking: a 30 year perspective. Adv Health Sci Educ Theory Pract 14:7–18
Fitts P, Posner M (1967) Human performance. Brooke/Cole, Belmont
Gadamer HG (1975) Truth and method. Seabury Press, New York
Gallagher S (2007) Moral agency, self-consciousness, and practical wisdom. J Conscious Stud 14 (5–6):199–223
Gatens Robinson E (1986) Clinical judgement and the rationality of the human sciences. J Med Philos 11:167–178
Groopman J (2007) How doctors think. Houghton Mifflin Company, Boston/New York
Haggard P, Johnson H (2003) Experiences of voluntary action. J Conscious Stud 10(9–10):72–84
Heidegger M (1997) Plato's Sophist (trans: Rojcewicz R, Schuwer A). Indiana University Press, Bloomington/Indianapolis
Henry SG (2010) Polanyi's tacit knowing and the relevance of epistemology to clinical medicine. J Eval Clin Pract 16(2):292–297
Hofmann B (2002) Medicine as practical wisdom (*phronesis*). Poiesis Prax 1:135–149
James W (1890) Principles of psychology. Henry Holt, New York
Jeannerod M, Frak V (1999) Mental imaging of motor activity in humans. Curr Opin Neurobiol 9:735–739
Kahneman D, Tversky A (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131
Mangan J (1949) An historical analysis of the principle of double effect. Theol Stud 10:41–61
Marcum JA (2012) An integrated model of clinical reasoning: dual-process theory of cognition and metacognition. J Eval Clin Pract 18:954–961
Masters RSW (1992) Knowledge, knerves and know-how: the role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. Br J Psychol 83:343–358
McGee G (1996) Phronesis in clinical ethics. Theor Med 17(4):317–328
Montgomery K (2000) Phronesis and the misdescription of medicine. In: Kuczewski MG, Polansky R (eds) Bioethics: ancient themes in contemporary issues. The MIT Press, Cambridge, MA, pp 57–66
Montgomery K (2006) How doctors think: clinical judgment and the practice of medicine. Oxford University Press, Oxford
Norman G, Eva K, Brooks L, Hamstra S (2006) Expertise in medicine and surgery. In: Ericcson KA, Charness N, Feltovich PJ, Hoffman R (eds) The Cambridge handbook of expertise and expert performance. Cambridge University Press, Cambridge, pp 339–354
Osler W (1905) Aequinimatas. P. Blakiston, Philadelphia
Pellegrino E, Thomasma DC (1981) A philosophical basis of medical practice. Oxford University Press, New York
Polanyi M (1962) Personal knowledge: towards a post-critical philosophy. Routledge and Kegan Paul, London
Polanyi M (1966) The tacit dimension. Doubleday & Company, Garden City, New York
Proust J (2013) The philosophy of metacognition: mental agency and self-awareness. Oxford University Press, Oxford
Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD (2004) Empathy for pain involves the affective but not the sensory components of pain. Science 303:1157–1161
Schaffner K (1994) Discovery and explanation in biology and medicine. University of Chicago Press, Chicago
Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schilcht T, Vogeley K (2013) Toward a second-person neuroscience. Behav Brain Sci 36:393–462

Stanovich KE (2011) Rationality and the reflective mind. Oxford University Press, New York
Stein E (1989) On the problem of empathy. ICS Publishers, Washington, DC
Sulmasy D, Pellegrino E (1999) The rule of double effect: clearing up the double talk. Arch Intern Med 159:545–550
Thornton T (2010) Clinical judgement, expertise and skilled coping. J Eval Clin Pract 16:284–291
Toulmin S (1982) How medicine saved the life of ethics. Perspect Biol Med 25(4):736–750
Widdershoven-Heerding I (1987) Medicine as a form of practical understanding. Theor Med 8:179–185
Wiens S (2005) Interoception in emotional experience. Curr Opin Neurol 18:442–447
Woolley A, Kostopolou O (2013) Clinical intuition in family medicine: more than first impressions. Ann Fam Med 11(1):60–66
Zahavi D, Overgaard S (2012) Empathy without isomorphism: a phenomenological account. In: Decety J (ed) Empathy: from bench to bedside. The MIT Press, Cambridge, MA, pp 3–20

Meaning and Use of Placebo: Philosophical Considerations

43

Pekka Louhiala and Raimo Puustinen

Contents

Abstract

Confusion around the concept placebo and its derivatives is widespread, and the aim of this chapter is to elaborate the nature of these concepts in medicine. The historical development is first described. The current understandings and conceptual problems related to placebo are then examined. Finally, ways to clarify the ongoing conceptual disarray are proposed.

P. Louhiala (✉)
Department of Public Health, University of Helsinki, Finland
e-mail: pekka.louhiala@helsinki.fi

R. Puustinen
Medical School, University of Tampere, Tampere, Finland
e-mail: raimo.puustinen@uta.fi

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_34

> . . .placebo effect, probably the most fascinating and misunderstood aspect of human healing, which goes far beyond a mere sugar pill: it is counterintuitive, it is strange, it is the true story of mind-body healing, and it is far more interesting than any made-up nonsense about therapeutic quantum energy patterns. (Goldacre 2009, p. xi)

Introduction

> In clinical treatment of patients, the knowing prescription of placebos has not been used in medicine for many years. (Shapiro and Shapiro 1997)
>
> Placebos are commonly used in UK primary care. (Howick et al. 2013)
>
> It is suggested that in select cases, use of placebo may even be morally imperative. (Lichtenberg et al. 2004)
>
> Clinical placebo interventions are unethical, unnecessary, and unprofessional. (Hróbjartsson 2008)

As the quotations from recent medical literature indicate, there is fundamental disagreement both about the prevalence of the use of placebos in clinical medicine and the ethics of such practice. The disagreement is at least partly due to the ambiguous nature of the concept placebo and its derivatives. That is, when the authors of the above quotations use the word placebo, they mean different things. Confusion around these concepts is indeed widespread, and the situation seems not to be any better today than three decades ago when Grünbaum (1986, p. 19) wrote: ". . .the medical and psychiatric literature on placebos and their effects is conceptually bewildering, to the point of being a veritable Tower of Babel."

The aim of this chapter is to elaborate the nature of the concept placebo and its derivatives in medicine. The historical development of the concepts is first described. Then, the current understandings and conceptual problems related to them are examined. Finally, ways to clarify the ongoing conceptual disarray are proposed.

From an Everyday Word to a Medical Concept

In the placebo literature, the first appearance of the word placebo has been generally attributed to *Vulgate*, the fourth-century Latin translation of the Old Testament by St Jerome, where, in the Psalm 116:19, we find the expression "Placebo Domino in regione vivorum" ("I shall please the Lord in the land of the living"). Since the word placebo is a common Latin expression, the first-person future indicative of the verb *placeo* (to please), this attribution is not historically plausible. The expression must have been in general use since the birth of the Latin language, and the word placebo can, indeed, be found in several Latin texts predating St Jerome for centuries, such as Petronius' Satyricon, Seneca's De Consolatione, and Martial's Epigrammata, to name a few (see References for internet sources).

In the Medieval Catholic church, especially in France, the word placebo was associated to the tradition of singing the Psalm 116 during funerals, "Singing the Placebo." That practice was commercialized when professional mourners started to attend the ceremonies and charge a fee for their performance. Placebo gained thus a profane meaning as a cunning flatterer, as vividly expressed in a fourteenth-century Merchant's Story by Chaucer (2001), where a character in the story was named Placebo.

With its profane meaning, the word placebo found its way also to the medical literature in the wake of the eighteenth-century enlightened scientific acumen. For example, in a text published in 1763, a British physician wrote about a local charlatan how "Placebo never saw a professor in his chair, nor never made up a Doctor's prescription. Without knowledge chemical or practical, he was said to understand the waters better *than them all*" (italics original) (Sutherland 1763 p. xxiii–xxiv).

In medicine the meaning of the concept placebo was extended from indicating charlatans to the use of ineffective treatments in an attempt to merely please the patient. In 1776, a Scottish physician W. Robertson wrote in his book *Observationes Miscellaneae Inaugurales de Vino Praecipue* how "Ex modo et quantitate quibus administratur, nihil nisi placebo effe conludere volo" (the medication given had nothing but pleasing effect to his patient) (Robertson 1776).

A rather modern definition of placebo as an ineffective medication or treatment was given by another Scottish physician Andrew Duncan in 1752: "Where a *placebo* merely is wanted, the purpose may be answered by means, which, although perhaps reduced under the *materia medica*, do not, however, deserve the name of medicines. When a class of medicines, then, is said to be indifferent with regard to a morbid affection, nothing further is meant, than that is has no peculiar tendency to increase the evil; while, at the same time, no peculiar benefit can be expected from its employment" (italics original) (Duncan 1770).

The established use of the term placebo in the late eighteenth-century medical parlance is indicated, at least in the Anglophone world, with the term's entry to medical dictionaries. The first edition of the Motherby's New Medical Dictionary from 1785, for example, does not mention placebo, but in the third edition 1791 we find an entry "Placebo: A common place method or medicine" (Motherby 1785, 1791).

Similarly, Hooper's Medical Dictionary from 1798 does not include the term placebo, but it appears in the 1811 edition as "Placebo. I will please: an epithet given to any medicine adapted more to please than benefit the patient" (Hooper 1798, 1811). The explanation is identical with the one given in Coxe's Philadelphia Medical Dictionary in 1808 (Coxe 1808).

By the early nineteenth century, the term placebo seems to have been a part of physicians' everyday clinical vocabulary, as indicated in a scene in Sir Walter Scott's novel *St Ronan's Well*, published in 1823: "You mistake the matter entirely, my dear Mrs. Blower," said the doctor; "there is nothing serious intended – a mere placebo – just a divertisement to cheer the spirits, and assist the effect of the waters – cheerfulness is a great promoter of health" (Scott 1824).

The term placebo gained also pejorative meanings as can be seen in an Editorial of the Edinburgh medical and surgical journal in 1834 "...it would appear that the homoeopathic plan of treating diseases is totally inert, and can be useful only as a placebo to hypochondriacs and nervous women, by relieving them from swallowing the manifold drugs which they think is their duty to burden their stomachs" (Editorial 1834).

The term "placebo effect" has also long roots as can be seen in the 1776 quotation above. By the early twentieth century, the expression seems to have been in common use when discussing the outcome of treatments. An anonymous writer, for example, ponders in The International Journal of Surgery in 1900, how in the course of the assisted delivery "My rule is to prohibit all pulling in the first stage, and dispense with chloroform entirely, unless it be a little for its placebo effect – not enough to arrest contractions"(Anonymous 1900). Oswald, in turn, wrote in *The Sanitarian* in 1902, when discussing indigenous healing methods among Africans, that "there may have been a mere placebo effect about the procedure...where a victim of serpent bites was dosed with a decoction of boiled ants..." (Oswald 1902). In 1920 Graves published a case report of a 15-year-old boy with delayed puberty and epileptic seizures (Graves 1920). Various drugs had been given without a marked effect. Finally, Graves decided to try testicular extract in tablet form, and the boy's symptoms declined gradually, which, according to Graves, might have been related to the "placebo effects of the drugs given prior to admission."

The term "placebo effect" was made popular beyond medical circles in 1955 when Henry Beecher published his article "The powerful placebo" (Beecher 1955). Beecher was a strong supporter of randomized controlled trials (RCTs) designed with two arms, one receiving the active drug under investigation and one receiving some inert substance that was called placebo. His paper was both a review of the topic and a meta-analysis of 15 studies covering a wide variety of conditions. Beecher found a "relatively constant" therapeutic effectiveness of 35 % for placebos and concluded that it suggested "a fundamental mechanism in common" for placebos. That estimate became a standard reference to placebo effect both in subsequent trials and medicine in general. While in the 1950s RCTs developed rapidly to a "golden standard" in clinical research, placebo "changed from what was called the 'humble humbug' to an entity with occult-like powers that could mimic potent drugs" (Kaptchuk 1998). Since then, there has been a tendency to regard the placebo effect in the research context as a necessary background "noise" that must be subtracted from the results of a trial (Hunter 2007).

More recent derivatives of the term placebo entering into medical vocabulary are "pure" and "impure" placebo. The concepts were introduced in the 1940s at a Cornell Conference on Therapy, the proceedings of which were published in 1947 (Gold et al. 1947). In that publication, DuBois divided placebos into three classes: (1) pure placebos (e.g., bread pills or lactose tablets with no significant physiological effects), (2) impure placebos ("adulterated with a drug that might have some pharmacological action, such as tincture of gentian or a very small dose of nux vomica"), and (3) "the universal pleasing element which accompanies every prescription."

Current Understandings of the Concepts

Placebo

In November 2014 The *Wiktionary* defines placebo as "a dummy medicine containing no active ingredients; an inert treatment." This short definition seems to represent the overall current understanding of the concept placebo among both the lay people and the medical profession. The latter has, however, suggested wider and more complicated definitions for the concept placebo and its derivatives.

Chaput de Saintonge and Herxheimer (1994), for example, expand the realm of placebo "to the causes of the aggregated non-specific effects of treatments when specific effects have been segregated." This characterization seems to cover practically all elements of therapeutic encounter, except a specific pharmacological or other physiological mechanism. A table in the paper classifies placebos into eight main classes: (1) scars; (2) pills, tablets, and injections; (3) appliances; (4) touch; (5) words; (6) gestures; (7) local ambience; and (8) social interventions.

More recently, Benedetti (2009) has included the context of treatment into the definition of placebo:

> . . .a placebo would be better defined as an inert treatment plus the context that tells the patient a therapeutic act is being performed.

The American Medical Association brings the beliefs of the physician into the definition, when it defines placebo as "a substance provided to a patient that the physician believes has no specific pharmacological effect upon the condition being treated" (AMA 2007).

The concepts pure and impure placebo were hardly ever mentioned in medical literature for decades after their introduction, and even in the context of empirical placebo research, the concept has been used only recently (Fässler et al. 2009; Howick et al. 2013). According to Howick et al. (2013), "Pure placebos are interventions such as sugar pills . . . or saline injections without direct pharmacologically active ingredients for the condition being treated. Impure placebos are substances, interventions or 'therapeutic' methods which have known pharmacological, clinical or physical value for some ailments but lack specific therapeutic effects or value for the condition for which they have been prescribed."

Placebo Effect

Also for the notion "placebo effect," several different definitions have been proposed.

Shapiro and Shapiro (1997), for example, define placebo effect as "primarily the nonspecific psychological or psychophysiological therapeutic effect produced by a placebo, but may be the effect of spontaneous improvement attributed to the placebo."

Miller and Kaptchuk (2008) have suggested that the placebo effect should be reconceptualized as "contextual healing." By this they refer to the context of the clinical encounter, as distinct from the specific treatment interventions containing factors such as the environment of the clinical setting, the communication between patient and clinician, and the rituals of treatment.

Also Moerman (2002) has referred to the context when he has suggested that much of what is called the placebo effect is a special case of the "meaning response," which is defined as the physiological or psychological effect of meaning in the origins or treatment of illness. When such effects are positive, they include most of the things that have been called the placebo effect, and, when they are negative, they include most of what has been called the nocebo effect. Meaning response is attached to the prescription of active as well as inert medications and treatments.

Louhiala and Puustinen (2008) have suggested that "placebo effect" should be replaced with "care effect" to address the outcome of a therapeutic encounter that cannot be attributed to the specific physiological response to the treatment given.

Conceptual Problems in Placebo Literature

To be scientifically useful, theoretical concepts used in scientific enquiry need to be clearly defined and unambiguous. Yet, as can be seen in the examples above, there is no consensus within the scientific community on the definitions of placebo and its derivatives pure placebo, impure placebo, and placebo effect. In addition to the lack of consensus, some of the current definitions of those terms are internally incoherent.

When, for example, The American Medical Association (2007) defines placebo as "a substance provided to a patient that the physician believes has no specific pharmacological effect upon the condition being treated," it can be concluded that a substance (or method) may be a placebo today but not tomorrow (or vice versa), depending on the beliefs of the physician. Equally, a substance (or method) is a placebo when given by Dr. A (who believes it to be ineffective) but not when given by Dr. B (who believes the contrary).

On the other hand, if the Wiktionary definition of placebo as an inert treatment is agreed upon, the logical problem with using the term placebo effect is obvious. If placebo has, by definition, no effect, how could a placebo effect exist? This logical fallacy can be found even in a standard textbook *The Powerful Placebo* (Shapiro and Shapiro 1997), where placebo is defined as:

> any treatment . . . that is used for its ameliorative effect on a symptom or disease but that actually is ineffective or is not specifically effective for the condition being treated.

Further on, placebo effect is defined as:

> primarily the nonspecific psychological or psychophysiological therapeutic effect produced by a placebo, but may be the effect of spontaneous improvement attributed to the placebo.

If the word placebo in the latter definition is replaced with its definition above, the following "definition" of placebo effect is obtained:

> . . .therapeutic effect produced by [a treatment] . . . that actually is ineffective or is not specifically effective for the condition being treated. . .but may be the effect of spontaneous improvement attributed to the [treatment].

The first part is not meaningful and the second part limits the therapeutic effect to spontaneous improvement only (Moerman 2002; Puustinen and Louhiala 2014). Shapiro and Shapiro, like many other authors, use the term "nonspecific" referring to the alleged result of the placebo effect. This, however, refers only to the fact that we do not know what takes place in a therapeutic encounter when a patient feels better even he or she has not received any biologically plausible treatment.

The same confusion resides in using terms pure and impure placebo. From the practical point of view, the concept "pure placebo" is usually clear and meaningful. Technically, however, "pure" placebos are not without *any* effects since biological substances are never completely inert. Saline as such, for example, is practically inert when administered in small doses, but in intramuscular dosing the *procedure itself* is far from inert. Vitamin C and lactose have been used as "placebos" although they certainly have meaningful biological effects in many circumstances.

The category of "impure placebos" is more ambiguous and, in fact, extremely problematic. The list of treatments that have been categorized as impure placebos is long (Howick et al. 2013), but only three examples are enough to demonstrate the problematic nature of the concept: antibiotics for suspected viral infections, non-essential physical examinations, and positive suggestions. All these have been mentioned as examples of impure placebos in several empirical studies during recent years.

Antibiotics for suspected viral infections. If a physician *knows* that an infection is caused by a virus, prescribing antibiotics is clearly unethical. In real life, however, it is practically impossible to be certain about the cause of an infection. Medical decision-making is always based on probabilities, and in individual cases the physician weighs the potential gains and harms of the prescribed treatment (Louhiala 2009).

Nonessential physical examinations and nonessential technical examinations of a patient. The spectrum of physical or technical examinations that are "essential" for a particular patient is highly dependent on the context. The experience of the physician and the setting of the consultation, for example, define the variety of examinations, and there is no clear line between "essential" and "nonessential" examinations.

Positive suggestions. The role of a physician is to inform, comfort, and give hope to the patient, and positive suggestions are an essential element of this activity. It is not meaningful to describe it as a "'therapeutic' method" which lacks "specific therapeutic effects or value for the condition for which they have been prescribed."

As the category of impure placebos is highly ambiguous, we may ask whether dividing the concept placebo into categories "impure" and "pure" placebos is relevant in any scientifically fruitful way. While clinical practitioners may

prescribe, give, or recommend treatments that can be considered ineffective, labeling all such treatments as impure placebos paints a simplistic picture of clinical reality.

Placebos in Clinical Practice

Several studies have suggested that the deliberate use of placebos is a common and widely accepted practice among physicians. The lowest and highest reported proportions of doctors who have prescribed or administered placebos in their clinical practice have been 20 % and 97.5 %, respectively (Louhiala 2012; Howick et al. 2013).

A closer look at these studies shows, however, that the conclusion about the popularity of the use of placebos is false or at least seriously misleading. Some of the studies have not provided a definition for placebo and it was thus up to the respondents to interpret what they considered to be a placebo in their practice. In an oft-cited questionnaire survey from Israel, for example, 53 % of the physicians reported using a placebo (Nitzan and Lichtenberg 2004). Because the key concept was not defined, the respondents may have understood placebo in at least four different ways: "First, placebo may have meant deliberate deception through the administration of an inert substance. Second, it may have meant giving an inert substance openly. Third, some may have thought of a situation in which the doctor or nurse believes that the drug works even though there is no supporting scientific evidence. Fourth, some respondents may have thought more about the placebo effect than the nature of the substance given" (Louhiala 2009).

Furthermore, studies that have defined placebo show a large variation in their definitions. In a questionnaire survey among internists in Chicago (Sherman and Hickner 2008), for example, the respondents were given several alternatives for the definition of a placebo. They could either give their own definition or choose between an intervention that is not expected to have an effect through a known physiologic mechanism, or an intervention not considered to have a "specific" effect on the condition treated, but with a possible "unspecific" effect, or an intervention that is inert or innocuous. The main finding of the study was that 45 % of the respondents reported that they had used a placebo in clinical practice. Given this broad variety of definitions and interpretations of the basic concept, the finding is not informative.

A neglected aspect in all of the empirical studies addressing the use of the placebo has been the clinical prescription or administration process. The choice of words is not trivial here: "administering," "prescribing," "recommending," and "ordering" are different issues.

Given the conceptual confusion in defining the concept placebo, it is not surprising that opposite views have been presented also about the acceptability of the use of placebos in clinical practice. Walter Brown, an American psychiatrist, wrote in 1998 that "we should respect the benefits of placebos – their safety, effectiveness and low cost – and bring the full advantage of these benefits into

our everyday practices" (Brown 1998). Asbjorn Hróbjartsson, a Danish clinical epidemiologist, concluded in his paper that "Clinical placebo interventions are unethical, unnecessary, and unprofessional" (Hróbjartsson 2008). Liechtenberg et al. (2004) have suggested that "in select cases, use of the placebo may even be morally imperative." Howick et al. (2013) have proposed further investigations to develop "ethical and cost-effective placebos."

The arguments in support of the use of placebos can be summarized as follows: "they work in clinical trials, are cheap, and have no side effects." This statement is not only an ethical argument but demonstrates also the conceptual problem that is so common: "they" work and "they" do not have side effects, even though "they" are supposed to be, by definition, inert.

It is widely believed that a beneficial response to placebo treatment requires deception or at least some kind of a "white lie" to the patient. Lying to the patient is, however, ethically problematic, to say the least. For example, if the patient later finds out that the physician has not told the whole truth about the treatment, there may be serious consequences not only for the present physician-patient relationship but also for future relationships with other health-care professionals.

Some recent studies examining open-label use of placebos have suggested that a beneficial response to placebo treatment is not necessarily limited to settings where the patients have been deceived. As our knowledge on this topic is thus far very limited and an open-label use of pure placebos remains anyway a special case, no general recommendations can be given on such use.

Clarifying the Conceptual Confusion

The conceptual problems related to placebo and its derivatives have been acknowledged, and different solutions have been proposed to resolve the confusion.

On one hand, it has been suggested that the concept of placebo should be discarded altogether (Götzsche 1995) or that it should be limited to research context only (Louhiala and Puustinen 2008). In the latter case the term would refer only to the procedures and substances that are used as biologically inert controls to active treatments in medical research. If inert treatments are used in clinical practice, they should not be called placebos but ineffective treatments.

As reported earlier, several alternative concepts have been proposed to replace "placebo effect." Miller and Kaptchuk's *contextual healing* and Moerman's *meaning response* avoid the logical problem within the term placebo effect. Contextual healing refers to the therapeutic encounter on the whole which is, by necessity, always contextual in one way or another. Meaning response goes even further in addressing the essence of therapeutic encounter as a process attempting to create meaning to understand and solve the patient's problem. These concepts have not, to our reading, been analyzed in depth nor adopted for wider use in medical writing (Puustinen and Louhiala 2014).

Care effect refers to the phenomena that take place within both research settings and clinical consultations leading to beneficial therapeutic outcomes in cases when

the medical treatment given cannot explain those outcomes in full (Puustinen and Louhiala 2014). "Care" and "caring" carry positive connotations (Tudor Hart and Dieppe 1996), but not the burden "placebo" has obtained as referring to something unreal or negative ("dummy," "sham," "inert"). The patient's experience of having been cared for is always real. A care effect may be evoked in a clinical trial, too, but clinical practice and research are fundamentally different settings. In a trial, care effect may be considered a confounding factor, in clinical medicine an ally.

Definitions of Key Terms

Placebo	Is commonly defined as an inert treatment. However, it is often understood more broadly, even to cover practically all elements of therapeutic encounter except a specific pharmacological or other physiological mechanism.
Placebo effect	Also has several different definitions. In the research context, it may refer to the change in a placebo group. In the clinical context, it usually refers to the changes in the patient's condition that cannot be explained by a specific pharmacological or physiological mechanism.
Contextual healing	Refers to the context of the clinical encounter, as distinct from the specific treatment interventions containing factors such as the environment of the clinical setting, the communication between patient and clinician, and the rituals of treatment.
Meaning response	Is the physiological or psychological effect of meaning in the origins or treatment of illness.
Care effect	Is the outcome of a therapeutic encounter that cannot be attributed to the specific physiological response to the treatment given.

Summary Points

- A placebo is commonly understood as a dummy medicine or an inert treatment.
- Within medicine, however, much wider and complex definitions have been provided.
- The so-called placebo effect is a complex phenomenon, and the use of a placebo is not a necessary condition for a placebo effect.
- Because of the problematic nature of the concept placebo effect, several alternatives have been suggested to replace it (e.g., contextual healing, meaning response, and care effect)
- The category of "impure placebos" is highly ambiguous, and dividing the concept placebo into categories "impure" and "pure" is not meaningful.

- Although several studies seem to suggest that the deliberate use of placebos is a common and widely accepted practice among physicians, a closer look at these studies shows, however, that the conclusion about the popularity of the use of placebos is false or at least seriously misleading.
- The deliberate use of placebos – understood as inert treatments – in clinical practice is not ethically justified.

References

American Medical Association Council on Ethical and Judicial Affairs (2007) Placebo use in clinical practice. http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion8083.page. Accessed 19 Nov 2014

Anonymous (1900) Int J Surg 9:238

Beecher H (1955) The powerful placebo. JAMA 159:1602–1606

Benedetti F (2009) Placebo effects – understanding the mechanisms in health and disease. Oxford University Press, Oxford

Brown WA (1998) The placebo effect. Sci Am 278(1):90–95

Chaput de Saintonge DM, Herxheimer A (1994) Harnessing placebo effects in health care. Lancet 344:995–998

Chaucer G (2001) The merchant's prologue and tale. Cambridge University Press, Cambridge

Coxe J (1808) The Philadelphia medical dictionary. Thomas Dobson, Philadelphia

Duncan A (1770) Elements of therapeutics. Drummond, Edinburgh

Editorial (1834) Edinb Med Surg J 42:483

Fässler M, Gnadinger M, Rosemann T et al (2009) Use of placebo interventions among Swiss primary care providers. BMC Health Serv Res 9:144

Gold H, Barr DB, Cattell M et al (eds) (1947) Cornell conferences on therapy: use of placebos in therapy. Macmillan, New York

Goldacre B (2009) Bad science. Fourth Estate, London

Götzsche P (1995) Concept of placebo should be discarded. BMJ 311:1640

Graves TC (1920) Commentary on a case of hystero-epilepsy with delayed puberty. Lancet 196:1134–1135

Grünbaum A (1986) The placebo concept in medicine and psychiatry. Psychol Med 16(1):19–38

Hooper R (1798) A compendious medical dictionary. Myrray and Highley, London

Hooper R (1811) A compendious medical dictionary. Longman, London

Howick J, Bishop FL, Heneghan C et al (2013) Placebo use in the United Kingdom: results from a national survey of primary care practitioners. PLoS One 8(3), e58247

Hróbjartsson A (2008) Clinical placebo interventions are unethical, unnecessary and unprofessional. J Clin Ethics 19:66–69

Hunter P (2007) A question of faith. EMBO Rep 8:125–128

Kaptchuk T (1998) Powerful placebo: the dark side of the randomised controlled trial. Lancet 351:1722–1725

Lichtenberg P, Heresco-Levy U, Nitzan U (2004) The ethics of the placebo in clinical practice. J Med Ethics 30:551–554

Louhiala P (2009) The ethics of the placebo in clinical practice revisited. J Med Ethics 35:407–409

Louhiala P (2012) What do we really know about the deliberate use of placebos in clinical practice? J Med Ethics 38:403–405

Louhiala P, Puustinen R (2008) Rethinking the placebo effect. Med Humanit 34:107–109

Martial. Epigrammata. http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2008.01.0506:book=3:poem=51&highlight=placebo. Accessed September 12, 2015

Miller F, Kaptchuk T (2008) The power of context: reconceptualizing the placebo effect. J R Soc Med 101:222–225

Moerman DE (2002) Meaning, medicine and the 'placebo effect'. Cambridge University Press, Cambridge
Motherby G (1785) A new medical dictionary or general repository of physic. Containing an explanation of the terms and a description of the various particulars relating to anatomy, physiology, physic, surgery, materia medica, pharmacy &c. &c. &c. Printed for J. Johnson, London
Motherby G (1791) A new medical dictionary or general repository of physic. Containing an explanation of the terms and a description of the various particulars relating to anatomy, physiology, physic, surgery, materia medica, pharmacy &c. &c. &c. Printed for J. Johnson, London
Nitzan U, Lichtenberg P (2004) Questionnaire survey on use of placebo. BMJ 329:944–946
Oswald FL (1902) Cosmopolitan health studies. Sanitarian 49:500–505
Petronius. Satyricon. http://sacred-texts.com/cla/petro/satyrlat/satl130.htm. Accessed September 12, 2015
Puustinen R, Louhiala P (2014) The paradox of placebo – real and sham in medicine. In: Louhiala P, Heath I, Saunders J (eds) The medical humanities companion, vol 3, Treatment. Radcliffe, Oxford
Robertson W (1776) Observationes miscellaneae inaugurales De vino praecipue. Balfour et Smellie, Edinburgi, p 56
Scott W (1824) St Ronan's Well. Archibald Constable, Edinburgh, p 210
Seneca. De Consolatione ad Helvium. http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:2007.01.0017:book=11:chapter=4&highlight=placebo. Accessed September 12, 2015
Shapiro AK, Shapiro E (1997) The powerful placebo – from ancient priest to modern physician. The Johns Hopkins University Press, Baltimore
Sherman R, Hickner J (2008) Academic physicians use placebos in clinical practice and believe in the mind–body connection. J Gen Intern Med 23:7–10
Sutherland A (1763) Attempts to revive ancient medical doctrines. A Millar, London, pp xxiii–xxiv
TudorHart J, Dieppe P (1996) Caring effects. Lancet 347:1606–1608

Philosophical Issues in Nanomedicine

44

Christian Lenk

Contents

Abstract

In this chapter, the emerging field of nanomedicine is examined from a philosophical point of view. Firstly, the introduction works out the broader context of nanotechnology in today's scientific culture and shows some utopian undertones in the public discourse on nanotechnology. Secondly, in the following section on practical applications, some examples for research activities in nanomedicine are described and discussed. The third section gives a short introduction into the discussion on language and metaphors in nanobiotechnology. In the following section, the ethical issues in the context

C. Lenk (✉)
Ulm University, Institute for History, Theory and Ethics of Medicine, Ulm, Germany
e-mail: christian.lenk@uni-ulm.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_32

of nanomedicine are outlined. These selected issues are (1) risk assessment, (2) personal and human identity, (3) human enhancement, and (4) distribution of benefits and risks in the context of the implementation of innovative applications in medicine. Finally, some conclusions sum up the discussion on nanomedicine in philosophy and deal with characteristic hopes found in the public in the context of this biotechnology.

Introduction

In the late 1990s, an influential report of the US National Science and Technology Council (NSTC) was published with the title "Nanotechnology: Shaping the World Atom by Atom." The subtitle especially can be seen as characteristic for the report's line of argumentation. Referring to biological and molecular mechanisms in nature, the project is sketched out, to gain technological advances by following the example of nature for human purposes in the nano-area (nanometer, one billionth part of a meter). In the words of the Nobel laureate Horst Störmer, who is cited as follows in the report: "Nanotechnology has given us the tools ... to play with the ultimate toy box of nature – atoms and molecules. Everything is made from it ... The possibilities to create new things appear limitless" (NSTC 1999, p. 1). From the philosopher's point of view, this seems to be, at first glance, simply the vision or dream of an engineer; the expert on science and technology studies may regard this as mere speculation. However, beyond such rather negative associations, there are maybe some echoes of the beginnings of Western philosophy, namely, pre-Socratic natural philosophy (or speculation) with the two famous representatives of atomism, Leukipp and Democritus. In reading the hopes and ideas in the NSTC paper, one senses a similar spirit of speculation regarding the inner forces of matter and a similar enthusiasm about the enigmas of the world in some subsequent scholars influenced by these early thinkers.

However, the *motifs* of "shaping the world" or "to play with the toy box of nature" also point to human (or expert) control over nature or indeed over other human beings. Perhaps, this is the reason why a wider audience, especially those in nongovernmental organizations (e.g., the German *Bund für Umwelt und Naturschutz* or *Friends of the Earth*), regards the abovementioned visions rather with suspicion than with sympathy. However, especially in the field of nanomedicine, there seems to be a considerable gap between the original and lofty visions of nanotechnology at the conceptual level and concrete projects in the medical area, which do not aim at shaping humans or organs de novo "atom by atom" but rather aim at pragmatic translations of technological ideas into applied projects. Certainly, this does not mean that nanotechnology in medicine is per se a harmless endeavor. But at least one can say that it is obviously an endeavor which has similar aims to medicine, i.e., the healing of patients and the fighting of diseases, and has therefore to be evaluated in a similar way, for example, regarding proper risk-benefit assessment.

Examples of Practical Applications

The philosophical analysis of new technologies and scientific approaches must certainly take into account the visions, goals, and projects which are described by the protagonists of nanomedicine themselves. However, beneath the explanations of the actors, it is also necessary to examine the practical endeavors and concrete applications which result from research in nanomedicine. Like in other cases, such a double perspective will typically reveal a number of similarities between practice and theory but also a number of important differences. The difference between the visions and goals of nanotechnology in general and the practical applications of nanomedicine in particular may be considerable due to the strict demands which are normally presented to medical applications in human beings. This typically includes the pharmacological and medical proof that the risks of using the new application are acceptable in relation to the medical goals sought and that the relevant medical device or active agent is indeed effective in a specified and verifiable way. In this sense, in the following, a number of examples will be presented of new inventions in the field of nanomedicine. As the examples show, all these approaches depend on the special characteristics of the tiny, i.e., nano, particles used. On the other hand, the examples show a broad range of different approaches. These differences will probably lead to heterogeneous results of risk analysis.

A first example is the use of magnetic nanoparticles for tumor therapy. In the course of this therapy, a huge number of these nanoparticles are injected into the tumor tissue. Subsequently, the particles inside the tumor are heated by an alternating magnetic field. The aim is here to achieve a particular method of tumor treatment which is superior to conventional surgery. Due to this aim, the new method was tested in patients with glioblastoma, a kind of brain tumor (Müller-Jung 2009). Another approach based on the use of magnetic particles is their combination with modified viruses. These so-called *Lentiviruses* are equipped with modified genes for therapeutic purposes, dependent on the patients' pathology. In animal experimentation, researchers have tried to navigate such genetic "taxicabs" into the coronary heart arteries of mice, where the genes would be incorporated by the target cells and then develop their therapeutic function (Müller-Jung 2009).

The regeneration of body tissue and cells is a general theme of nanomedicine. For example, researchers developed a gel, based on nanotechnology, for the regeneration and the growth of the cartilage in the human body. Such active agents could be important in order to cure degenerative diseases in the aging populations of industrial states. A comparable approach involves the artificial building of protein structures for the regeneration of blood vessels and angiogenesis, i.e., the building and growth of new capillaries or blood vessels. In this context, researchers speak of "protein imitation for medicine and biotechnology," the production of body-like nanostructures which can serve as a starting point for the regeneration of body structures of vital importance (Kurz 2011). An example of the practical application of such an approach is research on the therapy of patients with paraplegia (currently also on the stage of animal experimentation). In the experiment, an

active agent was transported by newly developed nanomolecules to the paraplegic lesions (in the spinal marrow) of dogs to alleviate the symptoms the researchers had previously induced by toxic chemicals (DÄ 2009; Shi et al. 2010).

From a research ethics point of view, this overview already illustrates a number of ethical problems such as, for example, experimentation with vulnerable patients (patients with brain tumors); the first-in-human use of newly developed substances; animal experimentation for basic research; the possible risk of growth stimulus for body tissues, which can also result in uncontrolled cancer growth of tissue and cells; and finally the possible risk of allergic and rejection reactions of the human body (cf., e.g., the dramatic Gelsinger case, where a patient died due to the infusion of a huge number of genetically modified viruses (Kimmelman 2008)).

After this short introduction into the field of nanomedicine, the following philosophical analysis will be structured in two main parts. Firstly, how can the language and descriptions of nanotechnology be classified from the perspective of the humanities? What developments in terms of the history of ideas can be identified? Secondly, how does the approach of nanomedicine need to be evaluated from an ethical point of view, and what ethical presumptions and principles are involved in such an evaluation? Both areas are permeated by anthropological issues, for example, whether the human body can be seen as a machine or whether a possible "improvement" of human beings is seen as ethically acceptable.

Some Considerations on the Language of Nanotechnology and the Theory of Science

As Köchy points out in his article on the *Conceptualisation of living systems in nanobiotechnology*, the presentation of nanotechnology in scientific and popular scientific contributions strongly follows the traditional machine model (or metaphor) of life. The decisive references in classical philosophy in this regard are, for example, René Descartes and Julien Offray de La Mettrie in the seventeenth and eighteenth centuries. In the famous passage in the *Discourse de la méthode*, Descartes argues as follows:

> This will not seem strange to those who know how many different automata or moving machines can be devised by human ingenuity, by using only very few pieces in comparison with the larger number of bones, muscles, nerves, arteries, veins and all the other parts in the body of every animal. They will think of this body like a machine which, having been made by the hand of God, is incomparably better structured than any machine that could be invented by human beings, and contains many more admirable movements. Descartes 1999 [1637], 39 f

The human and animal body is here compared with a machine, and the body parts and organs appear as a *comparandum* to the "pieces" which are the constitutive elements of the machine. The imagined thinkers or observers ("those who know," "they") "will think of this body *like a machine*" (emphasis by the author), but this body machine is far more perfect than the everyday machines produced by human

craftsman, due to the far more perfect constitution of its creator (i.e., God). However, as this is conceived by Descartes, it seems to be rather a distinction of degree than of quality. It is therefore perhaps not surprising to come to the conclusion that such a model or metaphor in the end is closer to an identification as opposed to a comparison. This is suggested in the following passage by George Tombs: "For Descartes, the metaphor became an abstract *equation*, linking the human body and the machine. And that equation became one of the pillars of an entire philosophical system. [. . .] Expressed another way, metaphor as analogy *likens* one thing to another; metaphor as equation affirms that one thing *is* another" (Tombs 2002, p. 168; emphasis by the author). The result is then a kind of confusion between the heuristic/didactic function of the mere model and the ontological function of an identification of two different categories of objects (the human body and the machine).

Similarly, Köchy draws the conclusion in his text that "The old debate concerning the relationship of machine and organism gets a new topicality and meaning in the context of nanobiotechnology. At the same time, the content of the previous machine conception changes" (Köchy 2008, S. 186 f., translation by the author). However, as he (Köchy) shows in a number of examples, the model is used in a rather unclear and undefined way which points maybe to the fact that it is used by the authors in an unconscious manner. A lack of awareness regarding linguistic and metaphorical distinctions seems also to be a problem in the description of concepts of nanotechnology in high-ranking international publications, as the following citations show:

> That biological motors perform work and are engaged in well-defined mechanical tasks such as muscle contraction or the transport of objects is apparent in all living systems. Controlling motion using molecular switches is particularly attractive for the construction of nanomechanical valves.

> The exquisite solutions nature has found to control molecular motion, evident in the fascinating biological linear and rotary motors, has served as a major source of inspiration for scientists to conceptualize, design and build – using a bottom-up approach – entirely synthetic molecular machines. Browne and Feringa 2006, pp. 32 f

In the first citation, it is not clearly recognizable, whether the text refers to muscle fibers of the human body, to which the metaphor of the engine is applied, or whether the topic is the molecular engine, which takes on the task of muscle fibers. Both subjects appear in principle exchangeable or identical.

In the second citation, nature appears virtually as an engineer, an agent who controls molecular motion and is therefore a prototype for the scientists who want to construct molecular machines. What Descartes initially insinuated of the human body (the metaphor of the machine) is in this citation an inherent part of the body and firmly fixed in its perspective of bodily functioning. "Nature" here takes on the teleological function, which was occupied by God in the seventeenth-century conception of the world (see Descartes' citation above, "They will think of this body like a machine which, having been made by the hand of God, . . .").

When the message of the traditional machine model of the human body in philosophy was "The human body is also a kind of machine and has therefore to

be analyzed and understood as such," the message of nanobiotechnology seems to be twofold. Firstly (according to the classic Cartesian premise), "natural entities are also a kind of machines," but then secondly "molecular machines can also be transformed in natural entities." The subtext is in the second case not so much the ambitious, presupposed possibilities of nanotechnology but the very possibility of exchange between nature and technology. One of the implications of such a conceptualization is therefore the tendency to blur the line between the traditional concepts of *physis* and *techne*, i.e., things from nature and artificial products. In the science literature, this is then frequently phrased as a quasi-ontological statement, where special qualities are ascribed to the products of nanobiotechnology.

Ethical Considerations

In a previous review article on the ethical aspects of nanomedicine, four main points were identified: (1) the problem of an adequate risk assessment in the case of innovative and first-in-human nanotechnology applications; (2) the area of personal and human identity; (3) a possible enhancement of human beings by nanomedicine; and (4) the distribution of benefits and risks which might result during the course of the introduction of nanotechnology into medicine (Lenk and Biller-Andorno 2007). The four topics seem still to be essential components in the understanding of the ethical analysis of nanomedicine and will be described in the following with reference to recent developments.

Risk Assessment

The description of risks and attempts to control adverse effects has a long tradition in medical research. However, different possibilities and foci in the ethical analysis of risk are conceivable. In medicine, the risk of a new drug's application is usually described as a pharmacological side effect. For example, the consequence of testing a new pharmaceutical drug for the patient could be that her cancer is cured or the disease progression is stopped but she will lose her fertility. Such consequences are then described according to the pharmacological paradigm and analyzed biostatistically for the whole patient population. However, the significance and meaning of loss of fertility for the patient from the familial, social, and psychological point of view are not described. Although it is increasingly the case that questionnaires concerning the quality of life of patients are used in medical research, this is not a sufficient clarification of patient risk from the ethical point of view. In fact, the mere medical-pharmacological risk analysis obscures crucial dimensions of harm and has therefore to be complemented by further ethical considerations. Such an approach should also be integrated into research in nanomedicine.

In the area of nanotechnology, it is also known that nanoparticles can pass into the environment. This will probably not lead to the so-called gray goo scenario

(cf. Drexler 1986, the worry that nanotechnology-based little organisms or replicators could convert the biosphere in an uncontrollable way into dust or "gray goo") but could nevertheless cause further pollution of the environment or the impairment of other organisms.

A further difficulty in the area of nanomedicine is the combination of risks in medical applications in the human body, as in the case of gene therapy, whereby the gene drug is supposed to reach the body cells encapsulated in nanostructures or so-called gene ferries (cf. Lenk and Biller-Andorno 2007). Gene therapy is connected with a number of severe risks (Kimmelman 2008), so that such an approach leads to the combination of two innovative applications, whereby yet unknown side effects could occur in interrelation. Therefore, some commentators and science journalists see the issue of adequate risk evaluation of nanotechnology in medicine as the decisive problem for the further development of this form of research (Müller-Jung 2009).

This assessment is further highlighted by the fact that insurance companies partly exclude nanotechnology from their insured risks. As an insurance expert for risk assessment explained in a recent interview, from the perspective of insurances (who have for the sake of risk calculation a major interest in the proper determination of possible damages), the risk assessment for nanotechnology lags behind the technological innovation at the present point of time (Allianz Global Corporate & Specialty 2013). In the interview, it is also pointed out that the structure of some nanotubes resembles asbestos fibers, and it was demonstrated that some nanoparticles pose a danger for the health of water organisms. Therefore, the risk assessment of nanotechnology (understood as an interdisciplinary endeavor between the respective scientific disciplines and ethics) has to be developed further to keep pace with the technological development.

Personal and Human Identity

New applications from nanotechnology in medicine could also alter our perception of the natural human body. These considerations have to be seen in the context of the first part of this chapter, where the natural body was addressed as a kind of machine. This also has some ethical implications, for example, when the body as a machine is changed or complemented by applications from nanotechnology. There are a wide range of implants for different organs and functions which are currently already applied in the human body. To mention only two examples: for patients who have lost a knee and lower leg, these body parts can be replaced by a computerized and motorized so-called C-leg. Another example of modern implants is the subcutaneous defibrillator which automatically gives an electric shock to the patient's heart in case of cardiac arrhythmia. These are only two examples of cases where modern prostheses have the ability to react semi-autonomously and according to internal steering algorithms regarding bodily movements or dysfunctions. However, these examples show the accuracy of the machine metaphor because obviously the body machine is here complemented by other (helpful)

real machines. It depends on the individual patient's ability to adapt to these implants whether he or she will feel that his identity has changed as a result of these prostheses and implants.

In the case of nanomedicine, changes in the human body will be probably less obvious and clear-cut. However, this does not mean that they could not have an impact regarding personal and human identity. How will patients react on the possibility of infusing small nanomachines into their blood system with the aim of clearing blood vessels and capillaries? Is this substantially different from conventional drugs against plaque deposits in the blood vessels? The US National Science and Technology Council's report mentions the following possible applications in this regard:

> Nanotechnology will lead to new generations of prosthetic and medical implants whose surfaces are molecularly designed to interact with the body. Some of these even will help attract and assemble raw materials in bodily fluids to regenerate bone, skin or other missing or damaged tissues. New nanostructured vaccines could eliminate hazards of conventional vaccine development and use, which rely on viruses and bacteria. Nanotubules that act like tiny straws could conceivably take up drug molecules and release them slowly over time. A slew of chip-sized home diagnostic devices with nanoscale detection and processing components could fundamentally alter patient-doctor relationships, the management of illnesses, and medical culture in general. (NSTC 1999, p. 8)

These scenarios already point to the next ethical theme, namely, a possible enhancement of the human body (over and above mere therapeutic medical goals). Positively interpreted, some of the applications mentioned in the NSTC report (automatic and steady regeneration of body tissue, artificial drug secretion) could be seen as a kind of integration of therapeutic mechanisms into the human body which perhaps leads to a new form of "gentle" medicine. Other ideas, for example, the automatic monitoring of body functions and transfer of medical measurements to a physician or control center, could clearly change conventional ways of human living and could also have an impact on human identity. Surely, in the context of today's modern societies, this would be a voluntary decision of the person or patient her- or himself. However, there are, in the context of health care, a number of developments which show a societal dynamic of their own, where the individual person has to conform for not being excluded from the societal "fabric" or to be disadvantaged in the access to health-care services. In any case, a steady and automatic monitoring of body functions would lead to the situation that the concerned persons are permanently "accompanied" by a medical surveillance team. When the right to privacy also includes a "right to be let alone," this might lead to the abandonment of such a right and a decisive change in human living (cf. Hall et al. 2012, p. 769).

Human Enhancement

Ideas and doubts concerning the possibility of human enhancement by nanotechnology might be the area where there is the largest gap between a science-fiction

description of enhancement and currently existing applications of nanomedicine (cf. Lenk and Biller-Andorno 2007, p. 179). One has also to consider that nontherapeutic or enhancement activities in medicine exist independently of the possibilities of nanomedicine. The extension of biomedicine from the therapeutic occupation with existing diseases toward the improvement of bodily and mental qualities and functions is not initiated or fostered by nanomedicine as such. However, as was also mentioned in the last paragraph, there are a number of imaginable but yet not existing applications of nanomedicine which have to be classified not as a therapy but as an enhancement (for a further distinction between the two areas, cf. Lenk 2002). A number of applications of nanotechnology are described particularly in the field of rehabilitative medicine; they include, for example, the reconstruction of human tissue or body material which has deteriorated or disappeared due to degenerative or aging processes. In this medical field, nanotechnology applications could well lead to an "enhancement" of the human body, when regenerative mechanisms are strengthened or complemented. However, such a form of enhancement would probably not be seen as ethically problematic, although it could change, very gradually, important qualities of the "conditio humana" such as the normal process of human aging.

Distribution of Benefits and Risks

There are several points to consider in relation to justice in the introduction of new medical interventions, which are also relevant for nanotechnology in medicine. Firstly, international documents on research ethics such as the Declaration of Helsinki (cf. Art. 34) foresee that patients who take part in medical research studies should also profit from this participation and get a kind of reward for the connected risk and harm they might suffer. The authorization process of drugs in contemporary industrialized countries seeks to ensure that only verifiably efficient and safe drugs get to the drug market. However, this evidence does not exist in the case of early study phases in medical research studies. Therefore, patients in such research studies sometimes run a considerable risk of suffering from unexpected side effects. A well-known example is the so-called Gelsinger case for gene therapy, where a young volunteer with a slight disease, due to an idealistic motivation, took part in an experimental study and died because of a dramatic conjunction of scientific ambition, a problematic study design, and preclinical studies which were not significant enough. This shows, on the one hand, the urgent need to minimize risk for participants in such studies by appropriate measurements and, on the other hand, the need for an effective political regulation. Research promotion by laissez-faire deregulation can in such cases be extremely harmful for the patients concerned and study participants. Unfortunately, it is not always an advantage to be part of the scientific or medical avant-garde.

Secondly, some authors argue that the development of nanomedicine could contribute to widening the gap of medical supply between the industrialized and developing countries (Hall et al. 2012, p. 775). From the ethical point of view, such a claim makes sense if there is a general right to adequate health care and

participation in medical progress. From the medical point of view, such a claim for participation in medical progress is only reasonable when new applications bring a significant improvement in health care and the concrete health circumstances of patients. The authors of the named article give the example of a new medical device ("optical colposcope"), based on nanotechnology, which could enhance the diagnosis of cervical cancer (ibid.). In this context, they formulate the idea that the demand for justice in health care could be extended to include the research process itself (i.e., the selection and promotion of projects which focus on the health demands of the population in developing countries). Such an endeavor could then be integrated into national or international research funding, for example, on the part of the European Union. On the other hand, it seems to be rather unrealistic at the present point of time to expect that existing medical commercial companies would make such a commitment.

Conclusion

To draw some conclusions, one has firstly to see that all kinds of ethics and technology assessments are based on considerations concerning societal and technical developments in the future. The majority of the scenarios described in the literature will probably prove inaccurate because they are not based on a sound data basis, are more visionary than accurate from a methodological point of view, and are biased by the hopes and expectations of their authors and protagonists. If, for example, philosophers are invited to express a view of how the application of philosophy in education programs could change our society and make the world a better place, they would probably also draft visionary and mainly positive ideas. A comparable phenomenon occurs in nanotechnology, when protagonists as stakeholders of technology and possible recipients of research funding and investment present their technical plans and ideas.

This shows at the same time the importance of a proper, systematic, and interdisciplinary methodology of technology assessment, which also focuses on the broader societal implications and consequences. Because the real impact of a new technology can in most cases be revealed only in a broader perspective (this is also impressively demonstrated by the way the digital revolution and the internet change today's society and economy), a narrow approach which exclusively focuses on persons directly involved will probably not adequately describe the truly significant changes. Misguided expectations in regard to nanotechnology lead to what Wiesing and Clausen call in a recent article the "three dubious hopes in the context of nanomedicine":

- Firstly, that "[n]anobiotechnology will individualize therapy" (a dubious hope because also with nanomedicine it will be too costly to develop a drug for a single person or a small group),
- Secondly, that an "intervention [based on nanotechnology] is causal and therefore successful at the nanolevel" (a dubious hope because other therapeutic approaches

like for example the administration of insulin for diabetes patients are equally "causal", but hence not unproblematic or without unintended side-effects),
- and thirdly, interventions based on nanomedicine are "carried out with precision at the nano-level" (a dubious hope because such a promise is always brought forward with the introduction of any new medical technology, but this can only be evaluated after practical proving of a new approach). (Wiesing and Clausen 2014, p. 21)

This leads then, as Wiesing and Clausen explain, to the frequent doom and gloom scenarios for new technologies (ibid., 22). As was demonstrated in this chapter, for a comprehensive understanding of nanotechnology and its ethical implications, an interdisciplinary approach is necessary which in principle has to start from a linguistic perspective. The citations and remarks on the human-machine metaphor also show that the currently existing language in nanotechnology runs risk to confuse the metaphorical and the factual level and makes it difficult to distinguish existing scientific progresses from metaphorical and conceptual considerations.

Definitions of Key Terms

Nanomedicine	Medical applications which are based on mechanisms in the nanosphere (10 exp −9 m).
Scientific language	The specific language used in science in contrast to everyday language.
Risk assessment	A systematic assessment of an action's or project's negative side effects.
Applied ethics	Branch of ethics which is devoted to the analysis of ethical problems in practical circumstances, mostly in specific societal spheres (economy, medicine, science, trade, etc.).
Techno-utopianism	Ideology or system of belief which postulates the accessibility of utopian aims by technological means.

Summary Points

- Nanomedicine and nanotechnology did raise a number of fundamental questions in philosophy, ethics, and theory of science in the last 15 years.
- Characteristics are a metaphorical language of nanoscience, utopian undertones in the formulation of scientific goals, and especially comprehensive therapeutic hopes in the context of nanomedicine.
- An overview about current practical applications shows the arrival of nanomedicine in clinical applications with concrete patient groups.
- Among the ethical fields of discussion and analysis in the context of nanomedicine are risk assessment, personal and human identity, human enhancement, and the distribution of benefits and risks.

- Due to the evolutionary development of nanotechnology, a final ethical assessment of nanomedicine altogether is not possible at the present point of time, but a consolidated and comprehensive ethical assessment is carried out in this chapter.

References

Allianz Global Corporate & Specialty (2013) Die Nanotechnologie ist aus unserer Welt nicht mehr wegzudenken. [We cannot imagine our world without nanotechnology] Interview with Michael Bruch. Munich. www.allianz.com/de/presse/news/geschaeftsfelder/versicherung/news_2013-01-30.html. Accessed 16 Sept 2014

Browne FR, Feringa BL (2006) Making molecular machines work. Nat Nanotechnol 1:25–35

DÄ (2009) Nanomedizin: Heilung von Querschnittsgelähmten oder unkalkulierbares Risiko? [Nanomedicine: healing of paraplegics or incalculable risk?] DeutschesÄrzteblatt

Descartes R (1999 [1637]) Discourse on method [Discourse de la method]. In: Discourse on method and related writings. Penguin Books, London, pp 1–54

Drexler E (1986) Engines of creation: the coming era of nanotechnology. Anchor Press/Doubleday, New York

Hall RM, Sun T, Ferrari M (2012) A portrait of nanomedicine and its bioethical implications. J Law Med Ethics 40(4):763–779

Kimmelman J (2008) The ethics of human gene transfer. Nat Rev Genet 9:239–244

Köchy K (2008) Konzeptualisierung lebender Systeme in den Nanobiotechnologien [Conceptualisation of living systems in nanobiotechnology]. In: Köchy K, Norwig M, Hofmeister G (eds) Nanobiotechnologien Philosophische, anthropologische und ethische Fragen [Nanobiotechnologies. Philosophical, anthropological and ethical questions]. Verlag Karl Alber, Freiburg/München, pp 175–201

Kurz S (2011) Neue Blutgefäße mit Nanomedizin. Knorpelheilung, Tumortherapie und Wirkstofftaxis: Winzigste Partikel werden zu Helfern [New blood vessels with nanomedicine. Cartilage healing, tumour therapy and taxicabs for active agents: tiny particles become helpers]. Die Welt

Lenk C (2002) Therapie und Enhancement. Ziele und Grenzen der modernen Medizin [Therapy and enhancement. Aims and limits of modern medicine]. Lit Verlag, Münster

Lenk C, Biller-Andorno N (2007) Nanomedicine – emerging or re-emerging ethical issues? A discussion of four ethical themes. Med Health Care Philos 10(2):173–184

Müller-Jung J (2009) Heiler im Schattenreich [Healers in the realm of shades]. Frankfurter Allgemeine Zeitung

National Science and Technology Council (NSTC) (1999) Nanotechnology: shaping the world atom by atom. Washington, DC. http://www.wtec.org/loyola/nano/IWGN.Public.Brochure/

Shi Y, Kim S, Huff TB et al (2010) Effective repair of traumatically injured spinal cord by nanoscale block copolymer micelles. Nat Nanotechnol 5:80–87

Tombs G (2002) Man the machine: a history of a metaphor from Leonardo da Vinci to H.G. Wells. Doctoral thesis, McGill University, Montreal

Wiesing U, Clausen J (2014) The clinical research of nanomedicine: a new ethical challenge? Nanoethics 8(1):19–28

Philosophy of Sports Medicine

45

Silvia Camporesi and Mike McNamee

Contents

Abstract

The focus of this chapter is on the philosophy of Sports Medicine, that is, the practice of medicine in the context of sport. The chapter begins by examining ways in which a distinction in kind can be claimed between Sports Medicine and medicine per se. It does this by focussing first on the goals of medicine. This strategy proves to be indecisive, and it is concluded that a difference in degree only, rather than in kind, can be claimed for Sports Medicine. However, when the focus is directed to the normative aspects of medicine per se, in comparison with Sports Medicine, important differences can be identified. These differences

S. Camporesi
Department of Global Health and Social Medicine, King's College London, London, UK
e-mail: silvia.1.camporesi@kcl.ac.uk

M. McNamee (✉)
College of Engineering, Swansea University, Swansea, UK
e-mail: m.j.mcnamee@swansea.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_33

concern, especially, the way in which normative concepts central to medicine per se are operationalized in Sports Medicine. It is shown how norms regarding privacy, confidentiality, autonomy, and paternalism all apply in significantly different ways in the sporting context. Parallel differences are also identified in relation to the therapy/enhancement distinction. The problem of balancing current sporting goals against long-term health is also discussed.

Introduction

Sports Medicine is something of a paradox. On the one hand, a sufficiently similar practice to that which we now call Sports Medicine was practiced in the ancient cultures of Greece and Rome (Berryman 1992; Heggie 2011; Carter 2012). On the other hand, despite these venerable roots, it is fair to say that only during the latter half of the twentieth century that it started to seriously establish its professional credentials. Many sports and even some professional sports, even until very recently, had the most limited medical and healthcare resources (Howe 2004). What might have been called "Sports Medicine" in the highest football (soccer) leagues in Europe until the 1970s often consisted of a masseur and a trainer who carried on a bucket of cold water and sponge with (possibly) an analgesic spray. Team physicians were a much later advent.

Precisely who falls under the phrase sports medic is far from clear. The term sports physician is adopted here as the standard. This will typically refer to a medical doctor with some specialism in sports. Across the globe, there are a variety of standards and qualities of preparation, and some countries do not have a designated specialism with national standards and nomenclature. Thus, Sports Medicine more generally conceived and understood as the name of a community need not be restricted to registered medical practitioners but can also include physiotherapists (physical therapists), healthcare practitioners, dentists, and in some cases athletic trainers. Each of these occupations is likely to have varying professional standards, norms, codes of conduct, and other regulatory frameworks and goals. The focus in this chapter is on sports physicians, medically registered professionals, in order to bring some order and specificity to the discussion.

Merely being members of the medical professions brings a certain coherence and identity to the notion of Sports Medicine, but one cannot expect a high degree of overlap in aims and processes. The goals of medicine are, of course, contested (Allert et al. 1996; Callahan and Hanson 1999). Brulde's (2001) account of the goals of medicine is revealing. Surveying a range of institutions and policy frameworks, he identifies seven different and mutually irreducible goals. It is hardly surprising, then, that there is no agreement as to the nature and purposes of Sports Medicine. Some of the claims made on behalf of Sports Medicine range from political slogans to bloated commercial claims. Perhaps most bewildering of all is the claim that "exercise is medicine" (http://www.exerciseismedicine.org/support_page.php?p=113). This is of course a patently absurd idea. But it begs questions about the conceptual borders

of Sports Medicine that are rarely discussed since they are either taken for granted as unproblematic or relegated in priority by clinicians in order of a concentration on the main business of clinical work. Where sports physicians and scholars have made claims regarding the ethics of Sports Medicine, they have asserted that the ethics of Sports Medicine are "distinct" Green (2004) or "unique" (Johnson 2004; Testoni et al. 2013). The nature and ethics of Sports Medicine are critically discussed here, largely as a direct challenge to the unsubstantiated claims. A more modest proposal about the fiduciary obligations of sports physicians to their athletes and players (hereafter "athlete[s]") is presented and defended.

The Nature and Goals of Medicine and Sports Medicine

Although there is no uncontested essence to the concept, there is no good reason to think that is essentially contested. It is likely that there would be widespread agreement on very general ideas that the relief of suffering (Cassell 1982), or that the return to normal species functioning (Boorse 1975), are enduring features of the practice of medicine. The concept of medicine appears to have somewhat blurred and historically changing contours. Of course, even riverbeds shift – though very slowly.

Brulde notes seven independent goals: (i) to promote functioning; (ii) to maintain/restore normal structure/functioning; (iii) to promote quality of life; (iv) to save and prolong life; (v) to assist patients' coping with pathological conditions; (vi) to improve living conditions; and (vii) to promote children's growth and development (Brulde 2001). Each of these goals has something to recommend it as a claim to the nature of medicine as it is practiced today across the globe. Some appear more central than others; certainly much of Western medicine is in keeping with Boorse's general idea that health is to be understood as biostatistically normal functioning and that it is the job of medicine to secure and/or maintain this goal with and for the patient. Others, such as "improving living conditions" or "assisting patients coping with pathological conditions," might arguably fall more readily to associated branches of healthcare or welfare, respectively.

Some scholars, like Hoberman (2014) argue that we are witnessing the exportation of norms of Sports Medicine (enhancement) into mainstream medicine. Hoberman writes that "physician-assisted doping" has transformed high-performance sport into a "chronically overmedicated subculture" (Hoberman 2014, p. 572) that has been exported elsewhere ("the doping doctors of the sports world have pioneered 'entrepreneurial' medical practices that are now available to enormous numbers of people in search of hormonal rejuvenation").

This is not the place to substantively pursue the questions arising from the conceptual vagueness of medicine or Sports Medicine. It is just to note that there is no knockdown argument that we can employ about Sports Medicine's nature and ethics, without recourse to some nonneutral conception of medicine itself (Edwards and McNamee 2006). What can and should be done is to examine the claims made by the various constituencies of Sports Medicine on behalf of its medical status and its ethics.

Perhaps the boldest of all claims from within Sports Medicine is that exercise is itself a form of medicine, with or without physician assistance or intervention. But what, if any, sense is to be made of the slogan "exercise is medicine"? It can hardly be seen as some self-evident truth. First, it is noteworthy that the assertion is made not only by highly regarded professionals working in Sports Medicine in equally highly regarded scientific journals (e.g., Lobelo et al. 2014) on behalf of an international movement with a registered trademark "Exercise is Medicine®." So, perhaps it is best understood as nothing more than a slogan that captures a particularly modern set of pathological conditions that arise from sedentary lifestyles. Yet it should be noted, secondly, that the claim on the Exercise is Medicine (EIM) website, which has global policy and professional support, appears not merely to be that exercise is therapeutic or preventative of pathological conditions but that it is medicine in itself. Their mix of marketing and biomedical science appears to give the impression that exercise supplants traditional medicine in responding to the catalog of pathologies consequent upon inactivity. They continue, citing Robert N. Butler, MD, Former Director, National Institute on Aging, to the effect that "If exercise could be packed in a pill, it would be the single most widely prescribed and beneficial medicine in the nation." (EIM public presentation, slide 2, 20.3.15 http://www.exerciseismedicine.org/support_page.php?p=113). Thirdly, it is important to note that they propose implicitly, and explicitly on occasion, in their website pictures of physicians, and in publications, the idea that assessment and exercise referral is the province of the physician who is the legitimate mediator between the inactive (ergo pathological) populations and their exercise medicine. This of course is a highly contestable idea, one which physical educators, yoga practitioners, and health promotion officers might readily contest.

On more philosophical grounds one may query whether this colonization of leisure time is normatively justified or not. Though the idea of an obesity epidemic is questioned by some (e.g., Gard and Wright 2005; Gard 2010) there is widespread agreement that global health is indeed compromised by sedentary lifestyles. From this fact, if fact it is, the conclusion that exercise, presumably mediated by sports physicians, is the best or only response is of course highly contentious.

There is a further conceptual problem to consider. In the UK, and elsewhere, Sports Medicine as a profession has taken this turn towards exercise more generally rather than focusing exclusively on sport as a particular form of exercise. There may be excellent professional and political reasons for the adoption of this wider frame of reference. For example, a broader community of sport *and* exercise professionals could draw down greater funding from the state keen to keep individuals out of hospitals thus minimizing public expenditure; medical insurers in privately funded schemes might want to support this conceptual inflation because it is cheaper for them, and the fee-paying customer, to prescribe exercise over, for example, surgical intervention; by expanding their focus, the Sports Medicine community might acquire greater power over the lifestyles of citizens; and so on. This last benefit to the medical community has been more generally challenged under the construct of medicalization: the colonization of our lives by the medical profession (Parens 2013). Still, there are reasons pro and contra such conceptual inflation.

Nevertheless, the issues that arise from the adoption of a public health perspective into Sports Medicine are so heterogeneous that it is difficult to bring them into a singular conceptual framework. This heterogeneity brings further challenges in the context of ethical issues, since it would require an examination of public health ethics. In order to restrict the discussion, focus in the remainder of this chapter is on the medical issues arising from the more limited focus of Sports Medicine. In particular, it addresses the claims made regarding the distinctness or uniqueness of Sports Medicine among the family of medical professions.

Is There Anything Unique or Distinct About Sports Medicine?

The medical professions are many and varied. Nevertheless, it would be widely agreed that some of the occupations more readily claim to be at the center of medicine while others were more peripheral. For example, consider the contrast between general practitioners with cosmetic surgery. Bearing this in mind, few outside of Sports Medicine would not agree that it has enjoyed a kind of marginal existence and status. It is probably to be understood as undergoing what Habermas (1975) (albeit in a political context) called a "crisis of legitimation." In such a crisis, it is unclear how effective sports physicians might be in advancing their legitimacy claims. On the one hand, they might adopt a conservative strategy by advancing arguments that established their commonality with undisputed branches of medicine. On the other hand, sports physicians might formulate more ambitious claims regarding the distinctness or uniqueness of their clinical practice. If defensible, a claim regarding the "distinctness" or "uniqueness" (Dunn et al. 2007; Green 2004; Johnson 2004; Testoni et al. 2013) might be supposed to mitigate against the marginalization of Sports Medicine and the issue of its allegedly lowly status among the medical professions.

While the literature on the philosophy and ethics of Sports Medicine is not voluminous, there is widespread agreement on the central topics. A review of such a plethora of, largely spurious, claims includes: (i) treating pediatric athletes; (ii) medical advertising; (iii) innovative treatment; (iv) limits to patient confidentiality; (v) conflicting healthcare goals; (vi) enabling dangerous behavior; (vii) the physician-athlete relationship; (viii) privacy issues; (ix) concerns of autonomy; (x) informed consent; (xi) short-term gain, long-term risk; (xii) medical means to nonmedical ends; (xiii) drugs and the conflict of interest of the team physician; (xiv) effects of the cost of Sports Medicine care; and (xv) role of advertising in Sports Medicine. Time and space do not permit comment on all these claims, but a consideration of the more plausible contenders is presented below.

It will also be argued below that the claims of distinctness or uniqueness are overblown; what really exist are merely differences of degree, not differences of kind. Nevertheless, indeed *a fortiori*, a kind of transcendental argument can be used even before one considers these issues in detail. Suppose the claims to distinctness/uniqueness were true. One might reasonably ask how the proposers of the distinctness/uniqueness claim knew that the issues were then to be *bona fide* medical ones. Would

it not be the case, rather, that in virtue of being distinct or unique they would not be shared with other branches of medicine? And if that were the case, how could we vouch for their being *medical* at all? The claim to distinctness/ uniqueness thus turns out to be self-defeating. Indeed, a hope to solve the conundrums of Sports Medicine by analyzing the norms of Sports Medicine would be self-defeating as we would end up challenging the norms itself of medicine. By successfully demonstrating their difference, they must rescind claims to being medical. In any event, it would be worthwhile eschewing these claims and, after Wittgenstein (1953), considering the senses in which Sports Medicine shares family resemblances with others medical professions, displaying the degrees to which those resemblances are nuanced in particular cases.

Issues of Privacy and Confidentiality in Sports Medicine

It has been claimed that the physician-athlete relationship is a highly personalized one where the clinician must take the athlete patient's needs and goals seriously. On the one hand, the entire shift towards personalized medicine (chimerical or not) might undermine this bold claim. More prosaically, many general practitioners will say that their success or failure as a general practitioner may well hinge on the extent to which they treat the individual in front of them, and not the condition they present with, as the well-known saying goes. Moreover, certain parts of occupational medicine (such as might be enjoyed by pilots or chief executive officers in global businesses) would be predicated on their "personalized" approach. And of course the harrowing case of Conrad Murray, Michael Jackson's personal physician, regarding the claim to medicine's being personalized might well be framed as a professional failing. Part of a claim to highly personalized medicine will entail a consideration of the kinds of information that a physician may hold in relation to their patient.

Privacy issues are unique to Sports Medicine. On the face of it, this has little to commend this idea since privacy (or confidentiality to use a standard currency) is a widely shared norm across medical professions. But in Sports Medicine, like in many other branches of medicine, privacy is a nuanced issue. In some cases, the right of individual athletes is waived by contract, while in other cases it is breached by everyday norms of media reporting. So, in the first instance, National Football League players in the USA have – as part of their contract – waivers regarding privacy of data concerning injury status and treatment. This enables the media circus that attends most professionally commercialized sports to expose their product to the market in a variety of ways. And even where there is no contractual provision, such as in English Premiership Football, coaches, physiotherapists, and players discuss injuries and speculate all the time in public via radio or television (Ribbans et al. 2013). None of this is so different to discussions of politicians' health status or the fitness to perform in any given number of public roles. The claim qua personalized health seems unsustainable as a unique aspect of Sports Medicine.

There will also be occasions, similar to those experienced in occupational medicine or elsewhere, where a physician will divulge confidential health data to protect others. While cases such as sexually transmitted diseases are frequently used as

exemplars, team sports reveal a less discussed case in the light of athletes who have communicable diseases and ought not to share, for example, showers with other teammates or even simply sharing the field of play/court/ring and so forth.

Autonomy and Consent in Sports Medicine

A fairly counterintuitive claim has been made that concerns of autonomy generate uniqueness in Sports Medicine (Johnson 2004). Most medical ethicists or philosophers of medicine would think such a claim scarcely worthy of comment given the very widespread acceptance of the principle of respect for the autonomy of the patient. Now that seems almost trite were it not for the fact that many have queried athletes' desire to be autonomous in the face of complex, medically relevant, questions about their health and injury status; recovery times to training and participation in sports competitions; return to play decision (e.g., after concussions); and so on. Many athletes simply respond to their clinician when faced with a diagnosis and alternative treatment plans that they will go with whatever the "doc" recommends. And, of course, they are hardly unique in offering heteronymous responses. But if and insofar as athletes do want to be active and to have the final say in, for example, treatment interventions, then they will be aligned with general conceptions of best practice – at least within the mainstream of western medical ethics, where respect for autonomy is thought one of the foundational principles (Beauchamp and Childress 2012) and by some the first among those principles (e.g., Gillon 2003).

What may be present to an unusual degree in Sports Medicine is the extent to which individual athletes and players defer to their team doctors on treatment decisions. This should hardly surprise anyone since there is a considerable mutuality in their respective interests: the athlete/player wants to be at their fittest to compete, while the physician wants to enable optimal participation for the individual and/or their team. Nevertheless, two issues remain. First, the palpable existence of heteronymous athletes will trigger the well-known problem (Seedhouse 2008) of whether, or to what extent, it is the job of the physician not merely to respect autonomy but to foster it in their patients. Again, the problem is not unique to Sports Medicine but familiar. Secondly, in the increasingly globalized market for sports labor, it is interesting how issues of multiculturalism will affect the paternalistic-autonomy respectful dyad. Issues of linguistic competence (on behalf of the physician to explain and the patient to understand), wildly differing belief systems about causal efficacy from western pharmacology to witchcraft, and systems of authority and deference, combine to present sport physicians with exceptional challenges. Yet medical professionals working in general practice within multicultural societies will report sufficiently similar problems to undermine claims to uniqueness here.

What the increasingly multicultural nature of sports workforces highlights is the difficulties of gaining informed consent from their athlete patients. While informed consent reifies respect for autonomy, it may be overridden in conditions of incompetence. Incompetence (i.e., incapacity with respect to decision-making) in sports is likely to arise in a number of cases. Take just two: competence compromised

temporarily by head injury (McNamee and Partridge 2013; McNamee et al. 2015) and incompetence by virtue of immature reasoning powers. And of course there can be cases of the two (Webborn et al. 2015) but this does not generate new considerations, merely conjoining the two. In the first instance, there has been a surge in concern about concussion prevalence in contact sports (Clay et al. 2013) and the specific ethical issues that arise because of it.

Where paternalism might be thought obligatory in Sports Medicine is in the development of talent identification and development programs (Baker et al. 2013). Recent decades have witnessed the increasingly early specialization of athletic talents, at periods of life where children's life plans are both unformed and uninformed (Tymowski 2001). Given the complexity of the decision to focus or specialize on just one sport to the exclusion of other activities (including, but not limited to, other sports), the child or adolescent is likely to be thought incompetent in relation to the choice at hand. Can an average 8-year-old really tell that they want to become the next Andre Agassi or that they would prefer to specialize in gymnastics or playing a musical instrument where the choice is exclusive because of early specialization (Camporesi 2013; Camporesi and McNamee 2016)? This increasing problem is likely to be exacerbated by the claims of direct to consumer genetic testing in Sports Medicine (Webborn et al. 2015), which may attract "tiger parenting" in an attempt to secure the greatest marginal benefits for one's athletically gifted offspring.

Trading Present Sports Participation Against Long-Term Health

An issue that is likely to be found at the elite end of sports and Sports Medicine is the consideration of whether short-term gains are justifiable in terms of long-term risks. In his felicific calculus, Jeremy Bentham (1879) argued that ceteris paribus the nearness in time a pleasure was to be had – its propinquity – was a rational criterion for preference of one thing over another. But it seems that in the case of Sports Medicine there are different "goods" at play that become ranked in the utilitarian calculus of discounting future health for nearness of probability of winning. Cases like these abound because high-performance athletes are focused more on their athletic achievements now than their future health status. Therefore, they adopt a "win-at-all-costs attitude" as described by Krumer et al. (2011) that discounts future health for current athletic success.

Despite its ethical provenance, it is less easy, although not impossible, to find examples beyond Sports Medicine for this form of intervention that discounts future health for another nearer in time type good. Thus, for example, women may choose early IVF treatments with large doses of hormones that may compromise their health more generally conceived. In this case, the future health of their body is compromised for a different type of good, having a child. Moreover, self-harming behaviors such as smoking and alcohol consumption also harm the long-term health of the individual for the nearness of a different kind of good. Of course, though medically relevant, these are not interventions. Less mainstream examples might be drawn from cosmetic surgery where individuals seek interventions to satisfy

temporal desires for a particular physical appearance. Gender realignment surgeries and hormonal therapies may also harm the long-term health of the individual at the discount at a nearer in time kind of good viz sexual identity. A particularly challenging example might be elective amputation both in the context of Sports Medicine and outside. In the former context, it has been argued that some would elect to have transtibial surgery in order to become a paralympic athlete (McNamee et al. 2014). Outside the context of Sports Medicine, individuals also request for an otherwise limb amputation out of requests of "identity," under the umbrella of body identity integrity disorder (Müller 2009; Ryan 2009). The nature of the condition as a "genuine" medical disorder or not is currently under discussion (Giummarra et al. 2011). Uniqueness notwithstanding, it is certainly true that Sports Medicine more readily throws up cases where present high functioning is traded off against future good health. This risk-taking phenomenon is evident beyond sports in wider society of course. The extent to which famous, role model, athletes are driving this trend in cases of extreme sports, BASE jumping, solo mountaineering, as well as more prosaic activities such as football and rugby is a moot point, and certainly impinges upon questions of resources and public health. Equally uncertain is the role of Sports Medicine in facilitating risky endeavors.

Therapy, Enhancement, and the Use of Medical Means to Nonmedical Ends

The use of medical means to nonmedical ends in general philosophy of medicine raises again some particular status issues of the role of (sports) medicine in human enhancement (Edwards and McNamee 2006; Savulescu et al. 2011). It raises important questions about whether the traditional goals of medicine are therapeutic in nature (understood to include prevention) or whether they embrace nontherapeutic ends (Boorse 2015; Pellegrino 1999). A significant body of literature has arisen in the last decade concerning this issue generally and the normative force (or not) of the therapy/enhancement distinction.

Briefly, the therapy/enhancement distinction as referenced in the President's Council on Bioethics "Beyond Therapy Report" (2003) is based on Christopher Boorse's (1975, p. 77) definition of health as "normal species functioning," which defined enhancement beyond species typical functioning. According to Boorse's biostatistical theory of health (BST), health is "normal species functioning," which is the statistically typical contribution of all the organism's parts and processes to the organism's overall goals of survival and reproduction. Christopher Boorse argues that to be healthy is to function normally and that health is value-free.

Nevertheless, as demonstrated by many scholars including Scully and Rehmann-Sutter (2001), Kingma (2007), and Mills (2011), social and biological norms are inextricably linked and the biostatistical theory of species functioning cannot be not a value-free account of health, as the "norm" in a particular context contains social judgments together with biological facts. The concept of the "normal" which is considered to be value-neutral in Boorse's "normal species functioning" is not

actually value-free, as it has both descriptive and normative implications. The etymology itself of the word "normal" from the latin "normalis" is telling, as "normalis" was the word used to refer to "standing at a right angle," where "norma" was the carpenter's square. Indeed, in mathematics the word "normal" can still be used to mean "perpendicular." As argued by Catherine Mills (2011, 2013) building on Canguilhem (1978), biological and social norms are inseparable and irreducible, and the reference point for normal species functioning is not to be "deduced from nature" but it is a choice that includes social norms. In other words, the concept of the "normal" is a value judgment that cannot be grounded only in descriptive statements about nature. Hence, the therapy-enhancement distinction referencing to the normal species functioning as the demarcating axis implies a normative connotation and the "existence of a directed axis along which different human embodiments can be arranged in a proper order from "worse" to "better"" (Scully and Rehmann-Sutter 2001, p. 90).

The notion of how precisely "normal" is to be understood warrants a more extensive discussion that cannot be pursued here. For present purposes, however, it is worth considering the role the distinction plays in sport medicine. If we accept the definition of enhancement as going beyond normal species typical functioning (as in the *Beyond Therapy* Report), we could say that elite athlete serve as a benchmark for normal species functioning. By pushing the species boundaries to the limit in elite performance thanks to "physician-assisted doping," some scholars like John Hoberman (2014) argue that the benchmark for normal species functioning gets pushed too. Consider an example in which the therapy-enhancement distinction in Sports Medicine is challenged. Drugs prescribed for return to play such as cortisol are considered part of therapeutic use exemption and referred to as "recovery drug." But they actually represent a very good example of a drug that although used to "restore" a previous state of health (the state of health previous to the injury) confers a performance advantage which can be compared to the advantage conferred by a performance enhancing drug (e.g., testosterone, which is rarely if ever given a Therapeutic Use Exemption (TUE) certificate according to the protocol of the World Anti-Doping Code), even though its anabolic steroid effects are very similar to the one produced by cortisol, with the only exceptions being some sports such as power lifting and bodybuilding that have World Anti-Doping Agency compliant federations.

Hamilton and Dimeo (2015) provide a recent example of recovery drug that crosses the therapy/enhancement distinction. This is cortisol, a steroid that is used to enable a return to play after shoulder injury by baseball player Ryan Zimmerman. After injuring his shoulder in the summer of 2012, Zimmerman was able to go from "being one of baseball's worst hitters to one of its best" thanks to cortisol injections whose use was considered ethically justifiable as part of a recovery drug due to a therapeutic use exemption (in times of stress it allows the body to use stored energy in the muscles, liver, and fat tissue; it does not heal the injury but simply allows the athlete to play through it). The use of a recovery drug like cortisol under a TUE not only does not restore the body to a previous health state (as it simply allows the body to play through the injury without healing) but on the contrary has long-term implications for the health of the athlete. It has been demonstrated that athletes

often adopt a risky approach according to which they would sacrifice long-term health for short-term goal (Krumer et al. 2011), as highlighted above. The normative justification for this is far from straightforward. Indeed this is the conundrum of the Sports Medicine physician when confronted with the difficult decision of whether to prescribe or not cortisol (or similar "recovery drugs") to athletes who request it for a swifter return to play or training.

It should be born in mind that the T/E distinction, which is based on a biostatistical theory of health which itself presupposes a value-free concept of "normal species functioning", is in reality informed by social norms too. However, medicine and sports are two separate contexts, with different values at play. Douglas (2007) has highlighted that a drug which could confer performance advantage could be ethically justified in one context (outside of sport) but not in another (sport) because of the inherent values of the practice. Camporesi and McNamee (2012) argue along similar lines in reference to gene transfer to raise the tolerance to pain in the context of a clinical trial and of sports competition.

Returning to the testosterone case, we can see that in the context of sport it is allowed under guide of TUE (under the form of cortisol, which is an analogous of testosterone), but not for performance enhancement, because of the supposed validity of the T/E distinction. Outside the context of sport, testosterone is prescribed for supposedly "real" medical conditions such as hypogonadism where it functions as a recovery drug that also enhances performance. It is also increasingly prescribed as an "anti-aging" drug (Madrigal 2015). This second kind of prescription falls beyond the goals of medicine understood as restoration or preservation of health but presupposes a continuum between health and well-being as proper goals of medicine. This would lead us to discuss the goals of medicine and Sports Medicine and whether the traditional goals of medicine are therapeutic in nature (understood to include prevention) or whether they embrace nontherapeutic. According to Scripko (2010), the arguments that enhancement technologies do not belong to the proper scope of doctor's profession are historically inaccurate. Perhaps it will be best to follow Scully and Rehmann-Sutter (2001) who suggest abandoning the T/E as a global distinction and arguing on a case-by-case direct evaluation of the moral relevance of the distinction. So, for example, in the case of gene transfer to raise the tolerance to pain (Camporesi and McNamee 2012) one will have to evaluate the details of the biomedical technologies under discussion alongside the contextual values that inform the particular sporting practice that will form the basis of our ethical evaluation of each case (e.g., Green 2009; Murray 2009).

The Sports Physician and Their Fiduciary Relationship with Athlete Patients

In a notorious case of medical collusion with the team coach in order to help secure victory in a high profile European Cup rugby match, a British doctor once made an incision into the mouth of a player (at his request) in order to make it appear to third parties that his removal from the play had been for a legitimate blood injury. This had allowed the team to make an apparently legal substitution of a specialist kicker who

might win them the match in the dying minutes (Holm and McNamee 2009). The opposing team doctor, suspecting unfair play, followed them into the dressing room soon after the player's withdrawal and the plot was uncovered. Thereafter the scandal became known as "Bloodgate." The doctor attempting to cheat the officials was subsequently reprimanded, while the team physiotherapist who colluded in the deception was struck off the professional register of physiotherapists but reinstated on appeal. Interestingly, he vowed never to return to Sports Medicine and be confronted with pressures antithetical to the Hippocratic Oath.

Sohn and Steiner (2014) argue that the sports physician has an obligation arising from the Hippocratic Oath of nonmaleficence, and cases of assisted doping or return to play break this obligation as they harm the health of the athlete. But does a nonmaleficence obligation trump the other obligations that a sports physician may have (that arise out of the contract with the athlete/team), such as beneficence? This may need to be understood in the context of Sports Medicine as an obligation to optimize the athlete performance (make the athlete as fit as possible to compete). It could be argued that a broader understand of "benefit" needs also to be specified in this context which goes beyond the health to include other "goods" such as being as fit as possible to play.

Nevertheless, it raised the ire of the British Sports Medicine community many of whom had found themselves caught in the middle of the pressures to assist team performance (at any cost) and their traditional role to act as a fiduciary to their (athlete) patient. It even prompted the quoting of Shakespeare's Macbeth: "I am in blood, stepped so far …" wrote two physicians (Devitt and McCarthy 2010) acknowledging – after that the profession had been implicated in wrongdoing for so long that it could not see its way back.

Many of the problems that face Sports Medicine are highlighted in professional sports and perhaps exaggerated there under the influence of considerable sums of money. The issues that arise here, in addition to others concerning confidentiality and disclosure, license to practice and insurance cover for international sporting events beyond their registered jurisdiction, trustworthiness in the face of competing conflicts owed to players and their employers, are also exacerbated when the sport physician has no clear fiduciary duty to the best interests of their patient (Holm et al. 2011). Committing their services to the athlete patient will be the best means to assuage, though not necessarily to remove, the kinds of conflicts that arise when the sports physician serves two masters at the same time. But this too begs questions as to sports physicians' self-identity and vocation. To what extent should they be seen as a branch of occupational medicine, serving the welfare of co-employees for the employer, or acting as an independent fiduciary irrespective of the source of payment for their services (Holm et al. 2011).

Conclusion

The very nature of medicine and the role that health, illness and injury play in the lives of patients, means that ethical problems are likely to arise. Sports Medicine frequently resides in contested terrain because of the role that the body plays in

athletic performance, and the extremes of motivation to win with more or less attention to the welfare of players. Thus sports physicians must consider very general moral considerations that apply to all persons, but also how these are heightened in terms of the knowledge they have of the particular bodies of their athlete patients and because of the things they are allowed or requested to do *by* and *on* athlete patients. These problems are ethical but fundamentally conceptual too. It seems that it is precisely due to the lack of coherent self-understanding of the nature and goals of Sports Medicine that the ethical problems appear particularly, though not uniquely or distinctly, to be found there. Instead it has been shown here that the goals of Sports Medicine are no less contested than those of medicine itself.

Definitions of Key Terms

Sports Medicine	Medicine as practiced in the context of sport
Goals of medicine	That which medical practice hopes to achieve
Sports medic/physician	Medical doctor with some specialism in sports
Therapy/enhancement distinction	Referencing to the normal species functioning based on Boorse's biostatistical theory of health and used in applied ethics not without controversies to demarcate between ethically permissible and ethically impermissible application of a technology

Summary Points

- When focussing on the goals of medicine, no distinction in kind between medicine per se and Sports Medicine can be discerned; all that can be claimed is a difference of degree.
- But key norms in medical practice are operationalized differently in Sports Medicine.
- These include respect for patients' privacy and confidentiality for example.
- In Sports Medicine, these norms are standardly overridden and information regarding an athlete's health status may be given to third parties, for example, the sports media.
- Also, in Sports Medicine the relationship between achievement of current sporting goals is also controversial since athletes may compromise health status in later life by prioritizing short-term sporting success.
- The relationship between the doctor and the athlete also generates particular problems for the sports physician.

References

Allert G, Blasszauer B, Boyd K, Callahan D (1996) The goals of medicine: setting new priorities. Hastings Cent Rep 26(6):S1

Baker J, Cobley S, Schorer J (eds) (2013) Talent identification and development in sport: international perspectives. Routledge

Beauchamp TL, Childress JF (2012) Principles of biomedical ethics, 7th edn. Oxford University Press, New York

Bentham J (1879) An introduction to the principles of morals and legislation. Clarendon, Oxford

Berryman JW (1992) Sport and exercise science: essays in the history of sports medicine. University of Illinois Press, Urbana

Boorse C (1975) On the distinction between disease and illness. Philos Public Aff 49–68

Brülde B (2001) The goals of medicine. Towards a unified theory. Health Care Anal 9(1):1–13

Callahan D, Hanson MJ (eds) (1999) The goals of medicine: the forgotten issue in health care reform. Georgetown University Press

Camporesi S (2013) Bend it like Beckham! The ethics of genetically testing children for athletic potential. Sport Ethics Philos 7(2):175–185

Camporesi S, McNamee MJ (2012) Gene transfer for pain: a tool to cope with the intractable, or an unethical endurance enhancing technology? Life Sci Soc Policy 8(1):20

Camporesi S, McNamee MJ (2016) Ethics, genetic testing, and athletic talent: children's best interests, and the right to an open (athletic) future. Physiol Genomics 48(3):191–195. doi:10.1152/physiolgenomics.00104.2015

Canguilhem G (1978) On the normal and the pathological, vol 3. Springer

Carter N (2012) Medicine, sport and the body: a historical perspective. A & C Black

Cassel EJ (1982) The nature of suffering and the goals of medicine. N Engl J Med 306(11):639–645

Clay MB, Glover KL, Lowe DT (2013) Epidemiology of concussion in sport: a literature review. J Chiropractic Med 12(4):230–251. doi:10.1016/j.jcm.2012.11.005

Devitt BM, McCarthy C (2010) "I am in blood Stepp'd in so far. . .": ethical dilemmas and the sports team doctor. Br J Sports Med 44(3):175–178

Douglas T (2007) Enhancement in sport, and enhancement outside sport. Stud Ethics Law Technol 1(1)

Dunn WR, George MS, Churchill L et al (2007) Ethics in sports medicine. Am J Sports Med 35 (5):840–844

Edwards SD, McNamee M (2006) Why sports medicine is not medicine. Health Care Anal 14 (2):103–109

Gard M (2010) The end of the obesity epidemic. Routledge, Abingdon

Gard M, Wright J (2005) The obesity epidemic. Routledge, Abingdon

Gillon R (2003) Ethics needs principles – four can encompass the rest – and respect for autonomy should be "first among equals". J Med Ethics 29(5):307–312

Giummarra MJ, Bradshaw JL, Nicholls MER, Hilti LM, Brugger P (2011) Body integrity identity disorder: deranged body processing, right fronto-parietal dysfunction, and phenomenological experience of body incongruity. Neuropsychol Rev 21(4):320–333

Green SK (2004) Practice makes perfect? Ideal standards and practice norms in sports medicine. Virtual Mentor 6(7)

Green GA (2009) The role of physicians, scientists, trainers, coaches and other nonathletes in athletes' drug use. In: Murray TH, Maschke KJ, Wasunna AA (eds) Performance-enhancing technologies in sports. John Hopkins University Press, Baltimore, pp 81–96

Habermas J (1975) Legitimation crisis (trans: McCarthy T). Beacon Press, Boston

Hamilton L, Dimeo P (2015) Steroids in sport: zero tolerance to testosterone needs to change. The Conversation. https://theconversation.com/steroids-in-sport-zero-tolerance-to-testosterone-needs-to-change-48774. Accessed 13 Oct 2015

Heggie V (2011) A history of British sports medicine. Manchester University Press, Manchester/New York

Hoberman J (2014) Physicians and the sports doping epidemic. AMA J Ethics Virtual Mentor 16(7):570–574. http://journalofethics.ama-assn.org/2014/07/oped1-1407.html
Holm S, McNamee M (2009) Ethics in sports medicine. BMJ 339:b3898
Holm S, McNamee MJ, Pigozzi F (2011) Ethical practice and sports physician protection: a proposal. Br J Sports Med 45(15):1170–1173
Johnson R (2004) The unique ethics of sports medicine. Clin Sports Med 23(2):175–182
Kingma E (2007) What is it to be healthy? Analysis 67(2):128–133
Krumer A, Shavit T, Rosenboim M (2011) Why do professional athletes have different time preferences than non-athletes? Judgment Decis Mak 6:542–551
Lobelo F, Stoutenberg M, Hutber A (2014) The exercise is medicine global health initiative: a 2014 update. Br J Sports Med. doi:10.1136/bjsports-2013-093080
Madrigal AC (2015) Why testosterone is the drug of the future. Fusion. http://fusion.net/story/42619/why-testosterone-is-the-drug-of-the-future/. Accessed 3 Feb
McNamee M, Partridge B (2013) Concussion in sports medicine ethics: policy, epistemic and ethical problems. Am J Bioeth 13(10):15–17
McNamee M, Savulescu J, Willick S (2014) Ethical considerations in Paralympic sport: when are elective treatments allowable to improve sports performance? PM&R 6(8):S66–S75
McNamee M, Partridge B, Anderson L (2015) Concussion ethics and sports medicine. Clin Sports Med
Mills C (2011) Futures of reproduction: bioethics and biopolitics, vol 49. Springer
Mills C (2013) Reproductive autonomy as self-making: procreative liberty and the practice of ethical subjectivity. J Med Philos 38(6):639–656
Müller S (2009) Body integrity identity disorder (BIID)—is the amputation of healthy limbs ethically justified? Am J Bioethics 9(1):36–43
Murray T (2009) Ethics and endurance-enhancing technologies in sport. In: Maschke K et al (eds) Performance-enhancing technologies in sports. Ethical, conceptual, and scientific issues, pp 141–159
Parens E (2013) On good and bad forms of medicalization. Bioethics 27(1):28–35
Pellegrino ED (1999) The goals and ends of medicine: how are they to be defined?
Ribbans B, Ribbans H, Nightingale C, McNamee M (2013) Sports medicine, confidentiality and the press. Brit J Sports Med 47(1):40–43
Ryan CJ (2009) Out on a limb: the ethical management of body integrity identity disorder. Neuroethics 2(1):21–33
Savulescu J, ter Meulen R, Kahane G (eds) (2011) Enhancing human capacities. Wiley, Oxford
Scripko PD (2010) Enhancement's place in medicine. J Med Ethics 36(5):293–296
Scully JL, Rehmann-Sutter C (2001) When norms normalize: the case of genetic "enhancement". Hum Gene Ther 12(1):87–95
Seedhouse D (2008) Ethics: the heart of health care. Wiley, Oxford
Sohn DH, Steiner R (2014) Nonmaleficence in sports medicine. Virtual Mentor 16(7):539–541. http://journalofethics.ama-assn.org/2014/07/ecas3-1407.html
Testoni D, Hornik CP, Smith PB, Benjamin DK Jr, McKinney RE Jr (2013) Sports medicine and ethics. Am J Bioeth 13(10):4–12
Tymowski G (2001) Rights and wrongs: children's participation in high-performance sports. In: Berson IR, Berson MJ, Cruz BC (eds) Cross cultural perspectives in child advocacy. pp 55–93
Webborn N, Williams A, McNamee M, Bouchard C, Pitsiladis Y, Ahmetov I, Ashley E et al (2015) Direct-to-consumer genetic testing for predicting sports performance and talent identification: consensus statement. Brit J Sports Med 49(23):1486–1491
Wittgenstein L (1953) Philosophical investigations. 1967. Blackwell, Oxford

Part V

Medical Knowledge

Medicine as Art and Science

46

Kristine Bærøe

Contents

Abstract

Conceptual understanding of the essence of medical practice is important for many reasons. For example, it is crucial for how doctors interpret their role and effectuate it in practice, to help societies regulate and organize adequate provision of health care, and to enable critique of ongoing practice and identification of improved solutions for the future. Also, it is of importance to the medical

K. Bærøe (✉)
University of Bergen, Bergen, Norway
e-mail: kristine.baroe@igs.uib.no

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_35

profession itself as it helps distinguish medical practice from other healthcare practices as a way of supporting medical professionalism. Accounts of the essence of medical practice have extensively used the terms "art" and "science." However, the conceptual meanings of these terms are not obvious, and neither is it evident how one should perceive the relation between them. In this entry, various meanings of these terms will be addressed and their suggested internal relations in medical practice described. Finally, some practical and political challenges connected to one of the more comprehensive accounts are pointed out. In this way, the relevance of getting a firmer conceptual grip on the normative essence of medical practice is illustrated.

Introduction

Historically, discussions of medicine in terms of art and science are based on a conceptual understanding of medicine as *medical practice*. Thus, medical practice will also be the focus of this presentation. So what is the essence of practicing medicine? This question can be reformulated as both a descriptive and a normative question: What is the essence of medicine as it is in fact practiced? How should medicine ideally be practiced? The first question cannot be answered in isolation from descriptive accounts of how practicing medicine is actually organized and divided in real-world healthcare systems, and the latter question cannot be answered in isolation from normative accounts of what is considered to be the overall aim of medicine.

There is no direct access to the epistemological processes that support medical practice. Since one cannot gain knowledge of these processes by simply observing clinical work, one's understanding of them has to be based on conceptual analysis. Descriptively, one can try to account for what is actually going on in doctors' minds when they are practicing medicine. Normatively, one can discuss what should – ideally – be going on in their minds during this work. Importantly, these different perspectives must be kept apart to avoid the mistaken presumption that all doctors' medical practices coincide with ideal standards. (This assumption might be true but has to be explored empirically before being justified as an assumption.) Fortunately, much work has been carried out to elaborate accurate descriptions of processes of medical reasoning and normative ideals of medical practice. Central to many approaches are the concepts of "science" and "art" and elaborations on how these conceptualizations capture the essence of medical practice. The heading of this entry might invite one to think of these alternatives as apparent counterparts, but the general tendency in the literature is to acknowledge both categories as necessary parts of medical practice. Still, approaches may differ in how art and science in medical practice relate – or should relate – to each other.

Discussions of how to conceptualize medical practice on these terms are important for several reasons. The discussions have a bearing on how the role of being a physician is understood in general and more specifically on how doctors themselves interpret their role and effectuate it in practice. Conceptual clarification of medical practice is important for how society regulates and organizes the provision of

healthcare; this can only be done adequately insofar as it corresponds with a reasonable conceptualization of the ideal content of clinical work. Also, conceptual clarity of medical practice enables one to scrutinize and criticize the impacts of external organizational arrangements on real-world practice and, in turn, enable one to identify better organizational solutions. Furthermore, conceptual clarity is called for to delimit medical practice against other kinds of healthcare activities. It also enables decisions on relevant methods for developing and improving ongoing future practices. Conceptual clarification is also increasingly important for the medical profession itself in order to justify the privileged position it occupies in organized societies. It helps the professionals to be accountable to authorities and citizens and may support trust in that the medical profession handles its societal task of providing good medical care.

This entry is structured as follows: In the first section, a general epistemological framework for clarification of the fundamental conditions for the different approaches is presented. In the second section, meanings of "medicine as art" and "medicine as science" in relation to modern medical practice are presented. Next, versions of conceptual relations between art and science in medicine are described according to assumptions that the art and the science dimensions of medical practice are (a) independent of each other, (b) integrated with each other, or (c) the art dimension encompasses essentially different knowledge bases (including science) that supplement or complement each other. In the final section, philosophical and practical challenges involved in the art of balancing different knowledge bases in medical practice are described.

Epistemological Frame

Conceptualization of medicine as art and science gives associations to two basically different scientific traditions: science of humanities and science of nature. Since the Renaissance, humanistic disciplines have been concerned with disclosing and understanding the meaning of products created by humans through hermeneutical approaches, while science of nature traditionally has been taken to disclose and explain hidden facts about nature by experimental research. More recently, the social sciences have emerged as independent disciplines. Social sciences concern societies, human behavior, and social human relations and draw upon both methods of sciences of humanities and nature. These fundamentally different objects of scientific concerns imply different methods for reaching knowledge that is justified as scientifically valid. Depending on how the core tasks of medical practice are defined, seeking to establish knowledge within medical practice has the potential of calling on all of these traditions.

The Hippocratic Oath has for thousands of years served as a conceptual frame for defining the core tasks of practicing doctors. In the original version of the Oath translated into English, medical practice is basically referred to as "art." In the modern version of the Oath, the following statement is included: "I will remember that there is art to medicine as well as science, and that warmth,

sympathy, and understanding may outweigh the surgeon's knife or the chemist's drug" (Hippocratic Oath). In the old version, art refers to the whole practice of medicine considered as all-needed-capacities-included (Original Version Hippocratic Oath). However, it is described as art that can be taught to others. It is thus presumed that this art has some character of being *reproducible,* which is a criterion acknowledged for establishing knowledge within the science of nature rather than within knowledge production in the humanities. In the modern version of the Oath, art is basically related to the dimension of promoting understanding while science connects to actions involving the patient's body and that are based on knowledge that can be theoretically explained. Thus, historically, conceptualization of medicine as art within the medical profession's own constitutive declaration seems to differ with respect to its substantial meaning. In the following, medicine as art and science is basically understood according to modern medicine and existing tensions between conceptions of art and science.

Practicing medicine according to the ideal description of the modern Oath requires doctors to seek medically relevant knowledge along two different axes. They have to relate to nature in terms of seeking to identify and explain relevant features of the body in light of theoretical explanations. At the same time, they must seek to understand human products of meaning in terms of interpretations and explanations of patients' communication, reactions, and actions.

Most conspicuously, there is a fundamental epistemological gap between relating medicine to art – and by implication to the soft discipline of human science – on the one side and to science understood as the hard science of nature on the other (Snow 1998). Although the ideal description of modern medicine (the Oath) assumes that doctors base their knowledge on both, this gap allows for a different emphasis on these epistemologies and uncertainty with respect to how they should be taken to relate to each other. Empirically, emphasis on either dimension might depend on where in the medical process of identifying illness, treating or caring – and consequently, where in a specialized healthcare system – the practice to be described or assessed is found. The closer to the treatment of the bodily malfunction that medicine is practiced, the more the focus has to be on the explainable relations between intervention and expected outcome. When striving for identification of the medical issue or in providing nonphysical interhuman care, the more a focus on obtaining knowledge in terms of understanding is called for. However, one cannot conclude that in the first case medicine should be understood as science while in the latter case it is a matter of art. As the following sections will show, the science and art dimensions of practicing modern medicine have various interpretations, and the relation between them might be a bit less straightforward than suggested in the modern Hippocratic Oath.

Medicine as Science

In what sense is medical practice understood as science? One way to preliminarily clarify this dimension is to say that medical practitioners strive to be scientific and base their practice on scientific foundation (Sassower and Grodin 1987) or that

medical practice is scientific (Munson 1981). Another way of putting this is to say that medical practice requires the application of science (Munson 1981; Saunders 2000). In this sense, medicine is not taken to be a science itself; medicine is rather seen as an activity being based on translation of scientific knowledge into practice.

The question, then, is: what has been considered *relevant science* for medical practice? Again, descriptive and normative perspectives must be kept apart. For the following descriptive perspective on medicine as science, the focus is on what has been considered relevant science for medicine and thus has largely shaped the development of this practice. From a normative point of view, however, this historic perspective on medical science has been contested as representing an inadequate scope of scientific concerns (Malterud 1995).

Science Versus Nonscience

Scientific knowledge should be conceptually distinguished from nonscientific knowledge. Different criteria have been suggested (e.g., scientific knowledge must be empirically testable, explanatory, predictive (Sassower and Grodin 1987). However, as the history of science shows, criteria that qualify knowledge as science are not written in stone. So, from a normative point of view, some precaution is required when it comes to claiming absolute universal distinctions between science and nonscience in general and within disciplines, like medicine, in particular. From a general point of view, however, it might be uncontroversial to say that the aim to produce articulated and systematically justified knowledge is essential in science while it is not in nonscience.

In order to claim knowledge about a state of affairs, three criteria have been considered central since being discussed in Plato's dialogue *Theaetetus*: A proposition has to be true, one has to believe it, and one has to be able to justify it. Intuitively, these claims seem reasonable. From a philosophical point of view, however, the actual meanings of these criteria can all be scrutinized and discussed (What is truth? What is it to believe? What is it to justify?). This gives rise to various theories of science, which in turn base different methodological approaches to what is considered valid knowledge. Thus, in terms of science, modern medicine can descriptively be accounted for according to the dominating scientific view on how to reach valid knowledge in the field.

Medical science in modern times has unquestionably been dominated by biomedical science (Foss 1989). Thereby, the essence of medicine understood as science in this entry basically relates to biomedical knowledge and the criteria defining the scientific activity within this area. This approach can be traced back to Descartes and his dualistic account of the human mind as something distinct from the human body (Foss 1989). Hence, the human body and the mind were subjected to different fields of study. The concept of science applied on the body remained tightly connected with what can be derived from the laws of nature. The science of nature expanded into organic disciplines, like anatomy, biology, and physiology, and these approaches proved to be a helpful and effective means to understand and

develop tools to cure illness. Hence, science involved in medicine in modern times has basically been explained and practiced within a biomedical paradigm. (This applies to somatic medicine as the status of psychiatry as a science has been more contested.) At the same time, criteria defining scientific activity within this particular paradigm have also constrained the scope of what is considered valid knowledge on which to base medicine considered as a scientific medical practice.

Based on consensus, the medical community has broadly accepted the standards for evidence-based medicine (EBM). The ideal of EBM is to search for well-justified knowledge about efficacy and effectiveness of medical interventions based on experimental approaches within patient populations (Cochrane 1999). A basic principle of these clinical experiments is to strive for objectivity. For the results of the studies to be as objective as possible, one has to control for biases that might arise with respect to patient selection and outcome observations (and inherent interpretations). Therefore, participants are divided randomly into treatment and control groups. Also, the trials are double or triple blinded. In the first case, neither participants nor investigators know who receive the interventions being tested or who are in the control group. In the latter case, the groups of treatment assignments are also concealed for the team that analyzes the data. This approach is called a randomized controlled trial (RCT) and is referred to as the gold standard for medical research on clinical treatment; it tops the hierarchy of methodological approaches to knowledge ranked by the strength of evidence they produce. Scientific knowledge on which to base medicine correlates with research outcomes produced at the highest obtainable level of evidence. However, for pragmatic or ethical reasons, not all kinds of clinical research can be carried out as RCTs. Scientific knowledge can then be obtained by studies producing weaker evidence (e.g., controlled studies without randomization and observational, cohort, and case–control studies). At the bottom of the evidence hierarchy, and with very low scientific status, one finds expert opinion (e.g., expert reports of expert committees and experienced clinicians) (Essential Evidence Plus 2014).

The justification for the monopoly that the biomedical paradigm seemed to enjoy for a while has been contested (DiMatteo 1979; McWhinney 1986; Wulff 1986; Foss 1989; Malterud 1995; Saunders 2000). For instance, the recognition that medicine involves encounters between human subjects and not merely human bodies calls for a different kind of scientific approach than the one vindicated on the quantifiable conditions characterizing biomedical research alone (Malterud 1995). Human interaction is taken to be an essential part of medical practice. Thus, interpretive qualitative approaches developed within the tradition of humanities are called upon to inform medical practice. This acknowledgment also implies the need for including not only quantitative but also qualitative research approaches in the EBM framework.

From Science to Practice

Scientific results do not present themselves with a manual of how they should be used in medical practice. There is a gap between medical scientific research

(broadly construed) and medical practice that needs to be bridged. At least two fundamental challenges arise, and these are both connected to epistemic uncertainty. For one, how can practitioners be expected to gather all information and make use of the best available evidence in the myriad of published research? There is, of course, a practical side to this issue that has to do with time allocation. Philosophically, the core of this problem has to do with feasible expectations concerning individual assessments of strength of evidence. Proponents of basing medical practice on evidence have found a solution to the first challenge. Frameworks for systematically synthesizing knowledge and evidence assessment within medical research into guidelines have been developed (Woolf et al. 2012). The development of guidelines aims to reduce the messiness of the field of published research and provide healthcare personnel with tools for smoother and more feasible implementation of evidence in practice. It is worth noticing that the process of gathering and assessing knowledge cannot be considered as an objective and value-neutral activity in itself; clinical guidelines represent recommended policies for shaping practice and involve value trade-offs and judgment (Opel et al. 2013). Nevertheless, guidelines provide doctors with helpful manuals to handle the uncertainty related to the assessment of evidence. However, at the end of the day it is left to the doctors – and their clinical judgment – to choose whether to rely on these tools in their daily medical practice.

Proponents of EBM have been careful in pointing out that simply complying with evidence-based guidelines will not necessarily amount to adequate healthcare (Sackett et al. 1996). The evidence is based on population studies, and individual patients might present themselves with atypical conditions, comorbidity, and various personal preferences. Ultimately, this translational process has to lean upon an individual healthcare worker's judgment. It has to do so both to judge which recommending (synthesized) guideline is relevant in a particular case and then to assess whether this guideline actually covers the situation of the patient in question. Within this translational work bridging between general knowledge and particular cases, the art dimension of clinical work – or at least part of it – is located (Saunders 2000). This is independent of whether science is understood specifically according to an EBM framework or to a less specific knowledge concept. I will elaborate on this interpretation of medicine as art below. For now it is worth noting that art understood in the broad sense of representing a kind of translational judgment is also considered a crucial condition for adequately realizing science in successful evidence-based practice.

Medicine as Art: General

Attempts to grasp the content of medicine in terms of art can be a challenge. A reason for this is that medicine as art has, to a large extent, merely been negatively defined by pointing out what medicine as science does not cover. It has succinctly summed up how the art of medicine is often described by contrasts – being concerned with the particular rather than the general, practical knowledge rather

than theoretical; it includes the soul and is not merely focused on the body; it pays attention to mental processes and the unspecified effects of treatment (the doctor as a scientist tries to exclude the placebo effect; as an artist he/she makes use of it); it is concerned with values and not only facts; it concerns intuitions and affections and not merely rationality and knowledge; it provides courage and not merely medicine; it listens and not merely hears; it aims to restore rather than construes or generates; it integrates diagnosing and treatment (as science has separated) (Hofmann 2001).

The art of medicine is also accounted for independently of science. The art dimension has been described to encompass interpretations stemming from interhuman action (Malterud 1995); it can be taken to include tacit know-how based on experience (Malterud 1995), as well as any heuristics used to bring about practical conclusions under uncertainty (McDonald 1996). Moreover, it has been associated with the skill of bringing about a healthy outcome by technical interventions (i.e., according to the antique term *techne* (Hofmann 2003)) and the intellectual virtue *phronesis* (Gatens-Robinson 1986; Widdershoven-Heerding 1987; Davis 1997).

These ways of defining medicine as art can meaningfully be cataloged across two different accounts of how art comes into play in clinical care. This can happen, as already mentioned, within the work carried out by the judgment in translating general knowledge (broadly construed) into particular cases by practical reasoning and more specifically by involving and combining both nonmedical and biomedical knowledge in clinical care in order to bring about health.

Medicine as Art: Translating General Knowledge into Particular Cases

The process of translating theoretical knowledge into clinical practice cannot itself be labeled a scientific activity. From an epistemic point of view, particular clinical assessments are always subjected to some extent of uncertainty in knowing whether all relevant symptoms are uncovered, knowing which guideline – if any- to apply and in knowing how a particular body will react to treatment. In this translational process where the individual patient does not present him- or herself in any predefined manner, human reasoning cannot purposively work in a predefined automatic manner if the goal is to reach a certain health outcome. The literature describes heuristics available to the doctor's reasoning like rules of thumb and extrapolation (McDonald 1996). In sum, clinical judgment can encompass any ad hoc strategy or heuristic the individual doctor actual makes use of in order to bring the particular clinical situation of uncertainty to a practical conclusion. Thus, judgment can address issues concerning the patient's emotions; it can strategically produce health effects by comforting and not merely by medical theories (e.g., by actively alleviating fear and by downplaying the significance of observed anomalies); it can be based on values, experience-based intuitions, affections, and interpretative listening to what the patient – consciously or not – is communicating; it can encourage rather than provide medical fixes.

It is important not to confuse medicine as art with the idea that it represents a gift or some kind of esoteric knowledge. Strategies and heuristics can be learned through experiences (Malterud 1995). When they work automatically in experienced doctors, their clinical perceptions and conclusions may occur as being intuitive. This, however, does not necessarily make the emerging knowledge about the particular case tacit in the sense that it is impossible to articulate. Nevertheless, the translational reasoning process required to bridge between general knowledge and particular cases under uncertainty is not objectively controllable in the way scientific processes are required to be. The process is both context driven by features of the situation in question and personal in the way that trade-offs invoke a doctor's personal values. Thus, exercised clinical judgment does not follow any detectable systematic patterns that can be picked up, described, and reproduced in an objective scientific matter. In this sense, associations to uncontrollable, unforeseen reasoning processes supposed to be part of making art an aesthetic activity explain the labeling. But this alone does not promote any reasons to disregard the reasoning activity as something mysterious – it might simply represent another kind of rationality than the one presumed by the biomedical paradigm (Malterud 1995). The art of making clinical judgment along these lines can logically result in both failures and successes depending on the outcome. This is important to remember since one might be inclined to associate the art characteristic of medicine merely to clinical success stories.

Medicine as Art: Combining Contributions of Both Nonmedical and Biomedical Knowledge

As just pointed out, judgment is inevitably called for, even when translating science into practice. However, the interpretation of medicine as art is also distinguished from the interpretation of medicine as science in yet another way. In this version, the essence of medical practice considered as art is seen as being based on substantive contributions of knowledge coming from outside the biomedical domain. This conceptualization of medicine as art comes in at least two versions. On the one side, this conceptualization of medicine as art can be seen as referring to merely moral aspects of interhuman interaction (Saunders 2000). That is, the art elements refer to elements required for a morally justified medical practice where respectful treatment of the patient is emphasized.

In the other version, the elements involved in art are basically understood as everything involved in clinical encounters, including biomedical knowledge. Patients are fully recognized as human beings with lives and contextualized worries; they present themselves with both physical and mental attributes that must be taken into account in order for doctors to be able to respond with good and effective care. Malterud (1995) specified capacities that stem from interhuman encounters and that are considered crucial in order to adequately handle a patient's need together with biomedical knowledge. These capacities are not compatible with the construed rationality of the traditional biomedical perspective on medicine.

Malterud noted that these capacities should also be acknowledged for producing core knowledge for an ideal medical practice and as a consequence should be included in clinical epistemology.

Conceptual Relationships Between Medicine as Art and Medicine as Science

How is the conceptual relation between art and science in medicine described? Based on the literature, it seems apt to distinguish between three different versions of how art and science might relate conceptually in medical practice:

(a) *The art and the science dimensions of medical practice are independent of each other.*
The perspective reflected in the modern version of the Oath indicates some separateness between "art" and "science": Art is associated with promoting interrelational understanding while "science" is associated with skills required for technical interventions. Also, if art is basically considered as skillful treatment of patients merely in a moral sense, then art and science can be considered as distinct and independent elements in medical practice.

(b) *The art and science dimensions of medical practice are integrated with each other.*
When art captures the sense of translating general knowledge into particular cases, art is at the same time considered as an intrinsic part of practicing medicine on line with applying science. This would be the case independently of how successful the translation is according to any evaluative perspectives on medical performance. Analytically, any perspectives on medical practice that claim the inseparable nature of art and science, or claims that practical reasoning in principle can be broken down to such elements being inextricably bound together (like in conceptualizations of techne and phronesis), present the relationship between art and science as an matter of integration.

(c) *The art dimension encompasses essentially a different knowledge basis that supplements or complements the science dimension.*
The view that both biomedical and nonmedical constructions of knowledge are needed for *adequate* care and thus an adequate clinical epistemology presumes that knowledge emerging from interhuman encounters either supplements or complements scientific knowledge (i.e., biomedical science) in medical practice. In the first case, art will supplement biomedical knowledge if it provides nonbiomedical information that justifies nonstandardized interventions (e.g., a lack of a social network might justify a longer hospital stay or a patient's preference on intervention alternatives is taken into account). In the second case, art will complement biomedical knowledge if it is crucial in identifying what is at stake and what intervention is called for in order to achieve a beneficial outcome (e.g., when burdening social relations create physical symptoms). In both these cases different "types of knowledge construction are intimately interwoven in dialectic interplay" (Malterud 1995).

Synthesizing Approaches to the Role of Art and Science in Medicine

Exercising medicine as an art requires interpretive capacities which are called for in the translation of general scientific biomedical knowledge into particular cases; in acting as moral agents in encounters with patients; in establishing nonbiomedical knowledge with relevance for providing adequate care; and in the overall activity of combining all of these elements, including biomedical science, in the practice of medicine. This latter version of an all-things-considered art might very well equate with a broadly construed conception of practical, medical reasoning.

Concluding Remarks

Empirically, in medical practice all of the conceptually different relationships between art and science might very well be played out in a single clinical consultation. There are no logical bars to that. In that case, the conceptualization of art in the original version of the Hippocratic Oath as a comprehensive all-things-considered kind of art might in fact be closer to real-world medical practice than the more specified art concept presented in the modern version of the Oath.

In version (c) above, when the art dimension encompasses differently construed knowledge bases that either complement or supplement each other, careful balancing between the two categories is required. Structurally, evaluations of such a balancing process depend on what the aim of the medical practice is considered to be. This aim is rarely clearly stated in other than very general terms (like in legal regulations of provided healthcare). For instance, the aim of medical practice can be described as providing healthcare of high quality or healthcare according to the patient's best interest. In their clinical practice, doctors must both give this aim a substantive interpretation on a case-to-case basis and balance the concerns to emphasize accordingly. Uncertainty with respect to how balancing between different knowledge bases should be carried out within medical practice gives rise to various philosophical and practical issues. The list is not exhaustive but points to the fact that conceptualizations of medicine as art and science have relevance for the shaping of real-world healthcare provision and politics.

Epistemological Challenges

Malterud's account of a more adequate clinical epistemology requires supplementing/complementing qualitative research on premises of the tradition of the humanities. Still, the fundamental question concerning the normative limits of what to include/exclude in medical practice remains to be answered. Moreover, who decides on where to put the limits, i.e., what are relevant concerns and what are not?

Challenges in Organized Healthcare and Medical Education

The aim of medical practice may differ across different departments of a healthcare system, e.g., between primary and secondary healthcare. In primary care, diagnostic work may require doctors to take on a very broad perspective on what might be at stake before eventually referring the patient to the specialized care, i.e., for less broad approaches to specific domains of somatic or mental care. To correctly view the overall picture, GPs might be required to take more nonbiomedical information into account than their colleagues in secondary care specialities. Thus, adequate care might require unequal stress on the art dimension versus the scientific dimension depending on where in the system the healthcare is provided. How can this be handled by educational training?

Political Challenges

With a lack of clear instructions on how to balance the art and science dimensions of medical practice, unequal performance among clinicians is to be expected. For instance, clinicians might differ in what scope of nonmedical social concerns they find reasonable to include in their medical practice. This will, for one, lead to inequality in healthcare provided to patients with equal conditions and equal circumstances by different doctors. From certain positions on the social justice of healthcare, this will be unfair. Secondly, within public healthcare systems, doctors are given decisive discretionary power on distributional matters that ideally should be up to those with democratic powers to decide (Eriksen 2001). Should something be done to counter these "black holes" of democracy?

Challenges for the Medical Professionalism

The indeterminate nature of the overall goal of medical practice and its uncertain implications for how individual medical doctors should balance different knowledge bases in their practice also creates challenges for professional accountability. If there is no way to hold doctors accountable for the way they stress core elements in clinical epistemology relative to each other, there is nothing to support patients' trust in the professional's judgment in this regard.

Definition of Key Terms

Descriptive	Describes how something *is* without evaluating.
Normative	Describes how something *should be*/*should not be*, i.e., what would be ideal, good, right, fair, bad, wrong, unfair, etc.

Epistemology	Philosophical approaches concerned with the nature and scope of knowledge.
Biomedical paradigm	Set of broadly accepted premises structuring biomedical research.
Heuristics	Experience-based strategies for problem-solving and inquiries.

Summary Points

- Medical practice is often described in terms of "art" and "science."
- It is not obvious how these terms should be understood, neither how the relation between them should be described.
- Various meanings of medicine as science and medicine as art and the relation between them are presented.
- Medicine as science tends to refer to biomedical sciences, but an adequate clinical epistemology calls for supplementing/complementing this research with interpretive, qualitative research on phenomena occurring in interhuman encounters between doctor and patient.
- Exercising medicine as an art requires interpretive capacities, which is called for in the translation of general scientific biomedical knowledge into particular cases; in acting as moral agents in encounters with patients; in establishing nonbiomedical knowledge with relevance for providing adequate care; and in the overall activity of combining all of these elements, including biomedical science, in the practice of medicine.

References

Cochrane AL (1999) Effectiveness & efficiency: random reflections on health services. Royal Society of Medicine Press Limited, London

Davis FD (1997) Phronesis, clinical reasoning, and Pellegrino's philosophy of medicine. Theor Med Bioeth 18(1–2):173–195

DiMatteo MR (1979) A Social-psychological analysis of physician-patient rapport: toward a science of the art of medicine. J Soc Issues 35(1):12–33

Eriksen EO (2001) Demokratiets sorte hull - om spenningen mellom fag og politikk i veldferdstaten. Abstrakt Forlag, Oslo

Essential Evidence Plus. Available via http://www.essentialevidenceplus.com/product/ebm_loe.cfm?show=grade. Accessed 31 Aug 2014

Foss L (1989) The challenge to biomedicine: a foundations perspective. J Med Philos 14(2):165–191

Gatens-Robinson E (1986) Clinical judgment and the rationality of the human sciences. J Med Philos 11(2):167–178

Hippocratic Oath. Available via http://guides.library.jhu.edu/content.php?pid=23699&sid=190964. Accessed 31 Aug 2014

Hofmann B (2001) Legen som kroppstekniker. Tidsskr Nor Laegeforen 121(10):1266–1269

Hofmann B (2003) Medicine as techne – a perspective from antiquity. J Med Philos 28(4):403–425

Malterud K (1995) The legitimacy of clinical knowledge: towards a medical epistemology embracing the art of medicine. Theor Med Bioeth 16(2):183–198
McDonald CJ (1996) Medical heuristics: the silent adjudicators of clinical practice. Ann Intern Med 124(1_Part_1):56–62
McWhinney IR (1986) Are we on the brink of a major transformation of clinical method? CMAJ Can Med Assoc J 135(8):873
Munson R (1981) Why medicine cannot be a science. J Med Philos 6(2):183–208
Opel DJ, Taylor JA, Phillipi CA, Diekema DS (2013) The intersection of evidence and values in clinical guidelines: who decides what constitutes acceptable risk in the care of children? Hosp Pediatr 3(2):87–91
Original Version Hippocratic Oath. The Internet Classics Archive. Available via http://classics.mit.edu/Hippocrates/hippooath.html. Accessed 31 Aug 1014
Sackett D, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. BMJ 312(7023):71–72
Sassower R, Grodin MA (1987) Scientific uncertainty and medical responsibility. Theor Med Bioeth 8(2):221–234
Saunders J (2000) The practice of clinical medicine as an art and as a science. Med Humanit 26(1):18–22
Snow CP (1998) The two cultures. Cambridge University Press, Cambridge/New York
Widdershoven-Heerding I (1987) Medicine as a form of practical understanding. Theor Med Bioeth 8(2):179–185
Woolf S, Schunemann H, Eccles M, Grimshaw J, Shekelle P (2012) Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. Implement Sci 7(1):61
Wulff HR (1986) Rational diagnosis and treatment. J Med Philos 11(2):123–134

Biomedical Reductionist, Humanist, and Biopsychosocial Models in Medicine

47

S. Nassir Ghaemi

Contents

Abstract

In this chapter, three basic approaches to medicine are examined. The biological reductionistic model is criticized commonly but it is found to have important merits. The biopsychosocial (BPS) model is praised commonly but it is found to

S.N. Ghaemi (✉)
Department of Psychiatry, Tufts University School of Medicine, Boston, MA, USA

Mood Disorders Program, Tufts Medical Center, Boston, USA

Cambridge Health Alliance, Harvard Medical School, Boston, USA
e-mail: nghaemi@tuftsmedicalcenter.org

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_38

have many limitations. The BPS model is seen as representing eclecticism, which in turn represents a relativism about truth which is part of current cultural mores. Both perspectives are examined in the context of the history of medicine, where two basic tendencies are identified: the Galenic and Hippocratic approaches. The Galenic approach is based on biological speculation and is holistic and individualized to the patient. It held sway for most of recorded human history but caused much suffering through its false ideology. The Hippocratic approach is parroted but little understood: it is based on clinical observation, refusal to treat symptoms, and a commitment to identifying diseases. It is biologically reductionistic but humanistic. The evolution of modern medical breakthroughs, such as the antibiotic revolution, is seen as reflecting a rejection of Galenic models for Hippocratic ones. The BPS model is seen as a return to Galenic assumptions. A medical humanist model for the future, based on Hippocratic foundations and revised by awareness of the strengths and limitations of biological reductionism, is proposed.

Introduction

Three major approaches to a basic philosophy of medicine can be defined as follows:

Biomedical reductionist models take the view that all disease can be reduced to biological causes in the body; typically, treatments of those diseases are also biological in character, such as surgery or medications.

Humanist models take the view that illnesses sometimes may reflect diseases of the body, but sometimes they reflect problems between human beings of a non-biological nature, such as psychological or personal concerns. Treatment can be biological, but it often is not, entailing psychological or personal interventions such as counseling or self-help programs.

Biopsychosocial models take the view that all disease consists of an interaction between biological, psychological, and social causes; typically treatments of those diseases are also multiple, with biological (medications or surgery), psychological (counseling or self-help), and social (public health policy) interventions.

These three basic approaches to medicine have historical roots. This essay explores those historical roots and examines the concepts that evolved over time; it also will critique how those concepts have fared in medical practice.

Historical Background: Galenic Versus Hippocratic Approaches

The largest themes in the philosophy of medicine can be traced in historical sources to two basic lines of thinking: Galenic versus Hippocratic approaches. This claim is simplified necessarily, but simplification may help to clarify basic differences that

matter more to the general reader than smaller and more nuanced distinctions that matter more to specialists in the history of medicine. In this description, those nuances are being minimized not because they do not exist, but because they do not matter for the purposes of this chapter. There will be some combining of perspectives of different thinkers in one or the other of these basic philosophies of medicine, and the use of Hippocrates and Galen as the primary leaders in these schools of thought is based on their historical influence, with an awareness of the importance of other individuals before and after them in advancing or revising or clarifying their actual viewpoints. There also is awareness of the limits of historical documentation: As relates to Hippocratic writings, there is awareness of the fact that they mostly represent lecture notes, with lack of clarity at times of the exact identify of the lecturer; and they are incomplete. As relates to Galenic writings, they are extensive, but have passed through centuries of translation and revision. Basic sources exist for the overview below, to which the reader is directed (Jones 1931; Temkin 1973, 2002; Jouanna 1999; Porter 1999; Nutton 2001; McHugh 2006; Wootton 2007; Ghaemi 2008), as opposed to many repeated references after each statement made in the following summary.

With these scholarly caveats, the two basic trends of thinking in the history of medicine will be described.

Hippocratic Approaches

Hippocrates lived in the late fifth century BC, a contemporary of Socrates and Plato in a very exciting time of ancient Greek culture. His school of medicine, from the island of Kos, was set up as an alternative to the already prominent views of the school of Knidos. Both schools were influenced by the Egyptian tradition in medicine, centered in the city of Alexandria, as well as ancient sources from the Middle East (Mesopotamia). The school of Knidos, established around 700 BC, was the oldest and original source of the teaching of medicine in ancient Greece. Hippocrates was a successor to other teachers from Kos who revised and evolved ideas from Knidos, combined with new ideas influenced by the classic philosophers of Athens such as Socrates.

Most physicians pay lip service to Hippocrates and, if asked, will associate the man with the Hippocratic oath and the maxim "first do no harm." In fact, Hippocrates never said this; the phrase was invented in the mid-nineteenth century and falsely attributed to the Greek physician (Wootton 2007). Despite its historical falsehood, if we ask what this maxim means, most physicians, never having taken a history of medicine course, will tend to reply that it means that one should not harm the patient, first and foremost. Or perhaps they will translate it into standard risk-benefit analysis, where the benefits of treatment should outweigh the harm. This is all superficial. It would be like physicists saying that Newton sat under a tree and taught us that things fall. There was much more to Newton than the law of gravity; there is much more to Hippocrates than the Hippocratic oath.

There is a general misunderstanding of the term "Hippocratic," often associated with the ethical maxims of the Hippocratic oath, such as "first do no harm," later Latinized as *Primum non nocere*. A false claim, as noted, the full original quote was in the maxim of Epidemics I: "As to diseases, make a habit of two things – to help, or at least to do no harm" (Jones 1931). The Hippocratic tradition in medicine is thus identified simply with a conservative approach to treatment. While partly true, this popular simplification fails to capture the deeper genius of Hippocratic thinking, for its ethical maxims were not abstract opinions but rather grew out of its theory of disease.

The *basic* Hippocratic belief is that nature is the source of healing, and the job of the physician is to aid nature in the healing process. A non-Hippocratic view is that nature is the source of disease and that the physician (and surgeon) needs to fight nature to effect cure. Even in ancient Greece, physicians had many potions and pills to cure ailments; Hippocrates resisted that interventionistic medicine, and his treatment recommendations often involved diet, exercise, and wine – all designed to strengthen natural forces in recovery. If nature will cure, then the job of the physician is to hasten nature's work carefully and to avoid adding to the burden of illness.

Based on this philosophy of disease, the Hippocratics divided diseases into three types: *curable*, *incurable*, and *self-limiting. Curable* diseases require intervention, aimed at aiding the natural healing process. *Incurable* diseases generally were best left untreated, since treatments didn't improve illness and, due to side effects, would only add to suffering. *Self-limiting* diseases also didn't require treatment, since they improved spontaneously; by the time any benefits of treatment would occur, the illness would resolve by itself, again leaving only an unnecessary side-effect burden.

The Hippocratic approach emphasized clinical observation. Although the four humor theory was accepted, the presence or absence of disease was based on clinical sign and symptoms, not just speculation based on the four humors. This was another great breakthrough in Hippocratic thinking, an approach which would be submerged for over a millennium: base your thinking on observable clinical facts, not on your theories.

The basic Hippocratic insight into treatment was that the practice of medicine meant knowing when *not* to treat, not just assuming that one should always treat all symptoms or all patients.

Galenic Approaches

Galen came along in ancient Rome in the second century AD, over half a millennium after Hippocrates. But his influence would last almost two millennia, and in many ways, he can be seen as the most influential figure in the history of medicine. This is because Galen's approach surpassed Hippocrates' thinking and remained influential until the present day and, in fact, can be seen as still the most preeminent approach to medical practice today.

Although Galen wrote respectfully of Hippocrates and cloaked himself in the mantle of Hippocrates' reputation, he diverged remarkably from Hippocrates on the key topic of treatment. Galen loved drugs; Hippocrates distrusted them. They agreed

on the humoral theory, because there was no other conception of disease in ancient times. But they differed on everything else, with key differences on three matters: (a) the relative importance of clinical observation as opposed to biological theory, (b) the relative importance of natural history, and (c) the basic attitude of the physician toward treatment.

As noted above, the Hippocratic approach emphasized clinical observation as the most important aspect of medical knowledge. Galen's approach was quite different (Temkin 1973). He emphasized biological theory. Galen took the four humor theory very seriously, and inferred biological causes for the humors (the heart, for instance, heated the blood), and then would make treatment decisions on the basis of his biological theories (almost always some variant of bleeding). Galen was a theorizer, in contrast to Hippocrates who was an observer. The Hippocratic approach also emphasized the natural history of illness; nature was not seen as the enemy, and in fact it was noted that many illnesses resolved on their own, meaning naturally. Nature itself healed those illnesses. For Galen, nature was the enemy. Only the doctor healed. The doctor went to war against nature, whereas the Hippocratic doctor, seeing nature as a friend, was not going to war at all against anything. The attitude of the Galenic doctor then was aggressive: many treatments were given, and the patient suffered so the disease could be cured. This way of thinking is behind the old medical joke: The disease was cured, but unfortunately the patient died. The Hippocratic doctor would let a disease live, even for decades, as long as the patient survived. Hence the humorous comment from the nineteenth-century Hippocratic physician Oliver Wendell Holmes who remarked that the secret to longevity is to have a chronic illness and take good care of it (Holmes 1891).

An interesting aspect of Galenic medicine is that it was, in modern terms, "holistic" and "individualized to the patient." This is because the modern concept of disease was not present in ancient times. For Galen, there was only one "disease": an imbalance of the four humors. This imbalance could happen in an infinity of combinations in each individual person. So in a way, there was an infinity of diseases, one for each human being. Thus, he emphasized individualizing treatment to each patient: You get bleeding of this amount, this way, for this long; another person gets bleeding that amount, that way, for that long. You could add one of a thousand chemicals or herbs in varied combinations, and each person gets his or her own treatment – all based on pure speculation (Temkin 1973).

Galen wrote a great deal, and persuasively, and many of his writings survived. Hippocrates wrote little if at all; his ideas were recorded mostly as lecture notes, and most of them did not survive into the Middle Ages. Galen's ideas would take over the philosophy of medicine for a millennium.

The Evolution of Medical Thinking in the Middle Ages

From the fall of Rome until the Renaissance, a millennium of timespan represented the gradual explanation and transmittal of Galenic philosophy throughout the world. The main source of transmission was the Islamic world, and there the most

prominent medical thinker was Ibn Sina (Avicenna, eleventh century AD). (Much that follows can be sourced in a current text (Pormann and Savage-Smith 2007)). His main work, the Canon, which consisted of 14 volumes, would be taught as the primary medical text for over 500 years, not only in the Middle East but also throughout Christian Europe well into the seventeenth century. Although Ibn Sina has been held in high regard in the Islamic world, his basic philosophy of medicine is Galenic and not original to him. Ibn Sina expounded the four humor theory in detail and supported bleeding as a basic approach to treating many supposed illnesses; he also wrote extensively about over a thousand herbal or medicinal treatments, most of which was based on biological speculation rather than on clinical observation. He did make some clinical observations that stood the test of time, such as the sexual transmission of some diseases. By and large, though, the effect of Ibn Sina's work was to cement the influence of Galenic thinking throughout the Middle Ages.

In contrast to the powerful impact of Ibn Sina, some other Islamic physicians took up the basic philosophy of Hippocrates and tried to oppose Galenic orthodoxy. Among these, al-Razi (Rhazes, ninth century AD) and Ibn Rushd (Averroes, twelfth century AD) are prominent. Razi was the only medieval physician to directly attack Galen, in a book titled "Doubts about Galen." He directly attacked the legitimacy of the humoral theory and based his writing on disease purely on clinical observation, as in the Hippocratic tradition. In so doing, he described measles and smallpox. Like the Hippocratics, he emphasized the importance of *not* treating many patients and the need to avoid intervening in incurable diseases. Galen's influence was so profound that Razi couched his criticisms in the context of expressing great appreciation for the "master" Galen. Nonetheless, Razi was attacked widely for his temerity in criticizing the great Galen. Razi's long-term influence in the Islamic world may have been also limited due to his free-thinking and secular philosophy, as opposed to Ibn Sina's more mainstream religious orthodoxy.

Three centuries after Razi, Ibn Rushd would arise in Islamic Spain and have more influence. Ibn Rushd also was a liberal thinker, but this was not a problem in liberal Spain, where Islamic rulers were more tolerant than in the Persia of Razi or the Iraq of Ibn Sina. Ibn Rushd returned, like Razi, to Hippocrates and opposed the Galenic aggressiveness of Ibn Sina's philosophy. He did so through a Commentary on the Canon of Ibn Sina. Again, he emphasized clinical observation over the humoral theory, and he noted the importance of avoiding treatment in many patients. Like Razi, Ibn Rushd was more secular than Ibn Sina and more liberal in his political and social philosophy. His work was translated into Christian Europe from Islamic Spain, and thus he had longer-standing influence in the late Middle Ages.

The Seventeenth-Century Breakthrough

The Enlightenment did not occur out of the blue. It was an outgrowth of the impact in Christian Europe of more liberal elements of Islamic thought, combined with a rediscovery of ancient Greek and Roman texts. The key figure in modern

Enlightenment philosophy, its founder, was Descartes (sixteenth century), and he was influenced by the philosophical writings of thinkers like Ibn Rushd.

In medical thinking, the beginning of a change was heralded by the controversial Paracelsus (sixteenth century AD) of Switzerland. He picked up the mantle of Razi, without realizing it, when he became famous for rejecting Galen's authority and insisting on clinical observation. Like Razi, he was attacked widely and became quite embittered. A contemporary of Martin Luther, Paracelsus can be seen as a medical reformer who failed in his own time, but whose efforts represented the beginning of the end of the millennium-long reign of Galen. He rejected the humoral theory, arguing that many diseases were caused by outside causes, not by internal changes in the body. He strongly opposed bleeding, still by far the most prominent treatment in medicine in his era. He proposed many different minerals and medicines instead. He argued for cleanliness in managing wounds as opposed to the classic Galenic approach of frequent debridement.

Three major thinkers would soon follow in the seventeenth century, all contemporaries of each other, and together, they would succeed where Paracelsus had failed: they would kill Galen as the tyrant of medicine. These three men were Thomas Sydenham, William Harvey, and Giovanni Morgagni (Porter 1999; Wootton 2007).

Sydenham was an active support of the English revolutionaries and also was revolutionary in his medical thinking. Known as the English Hippocrates (Low 1999), he overtly returned to the basic Hippocratic idea of clinical observation and laid forth the general concept of clinical "syndromes" representing "signs and symptoms," a perspective now standard in clinical medicine, but quite radical in his age, since it rejected any reliance on the biological theorizing of Galen.

Where Sydenham rejected the whole concept of biological theorizing, Harvey replaced Galen's biological theorizing with a new approach of observation-based biological experiment. In so doing, Harvey famously identified the circulation of the blood. Sydenham opposed Harvey's biological theories, emphasizing the need for clinical observation, but both men together had the final impact of killing Galen, at least in the philosophy of medicine. If one was to be biological, then it would have to be based on experiment in the here and now, not based on what theories one might speculate about without testing them in experiment. Further, where biological knowledge was limited, clinical observation was seen as more scientifically valid than any other biological speculation.

Morgagni was Italian, unlike the other two English physicians. Morgagni became famous for his new concept of disease, described in his classic book "The seats and causes of disease." Morgagni made the radical claim, again now seen as commonplace, that diseases were caused by abnormalities in organs of the body. When combined with Sydenham's syndrome theory, it was a minor step to connect clinical syndromes identified through careful observation to abnormalities in organs of the body.

Finally, after more than a thousand years, Enlightenment medicine had a new explanation for diseases that could replace the four humor theory. It is important to note that this modern, radical, scientifically sound new approach was *not* holistic or individualized to the patient, as in Galen's theory. All patients with the same

abnormalities of the same organ would have the same clinical syndromes. Morgagni and Sydenham were identifying the group nature of disease and the fact that human beings were not individually different when it came to many diseases. Similarly, Harvey's discovery of the circulation of the blood was not different in one individual versus another: this expression of human biology was the same in everyone.

Statistics

Following these basic clinical and biological breakthroughs of the seventeenth century, the medical profession began to be influenced gradually by the birth of the new field of statistics in the eighteenth century. As exemplified by the French revolutionary thinker Pierre-Simon Laplace, the basic philosophy of modern statistics was founded on the view that mathematical methods would now be used to quantify, rather than ignore, error. Overlapping with Laplace in the late eighteenth century, and living well into the nineteenth century, the figure who founded medical statistics, and first applied those new mathematical approaches to medicine, was Pierre Charles Alexandre Louis, who introduced "the numerical method" to clinical medicine and applied it most famously to the first experimental study of the two-millennium-long practice of bleeding (Stigler 1986).

In 1828, he published the research study which has had, in my estimation, the most profound impact in the international medical practice. He examined the outcomes of bleeding for pneumonia in 77 patients in Paris (Yankauer 1996). Since it was considered unethical to withhold medical treatment that was thought to be effective, he did not try to see what happened in pneumonia with versus without bleeding; instead he compared outcomes in those who received more versus less bleeding. In some subjects bleeding was given early in the course of pneumonia, in others later. He observed that the longer one waited to bleed patients, the greater the percentage of those who survived.

The medical world was shocked. Even Louis did not try to make his claim straightforwardly. Like Razi a thousand years earlier, Louis was apologetic in his challenge to Galenic doctrine. In his paper, he stated that his study did not mean that bleeding was never effective, but rather that it might be delayed and used in later stages of pneumonia. For his cautious interpretation, Louis was pilloried in the letters to the editor of the French medical journal in which his study was published. Many physicians were enraged that Louis would place individual patients into a numerical test and that he would treat human beings as numbers. His methods were seen as unhumanistic and illiberal.

These critics were faced with the reality of the numbers, though. Louis attracted young physicians from around the world, bred in democratic nations, like the great Oliver Wendell Holmes, who, from his base as a Harvard professor of medicine, would go on to be a leading figure in American medicine. Holmes was a democrat and a humanist, and he saw that Louis' medical statistics, like science in general, was consistent with a humanistic outlook (Holmes 1891). If science could prove that bleeding was not effective, then so much the worse for bleeding.

Louis' work was the beginning of the end of bleeding. In about a decade, the importation of leeches to Paris declined from the millions to the thousands. It would still take half a century, until around the turn of the twentieth century, until the practice of bleeding ended.

What had begun back before the fifth century BC, when Hippocrates and his students roamed the island of Cos, would last until about 1900, when steam engines and electricity and large industrial factories existed. Galen's favorite treatment finally was let go by the world's physicians after more than 2,000 years.

The Impact of Pathology

Along with the development of medical statistics, the nineteenth century saw another new method in medicine that would prove to be the final step in leaving ancient methods behind and beginning the era of modern science: postmortem pathology. Morgagni's theory that diseases were due to abnormalities of organs was finally proven true when postmortem pathology evolved as a common medical practice in the late nineteenth century. One of the leaders in the pathology movement was the Canadian physician William Osler (Bliss 1999). Osler conducted tens of thousands of postmortem autopsies and carefully recorded the pathological findings of his patients. At the same time, he had examined those patients while living, applying Sydenham's method of meticulous clinical evaluation of signs and symptoms. He went back and forth between the two approaches: clinical syndrome observations were confirmed or rejected based on pathological findings at autopsy, and vice versa, autopsy findings were informed by prior clinical syndrome observations. Gradually, Osler and his colleagues and disciples evolved the "clinicopathological" method which has become the standard approach of modern scientific medicine. The development of the microscope in prior centuries, and its increasing use in histology in the nineteenth century, also facilitated the effectiveness of the clinicopathological method, as gross evaluation of organs was augmented by histological study of organ tissues.

Osler augmented his clinicopathological skills with an appreciation for the medical statistics of Louis, and the field of clinical research, with statistical evaluation of clinical and pathological findings, had begun. Osler's extensive knowledge of the history of medicine was added to the mix as he overtly taught a return to the basic Hippocratic philosophy of medicine and a rejection of the Galenic thinking that had led to so much bleeding for so long.

His ideas had a major influence on twentieth-century medicine through his professional role as the first chairman of the new department of medicine of the first modern medical school in the United States at Johns Hopkins University in Baltimore and later through his final position as Regius Professor of Medicine at Oxford. As importantly, he wrote a modern Canon of Medicine, as influential for a brief time as Ibn Sina's work had been for much longer: *The Principles and Practice of Medicine* (Osler 1912). Osler's textbook would be the last prominent textbook of medicine written by a single thinker. In it, he captured in detail the most careful

clinical syndrome observations and the most recent pathological evidence. He provided cautious, Hippocratic treatment recommendations, which were so sparse that he was accused of "therapeutic nihilism." From the late 1890s when he published his text, past his death in 1920, his textbook was the preeminent source for medical practice in the Anglo-American world. It would remain so into the 1940s, until the introduction of antibiotics, the next great revolution in medical practice.

A final impact on Osler was the influence of his literary and religious background in the Victorian late nineteenth-century period. He brought the humanistic ideals of the Enlightenment into the very core of modern medicine and showed how it could be applied along with rigorous scientific clinical and pathological methods (Osler 1932). Osler was both a biological reductionist and a humanist.

The Revolution of Randomization

Oslerian medicine focused on diagnosis, not treatment, because it was scientifically honest about the fact that there were not many effective treatments. Louis' methods had proven this fact. In the 1920s, Louis' numerical method was taken forward hugely through the development of the concept of "randomization" by Ronald Fisher, a statistician and geneticist. Fisher applied his idea initially in agriculture, but it didn't take long before it was picked up in medicine and applied for the first time by the British medical epidemiologist, A. Bradford Hill, in 1948. The first randomized clinical trial was conducted to prove the efficacy of an early antibiotic, streptomycin, in miliary tuberculosis (Hill 1971).

Antibiotics had been discovered before the Second World War, but began to be used widely after the war, into the 1950s. The work of Fisher and Hill was central in proving their efficacy, and the new statistical methods began to be used for many different medications in medicine.

None proved so effective and radical as the antibiotic class. These medications transformed the practice of medicine and saved millions of lives that for millennia were lost to infections.

We have lost historical memory of that period, what was only known in the memory of lost generations, of the grandparents and great-grandparents of the readers of this chapter. Readers can find a sense of the amazing impact of antibiotics in reading works of physicians from the mid-twentieth century, like Lewis Thomas in his classic book *The Youngest Science* (Thomas 1995). Thomas describes being a medical intern in the most prestigious Boston hospitals in the 1930s, just before the introduction of antibiotics. He describes how a young child would come to the emergency room with an infected cut of the hand; sepsis would ensue and the child would die. A decade later, no child would die of that cause, easily cured with penicillin. Tuberculosis was cured with streptomycin. Syphilis, which in its neurological effects caused the equivalent of schizophrenia in 1 % of the world's population, was cured (Shorter 1997). Addison's disease, which produced an expected lifespan of 30 years due to uncontrolled infections, became a chronic condition that

could be managed for a lifetime with steroids and penicillin. Thus could a young man like John F. Kennedy survive to become president in his 40s. Diabetes also was transformed, with the discovery of insulin, from a death sentence to a chronic manageable illness. Heart attacks and strokes, so commonly the cause of sudden death, became less common with the treatment of hypertension and, in later years, the reduction of cholesterol.

Clinical medicine was finally transformed by having biological treatments that actually cured diseases, unlike the thousands of treatments of Galen and Ibn Sina and their disciples. The Hippocratic method of careful clinical evaluation, modified by Sydenham and Harvey and Morgagni and Osler, could now be tested through the new statistical methods to identify truly effective treatments for real diseases. Speculation and theory were put in their rightful place of generating hypotheses, to be tested and possibly rejected based on experiment and clinical observation, rather than the reverse.

This modern neo-Hippocratic medicine proved immensely effective, and, for the first time in human history, physicians could save lives based on true knowledge, rather than guessing or, worse, harming based on false beliefs.

The Biopsychosocial Model: Late Twentieth-Century Dissatisfaction

It would be inconsistent with human history for this story to end here. By the late twentieth century, a new dissatisfaction arose with the mid-twentieth-century scientific transformation of modern medicine. We continue to live today with a backlash against these medical successes.

Along with the biological advances of antibiotics and other new treatments like steroids and insulin, there was a parallel development in medicine: Sigmund Freud's psychoanalysis. Freud discovered that certain apparent neurological problems did not have a biological cause in the body, but rather were produced in the body through purely psychological influences. He began his work with "hysteria," which related often to physical manifestations for which neurologists like Freud could find no brain basis. These included seizures and bodily paralysis. Some had found that hypnosis could produce or improve such hysterical symptoms, and Freud showed that a "talking cure" could produce the same benefit. Simply talking in what is now called psychotherapy had medical benefits.

Freud's work was taken in many directions, but in clinical medicine the main impact was in the field of "psychosomatic medicine" (Shorter 1997). Many internal medicine physicians became interested in how the mind might affect the body, especially for those medical conditions for whom physical causes were difficult to identify. Those physicians often obtained formal training in Freudian psychoanalysis and then would become specialists in psychosomatic medicine. One such person was the gastrointestinal specialist George Engel, who would later found the biopsychosocial model (Engel 1977). Engel specialized in ulcerative colitis and peptic ulcer disease. Both conditions were thought to be of largely psychological

origin for much of Engel's career, in the early to mid-twentieth century. By his later years, though, new work was discovering genetic causes for ulcerative colitis and infectious causes for peptic ulcers. The medical profession also began to become more interested in the biological mechanisms of inflammation that underlie those conditions. In other fields, an increasing use of testing of various kinds – x-rays, blood tests, and other machine-based measurements – was leading away from the clinical bedside observation that had characterized most medical diagnosis in prior centuries.

By the 1970s, toward the end of his career, Engel became disturbed by these trends (Ghaemi 2010) and wrote his classic paper contrasting a biopsychosocial model against what he called the bioreductionist model of medicine. Engel's critique struck a cultural nerve and became the standard approach taught in American medical schools in the end of the twentieth century and into the present time.

Definitions of the Biopsychosocial Model

The proposal made by Engel is that there are two basic models of medicine, the biological reductionist and the biopsychosocial (BPS). The former only looks at biological causes of disease; the latter argues that most (or all) disease is multifactorial, with psychological and social causes, not just biological ones. Engel came to this conclusion on the basis of his interest in psychosomatic medicine and functional bowel disorders and peptic ulcer disease. At the time he made this claim, it had been widely held that such conditions had important psychological causes.

It is important to keep in mind that by psychological causes, Engel was thinking in the Freudian/psychoanalytic paradigm. He meant unconscious mental states, often dating back to childhood, that produced physical symptoms (Ghaemi 2010).

It is also important to note that he hardly expanded on social causes in any of his writing, and he certainly was not thinking about the profession of social work in his theory, even though the BPS model has now become the mantra of the social work profession.

A typical case which Engel would present in lectures and writings is of a man with heart problems, who goes to the emergency room with chest pain (Engel 1980). There he is met by an inexperienced medical intern who does a terrible job with a needle trying to get an arterial blood gas in the patient's wrist. The patient becomes very anxious and is in pain and then has a ventricular arrhythmia. A code is called and he is given intravenous medications and electrical defibrillation such that he is resuscitated. Engel's point is that one could look at the whole story from the viewpoint of the physical and biological interventions, but the key to the story is the psychological impact of the intern's painful incompetence, which triggered the arrhythmia.

By the late 1970s, when Engel began speaking about these ideas, evidence began to accumulate both for and against his model. On the one hand, the new field of social epidemiology was identifying important social and psychological factors that predisposed to some illnesses, such as diabetes and heart diseases. On the other hand, the illnesses closest to Engel's interest, like peptic ulcer disease, were found not to be biopsychosocial at all. An infectious agent, H. Pylori, was found to be more important than the most complex psychosocial speculations about ulcers. Irritable

bowel syndrome was found to be importantly genetic, with a complex immunological pathophysiology; psychological and social factors were not confirmed as being central to those bowel conditions (Ghaemi 2010).

Evolution of the Biopsychosocial Model

Three groups latched onto the BPS model and continue to hold to it very strongly today: (1) primary care physicians, (2) psychiatrists, and (3) social workers.

Primary care physicians see many patients with many symptom complaints that often do not, after medical workup, have a physical basis in the body. Hence, the interest in psychological and social factors in the lives of patients that may bring them to the doctor (Weiss 1980).

Psychiatrists and social workers do counseling with many patients for life problems, like unhappiness after divorce or grief after the death of a loved one, that may not have any relation to a physical problem in the brain causing psychological symptoms (like manic-depressive disease or schizophrenia). At the time of Engel's cri de coeur, psychiatry was moving to more use of medications; this biological approach was met with much resistance by the psychoanalytic core of the profession. The BPS model has become the mantra which many psychiatrists use to push back against using medications or thinking about psychiatric symptoms as related to diseases of the brain or body (Gabbard and Kay 2001).

The social work profession sees its raison d'etre as tied to the BPS model; the word "social" in the phrase gives the social work profession a claim to relevance in medical care (Kerson 1987). This need not be the case, since social work as a profession predates the BPS model by over half a century. Further, as noted, Engel more or less ignored the social aspect of the BPS model.

Claims and Critiques of the BPS Model

The BPS model is attractive for professional reasons as given above, but its basic claims do not stand up to conceptual or scientific scrutiny. At times, it is hard to even clearly understand what the basic BPS theory is. This suggests that it is more a slogan than anything else, a label used by those who wish to maintain a humanistic attitude toward patients. But the latter wish, humanism, is not inherently in conflict with biological reductionism or inherently consistent with the BPS ideology. (The following critiques have been previously published by me in more detail at book length (Ghaemi 2010).)

Regarding the veracity of BPS claims, it is false to claim that *all* medical diseases have psychological and social factors in their etiologies. Many are purely biological; one can cite many purely genetic conditions, like trisomy 21 or phenylketonuria. Those conditions are 100 % genetic and biological in etiology; there are no psychological or social factors in their causation. Biological reductionism is correct in those diseases.

So the BPS claim would have to be weakened: One could claim that *most* illnesses are multifactorial in etiology. This also is a false claim. There are hundreds, if not thousands, of purely genetic or purely biological diseases of various kinds (infections, cancers, autoimmune diseases). Many of them may be rare, but they abound.

The BPS claim can be weakened further: One could claim that most chronic illnesses are multifactorial in etiology, like coronary vascular disease and diabetes and depressive conditions. This claim would be more defensible, but then many acute medical diseases would have to be excluded from the BPS model. Further, the influence of psychological and social factors in these illnesses still is exerted by a biological mechanism. For instance, social isolation is associated with diabetes; the mechanism may involve overeating and lack of activity leading to insulin resistance in the pancreas. Social factors are in the causal pathway but they always have to exert their effects through the biological mechanism, insulin resistance. The biological component is essential to the disease; the social components are not. One could get the insulin resistance by nonsocial means, such as genetic transmission of insulin resistance or a medication side effect. So even on this claim, though one can claim multifactorial etiology, not all the factors are equal in importance, and the biological factor still seems most important.

The BPS claim might be restated so as to move away from etiology altogether. The claim could be that all or most illnesses, whatever their etiologies, are affected by psychological and social factors. If you have a purely genetic disease, the course of your illness will still be impacted by your psychological state or your social condition. This is the most defensible version of the BPS approach to illness, but it cedes a great deal. It accepts biological reductionism in relation to cause for many illnesses, and it makes claims only regarding pathogenesis, or amelioration of the course of the biological illness. Though this claim is more defensible, it is rarely made in this limited way. Usually BPS advocates make etiological claims, as Engel did. The above critiques would then apply.

The Essence of the BPS Model

The above critiques bring out the vagueness and limitations of the BPS model as put forward by many of its advocates. In my analysis of this literature (Ghaemi 2010), I have come to the conclusion that there is one essence and core to the BPS model that is more central than any of the claims above. *The essence of the BPS model is eclecticism*. By eclecticism, I mean that the BPS model wishes to avoid any definitive assertion of causation or importance of any one of the three factors. It not only wishes to avoid biological reductionism, it wishes to avoid psychological and social reductionism. It wishes not to make definitive claims of any kind, except the claim that one can never be definitive.

This is what attracts clinicians, especially psychiatrists and social workers and some primary care physicians. One can always use the BPS model to criticize anyone else's claims about biological or psychological or social causation, and then one can defend whatever claims one wishes to defend with the same model. Essentially, it

allows clinicians to do whatever they want, under the cover of being "holistic" and biopsychosocial. The slogan of "individualizing care to the patient" is then brought out to further defend the clinician's wish to be free. All these attitudes are tied into the wish to be humanistic, to treat the patient as an individual human being who is unique and has feelings and a certain social context. All this can be true, but none of it proves or disproves a biological reductionist etiology to any putative disease.

Readers will note from the interpretation of the history of Galenic versus Hippocratic approaches to medicine that the BPS model is revisiting some very old territory. Galenic medicine, for two millennia, was holistic and individualized to the patient – but it was far from humanistic. Bleeding and purging for millennia only tortured many poor human beings who suffered from the biological ignorance of their physicians.

The BPS model can be quite anti-humanistic and dehumanizing (Ghaemi 2010). If you have purely biological disease, which has a biological cure, such as *H. Pylori*-related peptic ulcer disease, your clinician is harming your humanity by not diagnosing or treating it. He or she might be the most pleasant and humane person in the world, with the best of intentions, providing the best of counseling, and attuned to your personal life quite well. But your ulcer pain will persist until you get the right antibiotic.

In psychiatry, the same problem exists with dismissive BPS (Mojtabai and Olfson 2010) attitudes about the concept of disease. For instance, it is known that bipolar illness is almost completely genetic in etiology. And there is a treatment that essentially cures: lithium can produce complete remission in about one-third of persons with that condition. Yet it is only prescribed to about 10 % of diagnosed patients in the United States, and most patients receive instead symptomatic treatments with medications for depressive or anxiety or sleep symptoms, along with counseling. The latter approach is biopsychosocial, but it is ineffective, and patients continue to suffer needlessly while a much more effective cure is ignored because it entails thinking about treating an underlying biological disease, as opposed to seeing biological factors as limited in importance.

Postmodernism and Medical "Narrative"

Another aspect of the eclecticism of the BPS model is that it merges with an extreme skepticism about scientific truth. This is reflected in postmodernist attitudes in the past half century in Western culture, inaugurated by the 1960s counterculture (Ghaemi 2013). In this thinking, science is just another "narrative," not any more right or wrong than other ideologies, whether literary or political or cultural. There is no truth, only claims to truth which really reflect social and cultural interest groups. Science is what scientists say is true, but that is no more true than what theologians say is true. Michel Foucault and other postmodernists made trenchant critiques of the history of medicine on this basis (Foucault 2001). Their views have become very popular in modern culture. Even among those who have not read them, these postmodernist views have become part of the Weltanschauung of the current age. In medicine and psychiatry, they get played out through an attraction to extreme

eclecticism, with a dismissive attitude toward "the" truth, and a consequent adherence to an eclectic interpretation of the BPS theory.

Some also now speak of the importance of "narrative" in understanding medical illnesses. The professions of history of science and medical anthropology are now thoroughly postmodernist. Any history of medicine which seeks to claim eternal truth for any fact is dismissed *tout court*. The concept of progress is *verboten*. Nothing is allowed but the relativistic and at times nihilistic attitudes of those who wish to dissolve all medicine and science into a soup of eclecticism.

Even those who see themselves as scientifically oriented persons often succumb to the unconscious influences of the postmodernist spirit of our age. In medicine, the genetic revolution has led to the current mania for "personalized" medicine. This idea promises to succeed in achieving the Galenic goal: all illness will finally be individualized on biological grounds, just as Galen always wished. The genetic work is an empirical claim, and we can wait to judge it on empirical grounds. As epigenetics becomes incorporated, and the influence of environment on genes is better understood, it may provide another means to support some BPS intuitions. We will have to wait to judge these claims on empirical grounds. I would only comment that the history of medicine argues for caution against these claims, as they were strongly believed for 2,000 years and caused terrible harm. Future advocates of these ideas should keep in mind the errors of the past.

A Medical Humanism for the Future

While we await the playing out of the genomic revolution, I would suggest that if it fails to achieve its most grandiose goals, which it may, then we should have a more realistic alternative. I would like to propose here another perspective, not one that takes sides on these debates but one that seeks to learn from them (Ghaemi 2010).

I would suggest that future physicians and clinicians would do well to realize that the BPS model, as stated eclectically and forcefully, is false. So too is any strong formulation of biological reductionism. There are many diseases that can be understood reductionistically as being biological in etiology. There are also important chronic medical illnesses that have key social and psychological aspects, both to etiology and to course of illness. We all agree that humanism is important and that whatever the etiology of a person's disease or absence of disease, it is the person who either has the disease or does not. We still have to deal with a human being, with all her psychological and social and individual traits.

It is true that many patients who see primary care doctors and psychiatrists and social workers do not have any medical disease at all, but some do, and some even have purely biological diseases. Some persons even have purely social problems, or purely psychological problems, with no relevant biological component at all. How can we tell which is which?

I suggest that the main culprit here is eclecticism and underneath it the postmodernist relativism that is our current unconscious philosophy of life. We need to become conscious of the importance of the scientific attitude, in the straight

Enlightenment tradition, the notion that there are truths, that some facts are better proven than others. We need to be willing to commit to biological reductionism as true when the scientific proof is present for it and similarly for social and psychological reductionism and similarly for some cases where the scientific proof is antireductionistic and supports the importance of multiple factors.

In other words, we should not prejudge these matters, but let scientific methods tell us what to believe. We should believe in science and nothing else. But we should really believe, more so than in other ideologies, including the comfortable eclecticisms of the departments of literature and medical anthropology.

No matter what our science tells us, we also should be committed to the reality that dates back to Hippocrates, the fact that the individual patient as a human being must be understood as a human being. In this feature we are all individual, although even here, as the great psychoanalyst Harry Stack Sullivan once said, we are all much more alike than otherwise.

Definition of Key Terms

Biomedical reductionist models	Take the view that all disease can be reduced to biological causes in the body; typically, treatments of those diseases are also biological in character, such as surgery or medications.
Humanist models	Take the view that illnesses sometimes may reflect diseases of the body, but sometimes they reflect problems between human beings of a non-biological nature, such as psychological or personal concerns. Treatment can be biological, but it often is not, entailing psychological or personal interventions such as counseling or self-help programs.
Biopsychosocial models	Take the view that all disease consists of an interaction between biological, psychological, and social causes; typically treatments of those diseases are also multiple, with biological (medications or surgery), psychological (counseling or self-help), and social (public health policy) interventions.
Hippocratic approaches	To medicine emphasize clinical observation and not treating symptoms. Rather only some symptoms which are caused by diseases should be treated. Hippocratic approaches are disease oriented, not symptom oriented, and clinical observation based not biological speculation based.
Galenic approaches	To medicine emphasize biological theory and treating symptoms. Disease concepts are neglected, and all illnesses are seen as individualized to each person based on the specific combination of the four humors in that

	person. Galenic approaches are symptom oriented, not disease oriented, and biological speculation based not clinical observation based.
Medical humanism	Medical humanism is a biological reductionist approach to disease that recognizes the importance of also understanding each human being as a human being, not only based on psychological and social aspects but also based on existential aspects of the human condition.

Summary Points

- The history of medicine involves two basic currents of thinking, Galenic and Hippocratic, the former being biologically speculative and holistic but unhumanistic, the latter being clinically observational and disease based but more humanistic.
- Biological reductionism, though commonly criticized, is valid for many diseases.
- The biopsychosocial model, though commonly praised, is false for many diseases.
- The biopsychosocial model represents, in essence, eclecticism.
- Medical humanism is not incompatible with biological reductionism.

References

Bliss M (1999) William Osler: a life in medicine. Oxford University Press, New York

Engel GL (1977) The need for a new medical model: a challenge for biomedicine. Science 196 (80):129–136

Engel GL (1980) The clinical application of the biopsychosocial model. Am J Psychiatry 137:535–544

Foucault M (2001) Madness and civilization: a history of insanity in the age of reason. Routledge, London

Gabbard GO, Kay J (2001) The fate of integrated treatment: whatever happened to the biopsychosocial psychiatrist? Am J Psychiatry 158:1956–1963

Ghaemi SN (2008) Toward a Hippocratic psychopharmacology. Can J Psychiatry 53:189–196

Ghaemi SN (2010) The rise and fall of the biopsychosocial model: reconciling art and science in psychiatry. Johns Hopkins Press, Baltimore

Ghaemi N (2013) On depression: diagnosis, drugs and despair in the modern world. Johns Hopkins Press, Baltimore

Hill AB (1971) Principles of medical statistics, 9th edn. Oxford University Press, New York

Holmes OW (1891) Currents and counter-currents in medical science. Medical essays 1842−1882. Houghton-Mifflin, Boston

Jones WHS (1931) The works of Hippocrates

Jouanna J (1999) Hippocrates. Johns Hopkins University Press, Baltimore

Kerson TS (1987) Clinical social work in health care: new biopsychosocial approaches. Health Soc Work 12:77–78

Low G (1999) Thomas Sydenham: the English Hippocrates. Aust N Z J Surg 69:258–262

McHugh PR (2006) Hippocrates a la mode. In: The mind has mountains. Johns Hopkins Press, Baltimore
Mojtabai R, Olfson M (2010) National trends in psychotropic medication polypharmacy in office-based psychiatry. Arch Gen Psychiatry 67:26–36
Nutton V (2001) Hippocrates of Cos. In: Encyclopedia of life sciences. John Wiley and Sons. http://dx.doi.org/10.1038/npg.els.0002519
Osler W (1912) The principles and practice of medicine. D. Appleton and Company, New York
Osler W (1932) Aequanimitas, 3rd edn. The Blakiston Company, Philadelphia
Pormann P, Savage-Smith E (2007) Medieval Islamic medicine. Georgetown University Press, Washington, DC
Porter R (1999) The greatest benefit to mankind: a medical history of humanity. Norton, New York
Shorter E (1997) A history of psychiatry. Wiley, New York
Stigler S (1986) A history of statistics. Harvard University Press, Cambridge, MA
Temkin O (1973) Galenism: rise and decline of a medical philosophy. Cornell University Press, Ithaca
Temkin O (2002) On second thought and other essays in the history of medicine and science. Johns Hopkins University Press, Baltimore
Thomas L (1995) The youngest science. Penguin, New York
Weiss RJ (1980) The biopsychosocial model and primary care. Psychosom Med 42:123–130
Wootton D (2007) Bad medicine: doctors doing harm since Hippocrates. Oxford University Press, Oxford
Yankauer A (1996) Pierre Charles-Alexandre Louis: the impact of a clinical trial. Pharos Alpha Omega Alpha Honor Med Soc 59:15–19

Medical Theory and Its Notions of Definition and Explanation

48

Peter Hucklenbroich

Contents

Abstract

This chapter elucidates several special features of the usage of the notions *definition* and *explanation* in medicine and medical theory. As these special features are intimately connected to the key concept of d*isease entity*, the first section gives a short reconstruction of this concept. The second section presents three methods of defining disease entities, supplemented by a fourth, logically unsound method found in many medical textbooks. The third section shows that there are two senses of explaining symptoms and pathological conditions by referring to disease entities, i.e., a part-whole kind of explanation and a causal one. The relationship between *explanation* and *diagnosis* of diseases is analyzed by comparing their logical structure. In the last section, the very special kind of explanation found exclusively in medicine, viz., explaining *why* some condition is a disease or is pathological, is clarified by elucidating the concept of *pathologicity* and its criteria.

P. Hucklenbroich (✉)
Institut für Ethik, Geschichte und Theorie der Medizin, Westfälische Wilhelms-Universität Münster, Münster, Germany
e-mail: hucklen@ukmuenster.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_44

Introduction

Definitions and explanations are important conceptual tools in all scientific disciplines that contain full-fledged, well-articulated theories – hence, also in medicine and medical science. However, usage of these concepts in medicine exhibits several special features that are consequences of the conceptual structure of medical theory of disease (general pathology) and the logic and methodology of clinical reasoning. In particular, these special features play an essential role in the conceptual clarification and explication of the notions of *disease* and *diagnosis* and have a decisive impact on the handling of definitions and explanations. Hence, the following discussion consists of four sections: first, some basic concepts and principles of general pathology and theory of disease are outlined in order to improve understanding of the following sections (section "Some Concepts and Principles of General Medical Pathology"). Subsequently, the forms and characteristics of definitions of disease entities are presented and analyzed (section "The Definition of Disease Entities"). After this, the relationship between the concepts of *diagnosis* and *explanation* is elucidated (section "Explanation and Diagnosis"). Finally, the particular linguistic usage of explaining *why a particular medical condition is abnormal or pathological*, respectively, *why it is a disease*, is conceptually analyzed and reconstructed (section "Explanation of Pathologicity").

Some Concepts and Principles of General Medical Pathology

The expression *general pathology* (German: *Krankheitslehre*) is taken to refer here to the whole body of medical theories that refer to states and processes of health and disease as occurrences within an individual human life span. General pathology in this medical sense incorporates anatomical, functional, behavioral, mental, and subjectively experienced states and processes. Knowledge of all these conditions is spread over the whole body of medical theories and their representations in handbooks and textbooks. There are, however, very few attempts at giving a unified, systematic, and comprehensible overview and account of general pathology in this sense; hence, it has to be obtained from multiple sources (e.g., Büchner et al. 1969; Sandritter and Beneke 1974; Riede 2004; Siegenthaler 2007; Bickley and Szilagyi 2013; Hammer and McPhee 2014).

Even in everyday, prescientific discourse, some states and processes of life are characterized as pathological ones. Scientific pathology systematizes this knowledge and constructs a systematic nosology, entailing that all pathological conditions are parts and manifestations of particular diseases or, more precisely, of particular kinds of diseases called *disease entities*: every case of falling ill, every illness or sickness, and every disorder or malady are, in the view of medical science, a case of a disease entity. Disease entities comprise not only the "diseases" of lay understanding, such as infectious or metabolic disorders, but also congenital or acquired disfigurements, malformations and mutilations, wounds, burns and injuries, intoxications, cancer, addiction and dependency, and mental disorders like schizophrenia – to mention only

some of the broad varieties of disease entities. Currently, the number of disease entities known to medical science totals, at least, a five-digit figure. These disease entities are *kinds* or *types* of diseases, terminologically designated by *disease entity*, *unit of disease*, *nosological unit*, or *disease pattern*. As *types*, they must be conceptually distinguished from the individual cases of disease that form their instances (*token*).

Every individual case of falling ill is a case of at least one disease entity that is at its bottom and forms its basis. Particularly, all subjectively and objectively perceptible and observable pathological signs, symptoms, and findings are *manifestations*, hence *parts* of one disease entity (or concomitantly occurring disease entities). The term *disease entity* should not be misinterpreted as designating something like a physical object or body or a *kind* of physical objects or organisms. As Caroline Whitbeck puts it, "[. . .] a disease, in the sense of a *disease entity* (or *disease type*) could not be [. . .] very much like a body [. . .] Diseases are not particular physical objects, but this does not prevent their existence being as objective as types of rock or species of trees" (Whitbeck 1977, p. 623).

At the present time, the entirety of all existing disease entities is not yet completely known and discovered. Hence, it is possible that certain symptoms, or constellations of symptoms, are already known to medical science but the disease entity (or disease entities) of which they are manifestations is unknown or not yet ascertained. In the history of medicine and medical science, a typical pattern of discovery takes its course from primarily observing some single, isolated *symptoms* or clusters of symptoms, to secondarily lumping them together to typical constellations of symptoms called *syndromes*, to eventually identifying one *disease entity* by discovering the *causal connection* between them and thus identifying the consistent, unifying basis of all observed symptoms and findings in that syndrome. This typical course in history of medicine, together with the definitions of *disease entity* and of *syndrome*, was put forward in the handbook *Die klinischen Syndrome* by Leiber and Olbrich, founded in 1957. This huge handbook was an attempt at collecting all known syndromes of medicine and was continued unto the eighth edition in 1996; at present, it is transformed into an electronic resource and database (Leiber and Olbrich 1957; Burg et al. 1996).

On account of these structural properties, *identification* and *definition* of disease entities depend essentially on identification and recognition of its *primary cause*. As a first approximation to an explanation, the *presence* of a certain disease entity *explains* the *occurrence* of all symptoms and findings that form its parts and manifestations. From these two statements, guidelines regarding the explication of the notions of definition and explanation in the context of medical pathology are derived (Engelhardt 1975; Gifford 2011; Schramme 2012; Wulff et al. 1990).

The Definition of Disease Entities

Disease entities are kinds of processes in the course of individual human lives that exhibit a beginning (onset) and an end (outcome) in time and, consequently, a definite temporal extension (duration). In borderline cases, the onset may coincide

with the onset of individual life itself, viz., the time of procreation or of formation of a zygote. Likewise, the outcome of a disease may coincide with the end of life itself, viz., in the case of a lethal outcome. With the exception of congenital disease, the onset of a disease forms a *transition* within individual life, viz., the transition from the state when the disease is *not* (*yet*) present (= "relative health") to the state when it is (= "diseasedness"). In the case of congenital disease, the onset is not a transition *inside* individual life but a transition *identical* with the *onset* of individual life. The process or event that brings about this transition is designated by *primary cause*, *first cause*, *etiological factor*, or *primary lesion* and is taken to be a *unique* event (*the* cause or *the* etiological factor). From the viewpoint of causal analysis, the primary cause of a disease is a *necessary* condition that is *specific* for this disease entity. The primary cause of a disease is followed by a specific, typical chain or cascade of pathological events inside the affected individual. This specific chain or cascade of events forms a temporal pattern that is called *natural course*, *natural history*, or *pathogenesis*. Generally, pathogenesis takes place simultaneously on multiple, diverse levels of the patient organism. Pathological phenomena and changes may occur on the levels of biochemistry and molecular biology; of morphology and function of cells, tissues, and organs; of function and development of whole systems of the organism; and of the perceptible and observable phenomena of behavior and experience of the ill person. Because all these levels are causally connected and intertwined, there is no sharp distinction between "clinical" and "pathological" levels, though for pragmatic reasons this distinction is retained. Even the subjectively experienced *illness* of the patient is, in the view of medicine, only a small part of the entire course of the disease (therefore, disease and illness are *in medicine* not mutually exclusive concepts, as they are in some *philosophical* theories of medicine). At the present time, there are only very few attempts to analyze and reconstruct the concept of disease entity from a philosophical point of view (Whitbeck 1977; Reznek 1987; Hucklenbroich and Buyx 2013; Hucklenbroich 2014a; 2016a, b).

The identification of a novel disease entity presupposes a comparison of its cause and course with the whole system of known disease entities – the medical *nosology*. It presupposes (i) identification of a novel, so far unknown, etiological factor, which (ii) causes a novel natural course of disease or pathogenesis that is not identical with and cannot be subsumed under an already known pathogenesis.

In the case of identified, well-established disease entities, there are, principally, three different ways to define them:

(i) Definition by its unique etiological factor
(ii) Definition by its typical, specific pathogenesis
(iii) Definition by parts or manifestations that form necessary and sufficient conditions (= obligatory and pathognomonic symptoms) of it

The account of disease entities in medical textbooks usually combines methods (i) and (ii), by describing its etiological factor as well as its typical natural course. Although this description is usually presented as the "definition" of the disease, this is not a correct method of defining the concept in the logical and philosophical sense

of definition, because definitions must not be *creative*, i.e., they ought not entail *empirical* consequences (Essler 1970, p. 71). But the statement – entailed by the "textbook method" of definition – that a definite etiological factor causes a definite natural course of disease *is* an empirical statement derivable from it. Hence, the textbook definitions are not definitions in the logical sense but are *empirical theories* about causal connections between etiological factors and their effects, constituting a definite disease entity (Gøtzsche 2007). As Henrik Wulff puts it:

> In order to define a disease, it is necessary to fix a set of criteria *which are fulfilled by all patients said to be suffering from the disease and by no patients not said to be suffering from the disease*. In some textbooks the description of a disease begins with a 'definition' but on closer examination it is usually found not to be a logically satisfactory definition but only an ultrashort description. (Wulff 1976, p. 50)

The third method of defining a disease entity, by specifying necessary and sufficient conditions, must not be confused with a similar but logically different method of presenting disease entities found in medical textbooks, namely, "defining" by specifying *diagnostic criteria*. The difference may be characterized as follows: diagnostic criteria of a disease comprise conditions that are conclusive evidence for its presence but are not necessary conditions; hence, they are not bound to be present in every instance of it and are not usable as defining criteria. Defining criteria, vice versa, may be but are not bound to be employable for diagnostic purposes, because they may be remote in time or practically inaccessible for diagnostic techniques. Diagnosis of a particular disease entity is proven or ascertained by so-called *pathognomonic* findings. A pathognomonic finding of a disease entity, or a set of findings that are, taken together, pathognomonic of it, is a finding or a set that forms a sufficient condition of it – thus the diagnosis is established. But these sufficient conditions need not be necessary conditions. Necessary conditions of a disease entity are called *obligatory* symptoms or findings that play a different role in diagnostics: They are useful for the *exclusion* of a diagnosis, if they are missing. Therefore, positive diagnostic criteria for a disease are not identical with a definition, unless they are, at the same time, obligatory findings.

To sum up, it may be stated that the meaning and usage of *definition* in medicine deviates from strict logical conventions (i) by calling characterizations of diseases (disease entities) *definitions* that are, in fact, *empirical theories* and (ii) by calling diagnostic criteria for diseases (disease entities) *definitions* that are, in fact, only *sufficient conditions* but need not be necessary conditions, as required for a proper, genuine definition.

Explanation and Diagnosis

It is common in medical communication to state that a particular symptom or finding S, or a particular set of symptoms and findings S in a patient, is *explained* by a certain disease (disease entity) D or, better, by the presence of D. This manner

of speaking may be reconstructed in the following way: as shown above (section "Some Concepts and Principles of General Medical Pathology"), all symptoms and pathological findings in a case of disease entity D are manifestations of D and, hence, *parts* of D. In the same sense, as the presence of a part is explained by the presence of its whole, the presence of S is explained by the presence of D ("part-whole explanation"). Furthermore, as shown above, all symptoms and findings of a disease entity D are *causally* connected by a causal chain or cascade that starts from the primary cause or etiological factor. Thus, it is possible to explain a symptom or finding S by antecedent members of the chain, ultimately by the primary cause. This kind of explanation is not a part-whole explanation but a *causal* explanation of S. These two kinds of explanations of S are not inconsistent with one another but are complementary, because they operate on different levels of description and conceptual resolution of the same unitary and consistent process.

The statement that, in a particular patient X, a case of disease entity D is present is called a *diagnosis* or diagnostic statement D(X). Therefore, it is possible to say that a part-whole explanation of S by D represents an "explanation *by diagnosis*," because the symptom S(X) that is to be explained (*explanandum*) is logically derived from the statement that D is present in X (diagnosis D(X)) and the theoretical description of disease entity D (together forming the *explanans*).

This way of explaining symptoms may be called "explanation by diagnosis," because diagnosis D(X) forms part of the explanans. But this *explanation by diagnosis* must be distinguished sharply from *inference to diagnosis*: inference to diagnosis D(X) is identical with *proof of diagnosis* (i.e., proof of statement D(X)), as sketched above (section "The Definition of Disease Entities"). Proof of diagnosis uses pathognomonic findings (sufficient conditions), and its logical direction runs inversely to explanation, viz., from statements S(X) and theoretical knowledge about D to diagnosis D(X). Thus, in *explanation* we infer from an established, known diagnosis D(X) to established, known symptoms S(X), whereas in *proof* we infer from established, known symptoms S(X) to a hitherto unknown particular diagnosis D(X) (Nordenfelt and Lindahl 1984; Schaffner 1985; Schaffner 1993; Stegmüller 1983; Wieland 2004.

Explanation of Pathologicity

A special case of explanation in medicine is formed by answering the question as to *why a particular condition C is a disease* or *why it is pathological*. To answer this question, what is required is not the concept of disease or disease entity but the concept of *pathologicity*, a technical term of medical science. A more common term for the same purpose is the term *disease value* (German: *Krankheitswert*). But the term *disease value* is seductive and may mistakenly lead to the opinion that the question of pathologicity is an evaluative question, in the sense of subjective, sociocultural, or ethical values and is dependent on cultural and historical variations

and changes, thus forming a culturally relative notion. However, for medical theory and medical understanding, it is an essential precondition that any judgment of pathologicity does not resort to evaluations of the kind mentioned above (Hucklenbroich 2016a). Instead, medical judgments of pathologicity refer to criteria that rely on objective or at least objectifiable facts. For example, one main criterion of pathologicity refers to the question whether condition C will cause a *premature*, early death of the person X affected by C: will X under condition C suffer death earlier than under condition non-C?

A second criterion of pathologicity, or better a second set of criteria of this sort, refers to the question of whether condition C implies or causes the presence of definite, certain natural signs (symptoms, complaints) that indicate a state of disease. The most prominent sign of this sort is *pain*, but there are lots of other signs such as nausea, dyspnea, dizziness, blackout (syncope), tremor, insomnia, hallucinations, etc. These natural, objectively identifiable signs of pathologicity are deeply entrenched in the psychosomatic nature of human beings; they are not due to any subjective or sociocultural values or norms but are universally valid in mankind. The complete, systematic account of all criteria of pathologicity forms an essential part of *general pathology* and *symptomatology*, as components of medical theory (Hucklenbroich 2014a, b, 2016a, b).

The whole system of disease entities recognized by contemporary medicine relies, in the last instance, on the system of criteria of pathologicity. Thus, the general question as to why a particular condition C is pathological, or is a disease, or possesses disease value, may be answered by a statement of the following form:

> Condition C is a disease, or is pathological, or possesses disease value, *because C falls within the scope of at least one criterion of pathologicity.*

Statements of this form constitute a unique kind of explanation that is specific to theoretical medicine.

Definitions of Key Terms

Disease entity	Kind of disease
Pathologicity	Property of being pathological
Etiological factor	The (unique) primary cause (of a disease)
Pathogenesis	Natural course (of a disease)
Diagnosis of a disease D	Statement "D(X)" (referring to a patient X)
Illness	The parts of a disease process that are subjectively experienced and evaluated by the patient
General pathology	Complete body of medical theories concerning general features of normal and pathological conditions (or concerning states and processes of health and disease)

Summary Points

- Diseases are instances (or cases) of disease entities.
- An illness is the part of a disease that is subjectively experienced and evaluated by the affected person.
- Disease entities may be defined:
  - By their unique, specific etiological factor
  - By their specific pathogenesis
  - By any subset of their symptoms and pathological findings that are necessary and sufficient conditions of their presence
- Symptoms and pathological findings S of a disease D may be explained:
  - By the presence of the disease D (part-whole explanation)
  - As a causal effect of the etiological factor E of D or of some subsequent pathological process in the pathogenesis of D that causes S (causal explanation)
- Explanation of S by (presence of) D must be distinguished from diagnostic proof of D by S.
- Pathologicity is established by a system of criteria that are made explicit in general pathology.
- Explaining why condition C is a disease, or why C is pathological, is proving that C falls within the scope of at least one criterion of pathologicity.

References

Bickley LS, Szilagyi PG (2016) Bates' guide to physical examination and history taking, 12th edn. Walters Kluwer, Philadelphia

Büchner F, Letterer E, Roulet F (eds) (1969) Prolegomena einer Allgemeinen Pathologie. Springer, Berlin

Burg G et al (eds) (1996) Leiber – Die klinischen Syndrome, 8th edn. Urban & Schwarzenberg, München

Engelhardt HT Jr, Spicker SF (eds) (1975) Evaluation and explanation in the biomedical sciences. Reidel, Dordrecht

Essler WK (1970) Wissenschaftstheorie I. Definition und Reduktion. Alber, Freiburg

Gifford F (ed) (2011) Philosophy of medicine. Elsevier, Amsterdam

Gøtzsche P (2007) Rational diagnosis and treatment. Evidence-based clinical decision-making, 4th edn. Chichester, Wiley

Hammer GD, McPhee SJ (eds) (2014) Pathophysiology of disease, 7th edn. McGraw Hill, New York

Hucklenbroich P (2014a) "Disease entity" as the key theoretical concept of medicine. J Med Philos 39:609–633

Hucklenbroich P (2014b) Medical criteria of pathologicity and their role in scientific psychiatry – comments on the articles of Henrik Walter and Marco Stier. Front Psychol 5:128. doi:10.3389/fpsyg.2014.00128

Hucklenbroich P (2016a) Die Normativität des Krankheitsbegriffs: Zur Genese und Geltung von Kriterien der Krankhaftigkeit. Analyse & Kritik 38:1–38

Hucklenbroich P (2016b) Disease entities and the borderline between health and disease: where is the place of gradations? In: Keil G, Keuck L, Hauswald R (eds) Vagueness in Psychiatry. Oxford University Press, Oxford, pp. 75–92

Hucklenbroich P, Buyx A (eds) (2013) Wissenschaftstheoretische Aspekte des Krankheitsbegriffs. Mentis, Münster
Leiber B, Olbrich G (1957) Wörterbuch der klinischen Syndrome. Urban & Schwarzenberg, München
Nordenfelt L, Lindahl BIB (eds) (1984) Health, disease, and causal explanation in medicine. Reidel, Dordrecht
Reznek L (1987) The nature of disease. Routledge, London
Riede U (2004) Color atlas of pathology. Pathologic principles, associated diseases, sequelae. Thieme, Stuttgart/New York
Sandritter W, Beneke G (eds) (1974) Allgemeine Pathologie. Schattauer, Stuttgart
Schaffner KF (ed) (1985) Logic of discovery and diagnosis in medicine. University of California Press, Berkeley
Schaffner KF (1993) Discovery and explanation in biology and medicine, 2nd edn. University of Chicago Press, Chicago
Schramme T (ed) (2012) Krankheitstheorien. Suhrkamp, Berlin
Siegenthaler W (ed) (2007) Siegenthaler's differential diagnosis in internal medicine. From symptom to diagnosis. Thieme, Stuttgart/New York
Stegmüller W (1983) Erklärung – Begründung – Kausalität. Springer, Berlin
Whitbeck C (1977) Causation in medicine: the disease entity model. Philos Sci 44(4):619–631
Wieland W (2004) Diagnose, 2nd edn. Hoof, Warendorf
Wulff HR (1976) Rational diagnosis and treatment. Blackwell, Oxford
Wulff HR, Andur Pedersen S, Rosenberg R (1990) Philosophy of medicine. An introduction, 2nd edn. Oxford, Blackwell

Cultural Influences on Medical Knowledge 49

David Hughes

Contents

Abstract

This chapter examines how culture influences the content and practical application of medical knowledge. The current state of knowledge about pathology and treatment is not simply the outcome of a neutral process of scientific investigation and discovery, but is shaped by changing theoretical frameworks affected by more general cultural perspectives. Just as the disease classification systems utilized by doctors emerge in a social context, so lay health beliefs reflect local cultural perspectives, and medical practice involves mediating between expert and lay belief systems. Moreover, medical practice is itself conditioned by the subcultural perspectives associated with the medical profession, its constituent specialisms, and the diverse hospital and community settings where healthcare is provided. The dual nature of medicine as both a scientific and practice-based discipline has resulted in tensions between the art and science of practice, with some doctors putting more weight on clinical judgment based on experience

D. Hughes (✉)
Department of Public Health and Policy Studies, Swansea University, Swansea, UK
e-mail: d.hughes@swansea.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_73

rather than the standardized application of codified knowledge. More generally there remains a divide between practitioners and laboratory-based medical research which reflects the history of medicine in Western countries.

Introduction

Culture powerfully shapes human understandings of the world, including knowledge about health and illness. "We seldom realize," wrote the popularizing philosopher Alan Wilson Watts (1989: 53–54), "that our most private thoughts and emotions are not actually our own. For we think in terms of languages and images which we did not invent, but which were given to us by our society." Culture provides the conceptual scaffolding via which people make sense of the objects and events around them. Members of different cultural groups see the world in different ways and cultural perspectives change over time, so that time and place crucially affect expert "knowledge."

Culture as understood by most scholars encompasses language and its associated classificatory taxonomies, social norms, customs, moral precepts, and other symbolic resources such as visual art, music, and dance. For the purposes of this chapter, we adopt a wide definition in which the ideational aspect of a culture includes its modes of analysis, its characteristic forms of problem solving, and its technologies for generating new knowledge. This rests on the proposition that the constellation of beliefs and values found in a given social group influences the search techniques and tools of discovery that it uses.

Cultural influences affect the domains of expert as well as commonsense knowledge and shape the behavior of professionals as much as the laity. This chapter considers medical knowledge from both expert and lay perspectives and examines how it is shaped by wider social and professional influences. It starts by considering the conventional image of medical knowledge as a progressive unfolding of scientific discovery and a contrary view from social science and philosophy that argues that such knowledge must be seen in a social and cultural context. Later sections deal with the changing nature of disease classifications; the complicated linkages between scientific advance, technology, and culture; patient cultures and illness behavior; the culture of medical practice; and the relationship between medicine and science.

Medical Knowledge and the Narrative of Scientific Discovery

A key question in considering cultural influences on medical knowledge is whether the latter emerges from a direct engagement with the facts of the natural world or must be regarded, at least in part, as a human product influenced by the wider society. From the perspective of mainstream Western medicine, understandings of disease are characterized by a progressive accumulation of knowledge over time.

Diseases are conceptualized as distinct entities that present a recurrent signature or natural history associated with known signs and symptoms and may be managed (more or less successfully) with a repertoire of treatments shown to be effective by evidence concerning past outcomes. According to conventional histories of medicine, science discovers better ways of treating known diseases and finds new diseases. With the discovery of effective treatments, the incidence of "old" diseases, such as smallpox, polio, mumps, and dracunculiasis, has reduced dramatically. As knowledge advances, medical scientists may find that what was considered to be a single disease has more than one variant or discover new diseases. Since the mid-twentieth century new conditions as diverse as AIDS (acquired immune deficiency syndrome), SARS (severe acute respiratory syndrome), Ebola fever, Marburg hemorrhagic fever, hantavirus pulmonary syndrome, post-traumatic stress disorder, and chronic fatigue syndrome have found their way into medical textbooks.

Yet even within the ranks of the medical profession, some observers questioned whether the image of step-by-step discovery of an obdurate, external reality told the whole story. While medical knowledge was undoubtedly advancing, some raised doubts about the conceptualization of diseases as stable entities that had existed even before science discovered them. Lester King (1954: 199), a former editor of the *Journal of the American Medical Association* (JAMA) and president of the American Association for the History of Medicine, observed that:

> We are faced with the problem whether certain relational patterns, like diseases, "exist in nature", while other patterns, like a melody or a poem, we can create arbitrarily by our own skill and ingenuity. The question becomes, does a disease, whatever it is, have real existence, somehow, in its own right, in the same way as the continent of Australia? Such real existence would be independent of its discovery by explorer or investigator. A disease exists whether we know it or not. The contrasting point of view would hold, that a disease is created by an inquiring intellect, carved out by the very process of classification, in the same way that a statue is carved out of a block of marble by the chisel strokes of the sculptor.

King was not arguing the case for nominalism over realism, but rather pointing to the difficulty of grasping an underlying reality in which the precise nature of disease – the patterns observed by physicians – varied over time and between individual patients. Not only, in his view, did systems of classification shift over time, but the disease entities themselves might change as humans interacted in different ways with their changing physical and social environments.

The uncertain relationship between causes identified by scientific medicine and the effects produced in particular individuals, as well as the constantly evolving nature of diseases, were central themes in René Dubos' (1959) celebrated work, *Mirage of Health*. Dubos described how humans provide a habitat for microbes that can easily transform into virulent pathogens and highlighted medicine's inability to explain why the presence of indigenous microbial flora led to infection and disease in one person but not another. For Dubos, human and bacterial populations are part of the same evolving biosphere, and a world in which drugs remove all

bacteriological threats is an unattainable goal. Just as humans change to cope with the threat posed by microbes, both via the discovery of new treatments and natural adoptive mechanisms such as immunity, so microbes mutate to exploit weaknesses in biological defenses and develop resistance to previously effective antimicrobial drugs. Thus in place of the image of linear scientific advance, Dubos articulated a vision of changing patterns of interaction between human hosts and evolving microorganisms in which previously eradicated diseases might reappear and established drug therapies might become ineffective. Given that humans share a biosphere with other living organisms, and that interactions may be mutually harmful, some scholars argue that it becomes hard to distinguish diseases from other natural processes.

It can be argued that a similar picture of progress counterbalanced by new challenges (or the return of old ones) is visible in many other areas of medical practice. Advances in the treatment of infectious and other acute diseases need to be weighed against a rising incidence of heart disease, cancers, Alzheimer's disease, and other chronic conditions, which is affecting many advanced countries because of factors such as increasing life expectancy, changes in lifestyle and diet, and growing social inequality (Nordenfelt 1990). Moreover many of the treatments that are developed offer incremental rather than "big step" health gain, so that skeptics write of "halfway technologies" and point to the high cost of interventions that may at best buy a few more months of life (Thomas 1971).

In the philosophy of science, the notion of linear scientific advance was challenged by Kuhn's (1962) seminal work *The Structure of Scientific Revolutions*. Rather than steady, cumulative progress, Kuhn identified discontinuities that arose as periods of "normal science" were punctuated by moments of revolutionary change, when one dominant scientific paradigm was displaced by another. In the "normal" phase a community of scientists who share a general perspective and a set of associated theories – a paradigm or "disciplinary matrix" – seek to fill gaps or resolve anomalies revealed by observations which do not fit with the existing paradigm. Generally this results in incremental modifications to theory or revision of faulty evidence. However, Kuhn points to a pattern where over time changing search technologies and new directions of inquiry throw up an accumulation of observations that do not fit with existing theory. This leads to increasing "debate over fundamentals" and a search for a new conceptual framework – a new world view. After a period of turmoil and controversy, an older paradigm such as Newtonian mechanics is replaced by its successor, quantum physics.

Although Kuhn himself did not use examples from medical science to advance his argument, many later scholars have taken up the idea of changing paradigms in medicine. Medical scientists working in a specialty such as cardiology can be seen as a community of specialists who share a "disciplinary matrix," in the sense of frequent ongoing communication between in-group members and relatively high consensus about the current state of disciplinary knowledge (Hai 2009). Over time major shifts in knowledge occur, such as the transition from Galenic theory to the theory of blood circulation or the emergence of germ theory, often with considerable social resistance from interest groups with a stake in the existing paradigm

(Stern 1927). In our own time advances in genomics and regenerative medicine (stem cell research) suggest that a radical transformation of disciplinary theories and therapies is on the horizon (Perpich 2004; Latimer 2013). One controversial issue in Kuhnian analysis as applied to medicine has been whether changes that may be highly significant for practice in particular specialties really amount to paradigmatic change or are better seen as modifications of middle-level theory that leave existing paradigms intact. Developments in ulcer treatment following the discovery of *Helicobacter pylori* (see below) sparked controversy about what constitutes paradigm shift and whether medicine is a special case.

There is a family resemblance between Kuhnian approaches and the approach of scholars in social history and sociology who write of the social framing of diseases (Rosenberg 1989). Just as Kuhnian analysis directs attention to the dominant modes of reasoning within the scientific community and how the paradigmatic glasses through which it views data reflect its norms and culture, so the idea of framing recognizes that medical knowledge must be seen in a social and cultural context. Rosenberg (1989: 4) is interested in "the nexus between biological event, its perception by patient and practitioner, and the collective effort to make cognitive and policy sense out of those perceptions." The choice of the language of framing is a deliberate attempt to distance this approach from that of social constructivist writers, who in Rosenberg's view have tended to underplay the materiality of disease and exaggerate the degree of arbitrariness in scientific disease classifications, with the consequence that many case studies focus on "socially resonant diseases" such as hysteria, chlorosis, and neurasthenia. Actually this may misrepresent the position of constructivists who acknowledge that real biological pathology exists (see Wright and Treacher 1982; Nicolson and McLaughlin 1987), but, compared with the idea of "construction," the more neutral concept of "framing" sits more easily with the approaches of many social scientists and historians studying disease classifications and their social and cultural connections.

Culture, Disease, and Classification

The concept of classification refers to the idea that human knowledge is not merely an ensemble of facts, but a complex system of categories and ideas about category relations. To know the important attributes of some given phenomenon, it will be enough to place it in a category which shares its general characteristics, and human actors need only memorize those unique features that distinguish it from other items in that category. Health and illness are also understood via a process of classification and categorization (Bowker and Star 1999). It can be argued that the classification system employed by medical professionals overlaps with and is influenced by the systems of classification used in other scientific domains and indeed with general cultural knowledge concerning matters such as practical reasoning, problem solving, political interest, and morality (White 1991). Moreover, it is evident that classificatory schemata used in medicine evolve over time and may differ

somewhat from place to place, even among Western countries (Fabrega 1974; Helman 2007; Payer 1988).

It is obvious that disease classifications change over time, but perhaps more difficult to separate simple medical progress from changing understandings shaped by prevailing social mores and cultural perspectives. Many scholars argue that culture, conceived in broad terms, shapes the explanatory frameworks that emerge within the science of the time, as well as the search procedures, instruments of discovery, and modes of problem solving that are deployed.

The way that changing understandings of etiology and treatment shape disease labels, as well as the nature of the condition and the patient experience, is vividly illustrated by Peitzman's (1989) study of the changing framing of renal disease between the eighteenth and twentieth centuries. Peitzman shows how physicians trying to make sense of observed signs and symptoms moved through a succession of explanatory theories and disease classifications. In the late eighteenth century, "dropsy" was a general diagnosis for patients suffering from bodily swelling through an excess of fluid – what medicine today calls edema – which was understood in terms of existing humoral theories of illness. Although dropsy was a familiar malady encountered regularly by physicians, the clinical skills of the day were unable to differentiate edemas arising from different causes. It was only in the 1820s that doctors began to understand these symptoms in a different way, after Richard Bright distinguished that subset attributable to kidney dysfunction. Bright was able to construct a disease entity out of the association between the clinical picture in life, postmortem findings, and chemical changes in urine. Later in the century clinicians redefined and refined the characteristics of Bright's disease using such techniques as microscopic examination of tissue and urinary sediment.

The essence of Bright's disease was that patients got sick through their kidneys, but in the twentieth century this same conceptual space came to be occupied not by one disease but several. A new generation of doctors looked to physiology rather than the older lesion-based anatomical knowledge and searched for functional indicators using laboratory methods. Techniques applied to the stomach and the heart were applied by analogy to the kidney. As doctors began to perform tests for such things as dye excretion and urea loads, new terms such as renal insufficiency and renal failure entered the medical vocabulary. Renal failure meant retention of urea and other substances normally discharged by the healthy kidney. Gradually it became clear that patients might progress to renal failure without showing all the characteristics attributed to Bright's disease. Peitzman outlines a complicated succession of, sometimes competing, pathologic classification frameworks put forward by physicians and clinical scientists as they tried to identify subtypes of renal disease.

Peitzman's analysis concludes with an examination of a common diagnostic category applied to many renal disease patients from the late twentieth century to the present time, end-stage renal failure. Many dropsy sufferers would probably fall into that category if transported forward in time, and yet the experience of illness is completely different for the ESRF patient. Most are treated long before they become "dropsical," so that the bloated bodies characterizing eighteenth-century

sufferers would not be encountered by modern physicians. Nephrologists are now unlikely to come across the severe edema and uremia that characterized dropsy, and indeed much of their time is spent managing problems of the "cure" – dialysis – rather than the underlying pathology. The specialist rarely sees a kidney. The renal doctor is only likely to see kidneys as shadows on an ultrasound image or, microscopically, biopsied a small slice at a time.

The history of renal disease illustrates how a basic clinical picture is defined and redefined over time using a succession of differently focused explanatory frameworks. It is not merely that new disease labels are applied to constant physical phenomena, but one overarching conceptual scheme displaces another. Moreover this change is not simply a reflection of the state of a cumulatively developing corpus of medical knowledge, but depends on changing technologies, social practices, and professional and societal world views. Peitzman (1989: 21) argues that the shift of frame from dropsy to Bright's disease is not merely a new diagnosis based on new data, but represents a change in the way that doctors name diseases – "a new, nineteenth-century way of thinking about and defining disease." The focus shifts from the patient history and experiences reported to the physician to laboratory investigations that only the physician can perform. Anatomical observations are replaced by "functional diagnosis," and emerging nineteenth-century technologies yielded up new clinical parameters that helped to define the disease entity. When the explanatory limitations of Bright's disease become apparent, no single disease category emerged to cover the spectrum of diffuse renal disease. The term that came into widespread use – end-stage renal failure – had its origins in a 1972 Act of Congress, which outlined a practical threshold for public financial support for Americans requiring chronic dialysis treatment. ESRF is an administrative category bound up inextricably with twentieth-century US social policy and societal attitudes toward disabled people, but it became a shorthand diagnostic label commonly used by clinicians. As Peitzman states, each of the disease labels examined "has had its use, its particular reality, and its message," and each is closely related to the way of seeing of the time.

Peitzman tells us that the transition between frames takes time and is not accepted by all practitioners but says little about the conflict and micro-political struggles sometimes associated with scientific advances – something that is emphasized in the Kuhnian approach. The social anthropologist Bernhard Stern (1927, 1941), an early critic of the conventional history of medicine, focuses more on opposition to scientific advances and the social and cultural factors that retard the diffusion of innovation. In a classic text that still remains relevant today, Stern (1927) shows that almost all the major medical advances of the eighteenth and nineteenth centuries – from Harvey's theory of blood circulation to Jenner's work on vaccination and the breakthroughs of Semmelweis and Pasteur – met with resistance at the time of their discovery. Forces that impede change operate at the individual, group, and institutional levels. Individuals resist change because it can mean personal inconvenience, temporary pain, more work, and an end to old comfortable habits. For the group it can lead to disruption of existing routines and customs that disturb the status quo. At the institutional level, existing patterns

of tradition and authority tend to protect established practices. Stern highlights the brakes on change applied when a new theory conflicts with established cultural ideas and also the self-protective behavior of professionals or powerful interest groups who regard innovation as a threat to their economic interests. He argues that scientific advances are often intertwined with political struggles, and technical knowledge may be redefined to suit the profession's interest. Thus Stern (1941: 216) maintains that "medicine cannot develop and never has developed in isolation, that the nature of its role and its achievements are circumscribed by the soil in which it is rooted."

Scientific Progress and the Social and Cultural Context

How far, the reader may ask, is opposition to scientific progress a facet of medical history that is now firmly in the past? A more recent case that suggests its continued relevance concerns the discovery of the *Helicobacter pylori* bacterium and the resultant shift from the excess acid theory of peptic ulcers to the bacterial infection theory. The case illustrates how reluctance to accept new evidence may be related to issues of both culture and power.

In the 1960s the view emerged that ulcers were the product of stress resulting in an excess of acid damaging the surface of the stomach or duodenum. It was assumed that high acidic concentrations in these organs made it impossible for bacteria to survive, but this was challenged in 1979 when the Australian pathologist Warren observed spiral bacteria in microscopic slides prepared from endoscopy tissue biopsies. Further work over several years by a team led by Warren and the gastroenterologist Marshall established the existence of a previously unknown bacterium that they named *Helicobacter pylori*.

The *H. pylori* case has come to be associated with debates about the nature of scientific and medical paradigms and possible refinements of the Kuhnian position that are beyond the scope of this discussion (Thagard 1998; Gillies 2005; Hutton 2012). However, scholars on both sides of this debate acknowledge that Warren and Marshall's discovery was met with skepticism and resistance from the scientific community. Even though the theory of bacteriological infection was well established in other domains, it was some years before the causal role of *H. pylori* in peptic ulcers was widely accepted by practitioners, so that treatments based on the acid excess theory continued to be prescribed.

Collyer (1996) suggests that the new theory ran counter to the interests of large pharmaceutical companies, such as GlaxoSmithKline, which had invested heavily in profitable H2-antagonist drugs. The new treatment approach did not initially involve a purpose-designed antibiotic that might have had commercial appeal and so failed to gain corporate sponsorship for research and dissemination that the drug industry could have provided. It was only when Procter & Gamble realized that one of their patented products had anti-bacteriological as well as acid-reducing effects that the company began to support the ongoing research.

Nor was there much support for the new theory in the medical profession. Collyer (1996) argues that the new infection theory encountered resistance because it suggested that physicians' existing ideas about the importance of stress, lifestyle, and individual responsibility for behavioral change had been flawed. The older approach aligned with prevailing cultural stereotypes about the association between overeating and poor diet with stomach disease, but the new one suggested that, rather than being a matter of individual responsibility, high rates of recurrence of ulcers among treated patients reflected the limitations of present medical knowledge.

Commentators skeptical about the importance of cultural influences point out that several recent studies of new diseases (e.g., Richman and Jason 2001; Young 1997) involve psychological symptoms, in which the nature of underlying pathology is difficult to identify. The argument is that ideas about social construction or cultural framing are easier to apply when no physical disease exists. However, this criticism cannot be applied to Greaves' (1998) study of acute myocardial infarction, a leading cause of death in Western countries, which surprisingly did not emerge as a recognized disease entity until the twentieth century. It was only in the 1920s that the "new" heart conditions found their way into medical textbooks, and Greaves reviews the competing theories about why "heart attacks" had not been recognized earlier, including failure of diagnosis and the non-appearance of a disease of affluence in advance of the "epidemiological transition" to modern lifestyles. He finds both explanations unconvincing but is also dissatisfied with social constructivist accounts that portray the discovery of myocardial infarction as the result of a new disease classification connected with the emerging specialism of cardiology and its need to establish an expert knowledge base. Greaves instead argues that material changes in lifestyles and risk came into play at the same time that doctors were moving toward new theories of heart disease and patients toward new ways of understanding their illnesses. There was a "looping effect" (see also Hacking 1996) whereby medical and lay definitions comingled in shaping societal perceptions of the new condition. This was reinforced by "a cultural climate and expectation which is conducive to and sustains part of the epidemic of these heart disorders" (Greaves 1998: 139). Greaves' analysis thus involves a complex composite of objective and subjective and individual and social factors, in which there is an overlap between science and the wider corpus of cultural knowledge.

Lay Health Beliefs

Greaves' mention of the interplay between professional and patient definitions suggests that disease classification systems need to be considered in conjunction with lay understandings of illness. Physicians are familiar with a range of disease entities and their characteristic manifestations but also have firsthand knowledge of how patients understand and respond to their experiences of illness. Medical practice involves taking account of these lay perspectives and finding ways to mediate between professional and patient understandings of health and illness.

Illness refers to a person's subjective experience of ill health. As Fabrega (1974: 120) writes, "People don't feel X-ray shadows, blood chemistries, or auscultatory findings. Rather, people feel or report weakness, coughing, and excessive urination." Thus illness refers to the individual's perception of feeling unwell – pain, discomfort, and so on – and any modification of normal behavior that results. This is heavily shaped by culture, as well as social position and personality. Like disease classification systems, lay health beliefs and illness behavior vary according to place and time.

The example of depressive disorders is often used to make this point. Early studies suggested that patients from certain non-Western cultures reported fewer symptoms related to internal mood states and more relating to physical symptoms. Arthur Kleinman (1980) studied depression and neurasthenia in Taiwan and mainland China and found that patients used terms referring or relating to the body while utilizing few categories corresponding to Western psychological states. This supported the notion that somatization, the physical presentation of psychological distress, was more common among non-Western populations, an idea that has been challenged in more recent debates. Contemporary scholars regard somatization as a worldwide phenomenon but argue that different groups present somatic symptoms in different ways related to wider patterns of cultural meanings and the various psychological and social functions that somatization serves (Kirmayer and Young 1998; Kohrt 2014).

Differences between lay belief systems and Western medicine are not confined to distant cultures. Chrisman (1977) suggested a framework for the cross-cultural analysis of folk ideas about illness, setting out some of the basic modes of thought about illness – what he calls "thought logics" – that apply in many countries. He identifies a logic of degeneration or the running down of the body, a mechanical logic concerned with blockages or damage to bodily structures, a logic of balance linked to the disruption of bodily harmony, and a logic of invasion involving germ theory and material intrusions.

These logics are readily apparent in the findings of some of the important British research in this area. A classic study by Blaxter (1983) interviewed a sample of middle-aged, working-class women in Scotland about their ideas on health and illness. When questioned about the causes of illnesses, most women gave explanations which bore only a loose resemblance to those of medical science. Infection was the most commonly cited cause, followed by heredity, then by environmental hazards, and then other factors such as the secondary effects of other diseases, stress, and childbearing. Another seminal study carried out by Pill and Stott (1982) in South Wales looked at women in their early 30s who came from skilled manual backgrounds. Again infection (or "germs") was the most commonly mentioned cause of illness, followed by lifestyle, heredity, and stress. About half the women in the sample utilized concepts of causality that implied that illness was associated with choices about behavior and a degree of individual responsibility. These women were more likely to be homeowners and to have had more education than the women in Blaxter's sample, and their feeling of greater control over their lives may account for the different emphasis.

The logic of invasion can be seen as the cultural result of the theories of microbiology that are central to mainstream medicine. But heredity seems to come up more in lay belief systems than in conventional medicine, and lay beliefs which emphasize "stress," worry, and tension seem closer to holistic approaches than to the mainstream biomedical model, which emphasizes physical processes.

One important point is that although lay beliefs often appear to contain illogical ideas or inconsistencies, they can be part of a wider system of beliefs that makes sense to participants. This is illustrated by Helman's (1978) study of patterns of belief about infectious diseases in a North London community. Helman argues that some common infectious diseases that involve raised body temperature are understood in terms of a folk belief system which is quite distinct from medical science. Patients distinguish the subjectively "hot" diseases that are usually thought of as fevers from the cold diseases that are classified as colds or chills. Each of these two categories is associated with a set of ideas about cause, the course of the illness, treatment, and the degree of blame attaching to the sufferer. Colds and chills are seen as a result of the interaction between the individual and unfavorable environmental conditions, particularly low temperatures, which through dampness, cold winds, and drafts penetrate vulnerable surfaces of the body such as the head and feet. Transitions such as moving into a cold room after a hot bath are believed to make the individual particularly vulnerable. Treatment involves restoring temperature balance by hot drinks or a warm bed. Individuals often believe themselves to be to blame for getting a cold because of irresponsible actions like going outside with wet hair and the like. Fevers on the other hand are due to invisible entities – germs or bugs – transmitted from individual to individual. One important treatment is fluid that flushes out germs. The individual carries less personal blame for fevers because they are unavoidably transmitted through contact with other people. Helman points to the similarities between the way people in Britain talk about germs and people in simple agricultural societies talk about spirits – from the point of view of folk beliefs, both are intangible and hypothetical and strike in mysterious ways.

These early studies have been supplemented by a corpus of later research that confirms variation in health beliefs and behavior among men and women from different social classes, geographical areas, and ethnic groups (for reviews see Lupton 2003; Stainton-Rogers 1991; Blaxter 2010). However, scholars differ in their views about whether divergent lay perspectives on illness are entirely a cultural phenomenon, depending on beliefs and attitudes passed from generation to generation. Sociologists in particular have often argued against the proposition that a culture of poverty, in which successive generations in disadvantaged populations engage in unhealthy behaviors, is the primary explanation for social class differences in morbidity and mortality. They have instead argued that material differences in the living conditions and life chances of richer or poorer social groups affect the resources available and everyday experiences of health and illness.

From the point of view of doctors and other healthcare professionals, understanding such differences is important for good practice. Sensitivity to these matters

pays off even in narrow terms of clinical effectiveness. For example, Iniu et al. (1976) examined how far patients treated for hypertension cooperated with treatment. Doctors who had received training about patients' health beliefs were encouraged to discuss patients' ideas more fully before explaining the diagnosis and arranging treatment. When these patients were compared with a control group who had not had the benefit of such a discussion, they were found to comply better with the prescribed drug regime and to achieve better blood pressure control. Harwood (1971) gives another example in a culture contact situation. Puerto Ricans in New York retain a belief system in which all illnesses, medicines, and foods are hot or cold. Vitamin supplements prescribed for pregnant women were frequently not taken because they were believed to be "hot" and to cause rashes and irritations to babies. However, there was no difficulty if supplements were taken with fruit juice – which is classified as cold. Other important studies show how, in areas as diverse as the implementation of Ebola control policies (Hewlett and Hewlett 2008), HIV/AIDS education programs (Lyttleton 1993), and the treatment of immigrant populations in developed countries (Fadiman 1997), the successful application of ideas from Western medical science depends on awareness of, and sensitivity to, local or migrant cultures.

Medical Culture

The practical application of medical knowledge involves interaction with other healthcare professionals in a range of clinical sites, typically associated with distinctive organizational subcultures. The culture of medicine, or more specifically the subcultural beliefs and practices of the specialism or healthcare locale in which the individual doctor is practicing, is an important factor that affects how expert knowledge is mobilized in real-world situations.

Medical culture is shaped by convergent influences such as medical school but also affected by the different career trajectories and work environments of practitioners. Professional cultures are transmitted both through the formal training process and the bedside experience of speciality work. Career advancement may be more dependent on normative compliance rather than technical excellence, and the literature suggests the importance of sponsorship and the existence of an influential patronage system (Bosk 1979; Atkinson 1981). Nor is the single hospital or clinic necessarily the unit of analysis; cultures may cut across organizational boundaries, as when medical consultants hold appointments or admitting rights in more than one facility. In cultural terms hospitals are becoming more rather than less complex entities as they adjust to an increasingly complex division of labor, a proliferation of special locales, and a range of new occupational categories. This has been associated by some scholars with "tribalism" and conflict, not just between different occupational groups such as doctors and managers, but within the medical profession itself. For example, sociologists argue that the medical profession has adjusted to oversight by general managers by "re-stratifying" itself to create a group of management-oriented doctors who act as mediators between clinicians and

hospital administration and may sometimes support rationalized policies opposed by professional colleagues (Numerato et al. 2012).

Interactions between doctors are guided by in-group norms and tacit rules. Both talk in medical consultations with patients and the language of case presentation among colleagues take a stylized form that serves to legitimize the expertise and authority of professionals (Atkinson 1995). Generally medical talk communicates the objective nature of decision making and the uniform competence of the practitioners. However, hierarchy and disciplinary rivalries also enter the picture. Thus, Atkinson (1995) found that the hematologists he studied erected subtle "us/we" and "them/they" distinctions when they considered the evidence assembled by the team compared with other more distant colleagues, weaving into their case narratives delicate attributions of differential credibility and sometimes blame.

Ethnographic studies suggest that a gap exists between the version of medical practice presented in public forums, or in consultations with patients, and the version communicated between colleagues behind the scenes. The discrepancy between public and private accounts may be especially clear when error is involved. A number of researchers describe how doctors distinguish between different types of error and determine what constitutes an error in particular circumstances. Clinical uncertainty and the unpredictability of treatment mean that what constitutes a mistake may be a highly contested matter. Bosk (1979) enumerates four types of error recognized by doctors. "Technical errors" occur when a surgeon is performing his role conscientiously, but his skill falls short of what the task requires. "Judgmental errors" occur when an incorrect strategy of treatment is chosen. "Normative errors" occur when a surgeon (usually a subordinate) fails in the eyes of others to discharge his/her role obligations conscientiously. "Quasi-normative errors" occur when subordinates fail to follow the practices or techniques favored by individual senior surgeons ("attendings"). Bosk found that the first two categories were usually seen as involving honest mistakes that were an accepted cost of training. But the last two types of error breached moral rules, specifically the etiquette governing role relations between senior surgeons and house staff, and were regarded in more serious light. Where a junior made repeated technical errors, he or she might still be regarded as a conscientious professional who could pursue a career in another branch of medicine, but repeated normative errors were taken to indicate unsuitability for the profession. Bosk describes how peer surveillance of performance takes place through a series of rounds, reviews, and conferences in which the moral meanings surrounding surgical error are reinforced. While colleagues take a supportive stance toward technical errors and perceive them as occasions for learning lessons, they are unforgiving and intolerant of moral errors.

For example, professional self-protection has been blamed for continuing high rates of iatrogenesis and several recent scandals about care in British NHS hospitals. After the discovery of high mortality rates in a pediatric cardiac surgery unit, the 2001 Bristol Royal Infirmary Inquiry Report suggested that professional culture has played a significant part in hiding and amplifying bad practices. It mentioned factors such as a mind-set of "professional hubris" in a teaching hospital, a "club culture" with insiders and outsiders, professional rivalries, the unwillingness of

senior doctors to engage with interdisciplinary teams except as team leaders, the covering up of patient deaths on the basis that surgeons were on a "learning curve," and the discouragement of "whistle-blowing." A decade or so later, similar issues were laid bare by the Francis Inquiry Report following revelations about poor patient outcomes at the Mid Staffordshire NHS Trust. The Inquiry wrote of lack of engagement between management and senior doctors, an over-preoccupation with targets, a tendency to "close ranks" to hide problems, and a "culture of fear" preventing disclosure of adverse incidents (Holmes 2013). These were perhaps the two most prominent in a string of recent British scandals that highlight how medical practice within modern multidisciplinary healthcare settings, subject to financial pressures and increasingly rationalized management regimes, is affected by social factors that influence how medical knowledge is applied in the treatment of individual patients.

Science and Art in Medical Culture

A final point for consideration is how medicine fits into the wider domain of science. Different cultures segment and organize their corpus of expert knowledge in different ways, including how they map domains such as religion, magic, philosophy and science, and the relations between them (Fabrega 1974). In Western Europe medicine did not always align itself with science, and indeed there is a degree of continuing ambivalence about the relationship among medical practitioners.

Before the 1840s Western medicine was practice based and mixed a romantic philosophy of nature with mystical ideas about spirituality and machine metaphors of the body (Verwey 1990). It was only in the mid-nineteenth century that medically trained researchers in German universities incorporated the new physiology into the medical curriculum and some years later before the British universities followed suit (Jewson 1976). The growth of the research laboratories depended on the institutional support of the medical schools. The rise of scientific medicine provides a new source of legitimation for the previously diverse and individualistic craft of healing. It was "a powerful and compelling means of conferring "expert status" on medicine, thereby consolidating its position as an "autonomous" learned profession" (Austoker 1988: 31). William Osler pioneered "science at the bedside" practice based on the growing body of laboratory-derived knowledge and brought about striking advances in the treatment of vitamin deficiencies and pernicious anemia (Beeson 1980).

However, within the medical profession there remained an undercurrent of resistance to a purely technical medicine that has persisted into the modern era. Thus the rise of scientific rationality led to a defensive counteraction through the reaffirmation of an older clinical tradition, which emphasized the individuality of patients and the indeterminacy of the practitioner's experiential knowledge and skills (Jamous and Peloille 1970). French hospital doctors responded to growing "technicality" by appealing to the mystery of clinical experience, a body of

knowledge not susceptible to precise codification. Over time a clinical tradition that had originally developed in eighteenth-century France had a continuing influence on physicians in Europe and North America and finds its voice in opposition to modern developments such as standardization and the application of clinical decision theory. This is manifest in the continuing tension between healthcare managers and medical professionals mentioned earlier and highlighted in events such as the scandal concerning patient care in the Mid Staffordshire NHS Trust.

The uneasy relationship between medicine and science reflects the fact that science is not fully under the medical profession's control. Clinical researchers do not enjoy clear superiority over nonclinical scientists working in chemistry, physiology, pharmacology, or genetics, who are often based in the research institute rather than the medical school (Strong 1984). The basic medical education is not a sufficient preparation for advanced bioscience research. Generally speaking medical graduates in Western countries learn only enough to be able to mediate between science and practice. This mediation effectively comes to mean controlling patients' access to the products of scientific research. For example, physicians function as "gatekeepers" to a variety of high-cost interventions or drugs. The medical profession's close contacts with the pharmaceutical industry and its legal monopoly in many countries over prescribing ensure that it does not share this strategic position with other occupations.

Definitions of Key Terms

Culture	Is the system of language, belief, and knowledge that shapes a given social group's understanding of the world; for the purposes of this chapter, this includes the group's characteristic modes of problem solving and the techniques via which it generates new knowledge.
Medical knowledge	Is the corpus of empirical observations, evidence, and theory transmitted via published literature, databases, and oral communications between medical practitioners and researchers and commanding significant support within the profession at a given time.
The medical profession	Consists of the persons formally licensed or registered by authoritative national bodies to engage in medical practice; generally speaking governments grant the profession authority to define what counts as legitimate medical knowledge and to regulate the conduct of its members.
Lay health beliefs	Are the beliefs that people hold about the health, illness, and disease; they are shaped by culture, social position, and individual biographies.

Medical culture	Refers to the language, thought processes, styles of communication, customs, and beliefs observable in medical education and practice; it is transmitted both in the medical school and via informal socialization in clinical settings.
The art and science of medicine	Refers to two strands within medical opinion that respectively emphasize the experience and intuitive clinical skills of individual doctors and the value of systematic scientific evidence and standardization and rationalization in medical decision making.

Summary Points

- Culture, the system of symbols and beliefs through which a group understands its world, influences the content of expert as well as everyday knowledge.
- Although medical knowledge may be seen as the product of scientific discovery involving direct engagement with the facts of the natural world, it has been built up within social and cultural contexts that influence its development.
- Rather than seeing medical advance as a linear process of progressive discovery, many philosophers and social scientists have argued that one theoretical paradigm will over time replace another and that these changing conceptual frameworks reflect the wider societal culture and prevailing modes of thought.
- Examples such as those presented in this chapter show that disease classifications and understandings of etiology are closely connected with other aspects of contemporary culture.
- Doctors are experts in the body of theory, evidence, and experience that comprises Western medical knowledge but must apply that knowledge in interaction with patients whose lay health beliefs about illness and its causes may not align with professional perspectives; thus the practice of medicine involves mediating between expert knowledge and lay belief systems influenced by local cultures.
- Medical knowledge is applied within the context of medical culture and the particular subcultural contexts of the clinical specialties and variegated locales in which practice takes place; studies have shown that sociocultural factors such as hierarchy, professional rivalry, and self-protective behavior impact upon medical practice.
- Medicine is both an art and a science and medical knowledge has both experiential and scientific components; professionals differ in their views about the relative importance of the clinical judgment of the individual doctor (rooted in traditional medical culture) and standardization based on codified, evidence-based knowledge.
- More generally there is a continuing tension between medicine as a practice-based discipline and medical science in the research laboratory, something which is the outcome of history and the way Western cultures segment domains of knowledge.

References

Atkinson P (1981) The clinical experience: the construction and reconstruction of medical reality. Gower, London
Atkinson P (1995) Medical talk and medical work. Sage, London
Austoker J (1988) A history of the Imperial Cancer Research Fund, 1902–1986. Oxford University Press, Oxford
Beeson P (1980) Changes in medical therapy during the past half century. Medicine 59:79–99
Blaxter M (1983) The causes of disease: women talking. Soc Sci Med 17:59–69
Blaxter M (2010) Health, 2nd edn. Polity, Cambridge
Bosk C (1979) Forgive and remember: managing medical failure. Chicago University Press, Chicago
Bowker GC, Star SL (1999) Sorting things out: classification and its consequences. MIT Press, Cambridge, MA
Chrisman NJ (1977) The health seeking process: An approach to the natural history of illness. Culture, Medicine and Psychiatry 1:351–377
Collyer FM (1996) Understanding ulcers: medical knowledge, social constructionism and helicobacter pylori. Ann Rev Health Soc Sci 6:1–39
Dubos R (1987) Mirage of health: utopias, progress and biological change (1st edn. 1959). Rutgers University Press, New Brunswick
Fabrega H (1974) Disease and social behavior: an interdisciplinary perspective. MIT Press, Cambridge, MA
Fadiman A (1997) The spirit catches you and you fall down: a Hmong child, her American doctors, and the collision of two cultures. Farrar, Straus and Giroux, New York
Gillies D (2005) Hempelian and Kuhnian approaches in the philosophy of medicine: the Semmelweis case. Stud Hist Philos Biol Biomed Sci 36:159–181
Greaves D (1998) What are heart attacks? Rethinking some aspects of medical knowledge. Med Healthc Philos 1:133–141
Hacking I (1996) The looping effects of human kinds. In: Sperber D, Premack D, Premack AJ (eds) Causal cognition: a multidisciplinary debate. Clarendon, Oxford, pp 351–383
Hai H (2009) Kuhn and the two cultures of Western and Chinese medicine. J Camb Stud 4:10–36
Harwood A (1971) The hot-cold theory of disease: implications for treatment of Puerto Rican patients. JAMA 216:1153–1158
Helman CG (1978) "Feed a cold, starve a fever" – folk models of infection in an English suburban community, and their relation to medical treatment. Cult Med Psychiatry 2:107–137
Helman CG (2007) Culture, health and illness, 5th edn. Hodder Arnold, London
Hewlett BS, Hewlett BL (2008) Ebola, culture, and politics: the anthropology of an emerging disease. Wadsworth, Belmont
Holmes D (2013) Mid Staffordshire scandal highlights NHS cultural crisis. Lancet 381:521–522
Hutton J (2012) Composite paradigms in medicine: analysing Gillies' claim of reclassification of disease without paradigm shift in the case of Helicobacter pylori. Stud Hist Philos Biol Biomed Sci 43:643–654
Inui TS, Yourtree EL, Williamson JW (1976) Improved outcomes in hypertension after physician tutorials: a controlled trial. Ann Intern Med 84:646–651
Jamous H, Peloille B (1970) Professions as self-perpetuating systems? Changes in the French university hospital system. In: Jackson JA (ed) Professions and professionalization: sociological studies 3. Cambridge University Press, Cambridge, pp 109–152
Jewson N (1976) The disappearance of the sick man from medical cosmology, 1770–1870. Sociology 10:225–244
King LS (1954) What is disease? Philos Sci 21:193–203
Kirmayer L, Young A (1998) Culture and somatization: clinical, epidemiological, and ethnographic perspectives. Psychosom Med 60:420–430

Kleinman A (1980) Patients and healers in the context of culture: an exploration of the borderland between anthropology, medicine, and psychiatry. University of California Press, Berkeley
Kohrt BA (2014) Somatization. In: Cockerham WC, Dingwall R, Quay S (eds) The Wiley-Blackwell encyclopedia of health, illness, behavior and society. Wiley, New York, pp 2248–2251
Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago
Latimer J (2013) The gene, the clinic, and the family: diagnosing dysmorphology, reviving medical dominance. Routledge, London.
Lupton D (2003) Medicine as culture: illness, disease and the body, 2nd edn. Sage, London
Lyttleton C (1993) Knowledge and meaning: the AIDS education campaign in rural northeast Thailand. Soc Sci Med 38:135–146
Nicolson M, McLaughlin C (1987) Social constructionism and medical sociology: a reply to M.R. Bury. Sociol Health Illn 9:107–126
Nordenfelt L (1990) Comments on Wulff's, Thung's and Lindahl's essays on the growth of medical knowledge, pp 121–130. In: ten Have H, Kimsma G, Spicker SF (eds) The growth of medical knowledge. Kluwer, Dordrecht, pp 121–130
Numerato D, Salvatore D, Fattor G (2012) The impact of management on medical professionalism: a review. Sociol Health Illn 34:626–644
Payer L (1988) Medicine and culture: varieties of treatment in the United States, England, West Germany, and France. Holt, New York
Peitzman SJ (1989) From dropsy to Bright's disease to end-stage renal disease. Milbank Q 67 (Suppl 1):16–32
Perpich JG (2004) The dawn of genomic and regenerative medicine: new paradigms for medicine, the public's health, and society. Technol Soc 26:405–414
Pill R, Stott NCH (1982) Concepts of illness causation and responsibility: some preliminary data from a sample of working class mothers. Soc Sci Med 16:43–52.
Richman JA, Jason LA (2001) Gender biases underlying the social construction of illness states: the case of chronic fatigue syndrome. Curr Sociol 49:15–29
Rosenberg CE (1989) Disease in history: frames and framers. Milbank Q 67(Suppl 1):1–15
Stainton Rogers W (1991) Explaining health and illness: an exploration of diversity. Wheatsheaf, London
Stern B (1941) Society and medical progress. Princeton University Press, Princeton
Stern B (1968) Social factors in medical progress (1st edn. 1927). AMS Publishers, New York
Strong PM (1984) Viewpoint: the academic encirclement of medicine. Sociol Health Illn 6:339–358
Thagard P (1998) Ulcers and bacteria I: discovery and acceptance. Stud Hist Philos Sci Part C Biol Biomed Sci 29:107–136
Thomas L (1971) The technology of medicine. NEJM 285:1366–1368
Verwey G (1990) Medicine, anthropology, and the human body. In: ten Have H, Kimsma G, Spicker SF (eds) The growth of medical knowledge. Kluwer, Dordrecht, pp 133–162
Watts AW (1989) The book: on the taboo against knowing who you are. Vintage Books, New York
White K (1991) The sociology of health and illness, trend report. Curr Sociol 39:1–40
Wright P, Treacher A (eds) (1982) The problem of medical knowledge: examining the social construction of medicine. Edinburgh University Press, Edinburgh
Young A (1997) The harmony of illusions: inventing post-traumatic stress disorder. Princeton University Press, Princeton

Hippocrates and the Hippocratic Tradition: Impact on Development of Medical Knowledge and Practice?

50

James A. Marcum

Contents

Abstract

What impact, if any, did Hippocrates and the Hippocratic tradition have on the development of medical knowledge and practice? For some, Hippocrates is the "Father of (Western or Modern) Medicine," and the Hippocratic tradition provides a framework for the development of contemporary medicine – especially a rational, scientific medicine. Hippocrates and the Hippocratic tradition are not only important in terms of the development of medical knowledge but also its practice, as exemplified by the Hippocratic oath. For others, modern medicine represents a rejection not so much of Hippocrates but only of the Hippocratic tradition, especially its vitalism and humoral theory of health and disease. In this chapter, the impact of Hippocrates and the Hippocratic tradition on the development of medical knowledge is explored first, followed by an examination of how they, especially the oath, shaped medical practice. The chapter concludes with a discussion of the lessons this exploration into Hippocrates and the Hippocratic tradition teach about the future of medical knowledge and practice.

J.A. Marcum (✉)
Department of Philosophy, Baylor University, Waco, TX, USA
e-mail: James_Marcum@baylor.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_82

Introduction

In this introductory section, it is briefly discussed who Hippocrates is and what the Hippocratic tradition is in order to provide the historical background necessary for the two main sections of the chapter. In the first section, it is explored what kind of impact Hippocrates' theory of medicine – especially as articulated in the Hippocratic tradition in terms of vitalism and humoralism – had on the development of medical knowledge. The impact's trajectory for that development is examined with respect to such diverse medical specialties, ranging from cardiology to spine surgery. In the second section, the impact of Hippocrates and the Hippocratic tradition on medical practice is explored, especially in terms of the Hippocratic oath. In a concluding section, the lessons this exploration into Hippocrates and the Hippocratic tradition teach about the future of medical knowledge and practice are discussed.

Who was Hippocrates? This question is difficult to answer, at best, because of limited resources on Hippocrates' personal life, medical practice, and literary output, especially by his peers. Unfortunately, the earliest extant biography of Hippocrates was not written until almost five centuries after his death by Soranus of Ephesus. Several other brief biographical works date from the tenth and twelfth centuries CE. "The material about Hippocrates was invented," claims Jody Pinault, "growing slowly during the Hellenistic period" (1992, 1). Indeed, the consensus today is that little is known about the historical Hippocrates in terms of his life, career, and writings (Jouanna 1999; King 2001; Levine 1971; Nutton 2013; Scarborough 1997; Schiefsky 2005; Smith 1990; Temkin 1991). "The commingling of legend, myth, and hagiography in the biography of Hippocrates attests to the fact," concludes Steven Miles, "that almost nothing is known about him" (2004, 28).

Traditionally, the birth of Hippocrates is assigned to the year 460 BCE, during the 80th Olympiad, on the Greek island of Cos. Hippocrates' father was Heraclides, who was also a physician and whose lineage is reputed to include Asclepius – the divine Greek physician. Hippocrates' mother was Phaenarete, who was supposedly a descendent of Hercules. Hippocrates lived during the classical or golden period of Greek culture under the aegis of the "first citizen of Athens," Pericles. Hippocrates' father was responsible for his early medical education, but, after his parent's death, he left Cos to further his medical education. According to Soranus of Cos, Hippocrates first traveled to Thessaly in obedience to a command given in a dream. When the plague that infested Athens erupted during the second year of the Peloponnesian war (430 BCE), Pericles invited Hippocrates to save the city's inhabitants from it. Hippocrates was successful, and a golden wreath was bestowed upon him, and he was made a citizen of Athens. He eventually returned to Cos where he was influential in the Coan school of medicine. He is alleged to have died in Larissa, although the exact date is uncertain.

The historical fact of Hippocrates' existence, however, is not in contention among scholars. Indeed, such contemporaries as Plato, Aristotle, and Aristotle's student Menon refer to him in their writings (Longrigg 1998). For example, Plato informs the reader that Hippocrates, a "member of the Asclepiadae," collected fees

from students for instruction in the medical arts (*Protagoras* 311b-c) and also discusses Hippocrates' clinical method in terms of assessing the "nature of the whole" with respect to a phenomenon's simplicity or complexity (*Phaedrus* 270c-d). Aristotle called him a great physician, although he found him small in stature (*Politics* 1326a13-16). Finally, Menon discusses Hippocrates' medical theory concerning the etiology of disease in terms of breaths or *physai* (*Anonymus Londinensis* 5.35–7.40). What is in contention among scholars, however, is the myth surrounding Hippocrates' personal and professional life, especially when the literary works attributed to him began to escalate during the Hellenistic period while the medical literary works were being compiled in Alexandria – , which leads to the next question.

What is the Hippocratic tradition? Briefly, the tradition represents the impact of Hippocrates on the development of medical knowledge and practice (Joly 1983; Mansfeld 1983; Michell 2010; Smith 1979). As Wesley Smith (1979) demonstrates in an analysis of the historical trajectory of the tradition, medical communities from every age – since Hippocrates practiced medicine – have been influenced by him. An example from the modern Hippocratic tradition is Thomas Sydenham, who is known as the "English Hippocrates." Sydenham championed Hippocratic observation to cure disease in contrast to theoretical constructs to explain it. Of course, Galen is an important figure in the development of the Hippocratic tradition, as are other ancient personages – especially from the Hellenistic period when the Hippocratic corpus was initially being collected in the fourth century BCE. Smith (1979) points in particular to the pseudepigrapha, a collection of letters and speeches, as chiefly responsible for establishing the mythical dimensions of the Hippocratic tradition. Throughout medical history, Hippocrates has been used to justify the current approach to medical knowledge and practice, and no single Hippocratic tradition captures its nature, but rather there are multiple traditions based on a particular historical era and country.

A major problem associated with the Hippocratic tradition – often called the Hippocratic question – is determining which, if any, of the books forming the Hippocratic corpus were genuinely written by Hippocrates. To date, over 60 volumes comprise the corpus (Corpus Medicorum Graecorum 2015). In a critical analysis of the question, Geoffrey Lloyd (1975) tackles both the external and internal evidence concerning various attempts to ascribe Hippocratic authorship to selected works in the corpus. With respect to external evidence, Lloyd concludes that it is simply too self-serving for the author citing Hippocrates as author to provide a standard by which to evaluate whether a book from the corpus was indeed authored by Hippocrates. Moreover, he acknowledges that a collection of medical treatises is extant by the early third century BCE but that the commentaries forming the corpus obfuscate rather than clarify Hippocratic authorship. With respect to the internal evidence, Lloyd concludes that it too cannot supply a standard by which to judge the authenticity of a book's Hippocratic authorship because of the "heterogeneity" of the doctrines expounded in the corpus – to the extent that even contradictions surface within it. "It may be that some of Hippocrates' work has come down to us in the Corpus," concludes Lloyd, "but we cannot now prove this,

nor determine which his work is" (1975, 189). This is certainly a sober warning to keep in mind as the impact of Hippocrates and the Hippocratic tradition on the development of medical knowledge and practice is explored in the next two sections.

The Development of Medical Knowledge

In this section, the impact of Hippocrates and the Hippocratic tradition on the development of medical knowledge is discussed. To that end, the notion of the nature of medicine from a Hippocratic perspective is examined initially – especially as that notion diverges from a traditional or pre-Hippocratic understanding of medicine in terms of the role of religion and the supernatural. Next, the commitment of Hippocratic medicine to the notions of holism and vitalism, which it shared to some extent with pre-Hippocratic medicine, is examined. Then, the commitment of Hippocratic medicine to a rationalistic and naturalistic approach to medical knowledge is discussed, especially with respect to pre-Socratic philosophy and the notion of humoralism. Finally, the section concludes with a brief overview of the impact of Hippocratic medicine on contemporary medical specialties and terminology, ranging from cardiology to spine surgery.

What is medicine, from a Hippocratic perspective? To answer that question requires a brief discussion of pre-Hippocratic or traditional Greek medicine. Pre-Hippocratic medicine was intimately linked to religion with respect to the etiology and treatment of illness (Jouanna 2012; Longrigg 1993). The classic example is the sacred disease, epilepsy. According to traditional Greek medicine, the cause of epilepsy was divine in origin as was the treatment, especially in terms of incantations and purifications. In contrast, the author of the Hippocratic text, *On the Sacred Disease*, claims that epilepsy is not divinely caused, but rather its cause was the result of blood flow blocked within the brain via the accumulation of phlegm within the vessels. The author goes on to chide those who claim a divine cause for the disease because they are ignorant of the disease's etiology. However, Hippocratic medicine was not inimical to religion, and it incorporated religion into the care and treatment of the patient. What it excluded from medical practice was the magic and superstition (Hankinson 1998; Martin 2004).

Although pre-Hippocratic medicine differs significantly from Hippocratic medicine vis-à-vis religious influence on medical knowledge and practice, they both shared a commitment to a notion of the whole – or what is contemporarily called holism (Smuts 1926) – in terms of understanding health and illness (Nutton 2013; Pitman 2006). This commitment to holism involves imaging the body as a whole (*holon*) or, as Jacques Jouanna articulates it, "a copy of the Whole" (1999, 276). Besides the wholeness of the body, the patient is also embedded within an environmental context. For example, the author of *On Airs, Waters, and Places* counsels the physician to take note of changes in the seasons and stars, when considering a disease's etiology. In other words, the patient is a microcosm functioning within a

macrocosm, and changes within that macrocosm can adversely affect the microcosm. As James Gordon explains,

> Hippocrates and the tradition out of which modern biomedicine has grown emphasized the environmental causes and treatment of illness; the etiological and therapeutic importance of psychological factors, nutrition and life-style; the interdependence of mind, body and spirit, and the need for harmony between an individual and his social milieu and natural environment. (1982, 547)

Thus, Hippocratic holism – especially as it is used to justify contemporary holistic medicine – is comprehensive in terms of embedding the whole patient within social and environmental contexts to address illness.

In addition, both pre-Hippocratic and Hippocratic medicine also shared a commitment to a vital principle or to a notion of vitalism for understanding life and living processes, especially health and illness. Although the term vitalism is not introduced until the late eighteenth century, the notion itself is prevalent throughout its complex history (Myers 1900; Wheeler 1939). Within the Hippocratic corpus, the vital principle is articulated with respect to a variety of terms for air, such as *aer*, *anemos*, *phusa*, and *pneuma*, with *pneuma* emerging as the chief term for the principle that animates the body (Frixione 2012; Lloyd 2007). According to the author of *On Breaths*, for instance, the breath of life involves the transformation of the breathed air or *aer* into the wind or *pneuma* that blows or circulates throughout the body and thereby animates it. Health and disease, then, depend upon the quality of this vital principle. As Charles Cumston summarizes Hippocratic vitalism, "Hippocrates admits without hesitation that life is a principle unknown to man, the necessity of which imposes itself as soon as one considers the unity, finality and harmonious plan of the vital phenomena" (1904, 314–315). It is the principle or *archeus* that organizes and makes life possible and that is at root of both health and illness. Finally, although the Hippocratic vital principle is not fully developed – or "incomplete" as Cumston acknowledges – it is often referenced with respect to further historical development of vitalism and neovitalism (Normandin and Wolfe 2013).

As Hippocratic medicine matured, especially in terms of its apex in Hellenistic medicine, it takes what is often called a revolutionary turn in terms of rationalism and naturalism (Boylan 2005; Heidel 1941; Langholf 1990; Longrigg 1998; Scarborough 2002; Schiefsky 2005; Sullivan 1996). As James Longrigg summarizes this turn of ancient Greek medicine,

> One of the most impressive contributions of the ancient Greeks to Western culture was their invention of rational medicine. It was the Greeks who first evolved rational systems of medicine for the most part free from magical and religious elements and based upon natural causes. (1993, 1)

Although Longrigg admits that incorrect rationalistic theories are no better than irrational religious superstitions, he argues that the former can correct itself while the latter cannot. Longrigg goes on to explain that Hippocratic medicine was

dependent on Ionian rationalism, which endeavored to explain phenomena in naturalistic terms. For example, Empedocles proposed the elements of earth, water, fire, and air to account for the composition of complex phenomena like life and tissue (Solmsen 1950). Besides the four elements, four qualities – heat, cold, dry, and wet – were also invoked in early Greek natural philosophy to explain phenomena, including health and disease. Hippocratic medicine qua rational, then, sought to identify the invisible causes of health and disease, and no better theory represented that approach than humoralism.

Within the Hippocratic corpus, there are a variety of humoral theories to account for health and disease, both in terms of the number and types of humors. For example, the author of *On the Sacred Disease* posited two humors – phlegm and bile. Eventually, the mature Hippocratic humoral theory, especially as Galen developed it later, posited four humors – blood, phlegm, yellow bile, and black bile (Balzer and Eleftheriadis 1991; Lonie 1981; Nutton 1993). The balance of these humors is responsible for a person's health, while an imbalance in them results in disease (Fig. 1). Moreover, the four humors were also associated with other approaches for explaining natural phenomena. The author of *On the Nature of Man*, for instance, correlated the humors with the four qualities, as well as with the four seasons. It was these correlations, as Vivian Nutton explains, that made the Hippocratic humoral theory attractive such that it "became the dominant medical philosophy" (1993, 287). Indeed, the theory dominated medical knowledge and practice for over a millennium. Finally, the origin of the contemporary notion of homeostasis is often located to the Hippocratic humoral theory in terms of the balance of humors (Bujalkova et al. 2001; Kontopoulou and Marketos 2002).

Besides homeostasis, many current medical specialties claim Hippocratic origins. For example, Hippocrates is called the "Father of clinical nephrology" (Eknoyan 1988), and he is credited with the foundations of modern cardiology and championed as the "Father of circulation" in contrast to Harvey (Cheng 2000, 2001). Moreover, Hippocratic origins are claimed for various surgical specialties,

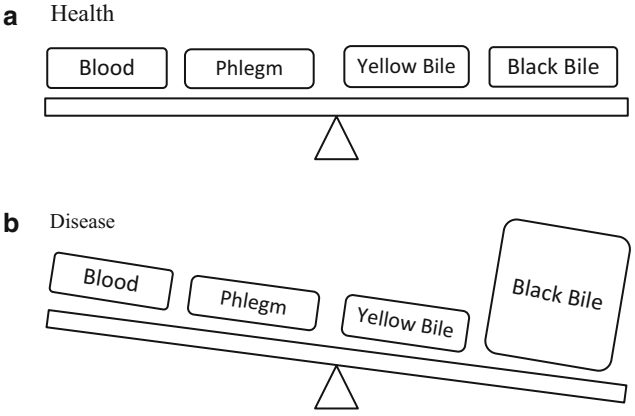


Fig. 1 Humoral pathology (**a**) health (**b**) disease

such as neurosurgery (Chang et al. 2007) and spinal surgery (Marketos and Skiadas 1999a). Interestingly, Hippocratic medicine has also been recognized as foundational for the emergence of genomic medicine:

> Genomic medicine's viewpoints on the biological foundations of human nature, the conceptualization of health and disease, the determinants of individuality in disease predisposition, and the personalized approach to diagnosis, prognosis, and treatment represent a revival of methodological and humanitarian Hippocratic principles. (Sykiotis et al. 2006, 181)

Finally, many contemporary medical terms, such as edema, ileus, and thorax; diseases, such as arthritis, eclampsia, and pneumonia; and notions, such as anesthesia and analgesia, are traced to their Hippocratic roots (Astyrakaki et al. 2010; Marketos and Skiadas 1999b; Yapijakis 2009).

The Development of Medical Practice

Besides medical knowledge, the Hippocratic tradition has had a major impact on the practice of medicine – particularly in terms of its clinical and ethical dimensions. With respect to clinical practice, the author of *Epidemics I* (▶ Chap. 9, "Goals of Medicine") identifies three components to the practice of medicine. The first is the disease, which can be explained in naturalistic terms; the second is the patient, who has the disease and represents a psychosomatic whole; and, the third is the physician, who endeavors to assist nature in helping the patient recover from the disease. These three components form the Hippocratic triangle (Fig. 2; Duffin 2005; Marketos and Skiadas 1999b). Although knowledge of the disease and its etiology is important in treating the patient, so is knowledge of the patient, which has a direct impact on how the physician treats the patient and which is vital for understanding the nature of medicine itself and its goals. For the Hippocratic physician, as well as for physicians throughout history, the Hippocratic oath provides an unsurpassed ethical means – the Hippocratic ethic – for discharging the duties of the profession associated with the interactions among the triangle's components, as well as for defining the profession and its duties to patients and society.

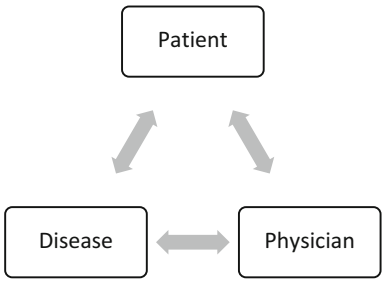


Fig. 2 The Hippocratic triangle

If the Hippocratic corpus is contentious among scholars, the Hippocratic oath is even more so – especially in terms of the oath's origins and authorship (Davey 2001; Miles 2004). For example, the date for the oath's composition varies from the sixth century BCE to the first century CE – although there appears to be consensus that it was composed in the fifth century BCE. Moreover, even though it is part of the Hippocratic corpus, Hippocrates is not considered its author. In fact, Ludwig Edelstein (1967), based on an analysis of the oath's textual content and its historical context, claimed that the oath represents a Pythagorean document. In "On Second Thoughts" on Edelstein's thesis, however, Owsei Temkin challenged Edelstein's analysis and thesis, although he conceded "Pythagorean influences might well have played a role behind the oath" (2002, 4). Finally, Plinio Prioreschi (1995) charged that many of the prohibitions Edelstein found in the oath, which Edelstein claimed were associated with the Pythagoreans, are also found elsewhere in ancient Greek culture and not necessarily unique to the Pythagoreans.

The Hippocratic oath consists of two main parts, after invoking the gods (Apollo and Asclepius) and goddesses (Hygeia and Panacea) associated with medicine as witnesses (Edelstein 1967; Nutton 2013). The first part concerns the duties of the physician to the profession, particularly in terms of honoring one's teachers, and the transmission of medical knowledge between generations. Specifically, the oath demands the inductee

> [t]o hold him who has taught me this art as equal to my parents and to live my life in partnership with him, and if he is in need of money to give him a share of mine, and to regard his offspring as equal to my brothers in male lineage and to teach them this art—if they desire to learn it—without fee and covenant; to give a share of precepts and oral instruction and all the other learning to my sons and to the sons of him who has instructed me and to pupils who have signed the covenant and have taken an oath according to the medical law, but no one else. (Edelstein 1967, 6)

Thus, "the Oath," as Lisa Keränen summarizes this part, "works to unite members of the profession into a tight-knit community" (2001, 59). In other words, as some commentators observe, it serves to demarcate the genuine physician from quacks and charlatans.

The second part of the Hippocratic oath concerns the covenant of the physician with the patient and contains at least half-dozen injunctions and consequences. The first pertains to a therapeutic injunction, "I will apply dietetic measures for the benefit of the sick according to my ability and judgment," which is associated with an ethical standard of patient non-maleficence, "I will keep them from harm and injustice" (Edelstein 1967, 6). The next involves a deep regard for human life in terms of prohibiting suicide or euthanasia, "I will neither give a deadly drug to anybody if asked for it, nor will I make a suggestion to this effect," and abortion, "I will not give to a woman an abortive remedy" (Edelstein 1967, 6). These injunctions are important for the Hippocratic physician, who in "purity and holiness" endeavors to "guard my life and my art" (Edelstein 1967, 6). As for surgery, "I will not use the knife, not even on sufferers from stone, but will withdraw in favor of such men as are engaged in this work" (Edelstein 1967, 6). In other words, as some

commentators have noted, Hippocratic physicians do not presume to practice what they have not been trained to do.

Again, the Hippocratic oath returns to an injunction concerning patient non-maleficence, in terms of not taking advantage of the patient's vulnerability:

> Whatever houses I may visit, I will come for the benefit of the sick, remaining free of all intentional injustice, of all mischief and in particular of sexual relations with both female and male persons, be they free or slaves. (Edelstein 1967, 6)

The following injunction concerns confidentiality, "What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself holding such things shameful to be spoken about" (Edelstein 1967, 6). Finally, the oath concludes with the consequences of fulfilling or transgressing the oath:

> If I fulfill this oath and do not violate it, may it be granted to me to enjoy life and art, being honored with fame among all men for all time to come; if I transgress it and swear falsely, may the opposite of all this be my lot. (Edelstein 1967, 6)

In sum, the oath for many physicians has functioned "as a powerful reminder and declaration that we are all part of something infinitely larger, older, and more important than a particular era, specialty, or institution" (Markel 2014, 29).

The Hippocratic oath contains many injunctions and statements, however, which are considered problematic vis-à-vis contemporary medical values and practice (Morgenstern 2008). For example, swearing to Greek gods is certainly awkward to those who believe in other religions. Another problematic part of the oath is teaching the art of medicine only to males and excluding females, even though two goddesses are invoked as witnesses. Other challenging parts of the oath include prohibition of suicide or euthanasia and abortion, albeit only for pessary. Moreover, the prohibition of the knife or surgery appears problematic, even though others defend it as physicians recognizing their limitations (Antoniou et al. 2010). In sum, Eugene Robin and Robert McCauley (1995) argue that the oath represents a "cultural lag" in which the contemporary medical community has failed to incorporate current changes in society, especially with respect to values and scientific innovations, into the oath.

Robert Veatch (1984) identifies another serious problem with the Hippocratic oath, which leads him to pronounce the death of any possible ethic based on it. The problem stems from the Hippocratic principle concerning the physician's promise to labor "for the benefit of the sick according to my ability and judgment" and to "keep them from harm or injustice." Although Veatch acknowledges the principle that appears praiseworthy prima facie, upon further reflection, it is problematic with respect to three points. The first is paternalism in that the patient's perspective of what is beneficial vis-à-vis therapeutic options is not taken into consideration. The next is individualism in that only the benefit of the individual patient is considered and not that of the larger community in which both the patient and physician reside.

And, the final is consequentialism in that the oath is only concerned with the consequences of a physician's actions and not with their inherent morality. Veatch then argues for a shift from a Hippocratic ethic based on paternalism, individualism, and consequentialism to an ethic based on "principles such as autonomy, truth-telling, avoiding killing and justice" (1984, 48).

Given these problematic injunctions and statements of the Hippocratic oath, the contemporary response to it is often controversial and ranges along a spectrum of those who dismiss it and propose revised or alternative oaths to those who defend and champion it. For those who dismiss and replace it with a revised or an alternative oath, the Hippocratic oath represents an archaic and irrelevant document, especially in terms of contemporary medical values (Hurwitz and Richardson 1997; Meffert 2009; Robin and McCauley 1995; Rosalki 1993; Wagley 1987). Louis Lasagna (1964), dean of Tufts Medical College, introduced one of the more well-known alternative oaths. In it, Lasagna stresses the humanity not only of the patient but also of the physician, without compromising the advances of scientific medicine. For example, one statement claims, "I will remember that there is art to medicine as well as science, and that warmth, sympathy, and understanding may outweigh the surgeon's knife or the chemist's drug." Finally, Keränen (2001) argues that such drastic revision of the oath is warranted since the oath represents an Aristotelian epideictic rhetoric, which operates ceremonially to construct especially a moral community. Incorporating contemporary moral and social values into a revised Hippocratic oath, then, is critical for the "long-term function of moving its audience from core values rooted in the past to principled action in the future" (Keränen 2001, 67).

For others who defend the Hippocratic oath and champion it, the oath is still relevant vis-à-vis contemporary medical values and practice, and it represents a means particularly for defining medical professionalism (Heubel 2015; Kravitz 1984; Markel 2014). For example, Spyros Marketos and colleagues criticize the revised and alternative oaths as often too legalistic and argue that the original oath's "true meaning, overall respect for the patient, can be accommodated in different cultures and historical periods" (Marketos et al. 1996, 101). Indeed, Lycurgus Davey goes so far as to argue that

> Hippocrates' oath is as fitting an ideal today as it was 2500 years ago. Let not the quaintness of its language nor the terseness of its terms obscure or distract from the highly principled guidance offered by the oath of Hippocrates. Only with his guidance can we hope to restore American medicine's golden age. (2001, 564)

Indeed, for some, the oath also serves as the basis for a robust medical ethics – in spite of Veatch's pronouncement of the Hippocratic ethic's demise (Orr et al. 1997). For example, Edmund Pellegrino claims that the Hippocratic ethic "is marked by a unique combination of humanistic concern and practical wisdom admirably suited to the physician's tasks in society" (2000, 42). Based on this characteristic of the ethic, Pellegrino then expands it in terms of its axiology to include ethical issues that face contemporary medicine, such as patient participation

in treatment, physician competence and duties, and the institutionalization of medicine. In sum, the oath serves as an ideal of who the physician is and of how the medical profession should provide the best possible healthcare.

Hippocratic Lessons for Contemporary Medicine

The present exploration of Hippocrates and the Hippocratic tradition teaches several important lessons not only about the development of medical knowledge and practice but also about their future advancement. Although there are a number of important lessons that can be derived from them, especially for contemporary medicine (Fabre 1997; Marketos 1993), only four are discussed in this final section. The first concerns medical knowledge and its rational approach to explaining disease as well as its relationship to alternative epistemic approaches of medicine; the second relates to medical practice particularly in a highly technologized medicine and the suffering associated with the patient illness experience; the third involves medical professionalism with respect to the duties required of the physician; and, the fourth pertains to medical philosophy in terms of the ethical challenges facing contemporary medical care and the role of virtue theory in addressing those challenges. The section concludes with a brief discussion concerning the role of the Hippocratic spirit for medicine in the twenty-first century (Daikos 2003; Helidonis and Prokopakis 2001).

With respect to medical knowledge, the Hippocratic tradition of explaining disease rationally, particularly in terms of their natural causes, is certainly an important lesson for contemporary medicine. Religion and spirituality may be important dimensions of a patient's illness experience and certainly need to be addressed (Koenig et al. 2012). However, the material and physical mechanisms underlying the etiology of disease often – but not always – demand priority. Although medicine should be based on the best empirical evidence available at the time, the Hippocratic tradition teaches a certain level of humility concerning the veracity or uncertainty of medical knowledge. After all, both vital spirits and humors are no longer viable for explaining health and disease or for treating patients. But, on the other hand, the epistemic humility that Hippocratic medicine teaches should provide sufficient motivation to support and continue the biomedical research needed to develop the medical knowledge required to provide the best technical care for treating patients. Moreover, medical research should be open-minded, what Grant Gillett (2004) calls a Hippocratic attitude, toward alternative epistemic approaches to investigating and explaining disease. Finally, Hippocratic medicine teaches contemporary practitioners how the limits to medical knowledge and its application to the clinic can help to avoid medical futility (Jecker 1991).

With respect to medical practice, the Hippocratic tradition teaches that the emphasis should be not simply on the disease but on the patient who is experiencing the disease. In other words, medical practice involves the patient and the disease, as well as the efforts of the medical professional to treat the patient, i.e., the Hippocratic triangle (Fig. 2). It also teaches that medical professionals must be fastidious

in their integrity toward the goal of relieving or attenuating suffering associated with illness. Unfortunately, a triumphal, positivist perspective of medical progress in which disease can be cured – either through pharmaceutical or surgical therapeutics – or even prevented, through genomic counseling or engineering, often blinds the healthcare professional to outcomes of greater patient harm and suffering and of a poorer quality of life for the patient. In other words, the treatment is often more painful and injurious than the disease – even to the extent of death. Progressive metaphysics with a verisimilitude epistemology – or what is generally called scientism – can be detrimental to the medicine's goal of reducing patient suffering and not adding to it (Cassell 2004). Moreover, the lesson that Hippocratic medicine teaches is a holism in terms of treating the whole patient within a given context. The patient does not simply have a disease but is living with it, and the medical professional must take into consideration not only the patient but also the context in which the patient is living with the disease.

With respect to medical professionalism, the Hippocratic tradition, especially in terms of its oath, teaches that to define medical professionalism requires a moral sense of duties and obligations not only to its patients but also to the profession itself (Coulehan 2006). As Daniel Sulmasy (1999) argues, an oath in general involves performative statements that carry considerable moral weight and consequences, compared to promises or codes. Given the drastic changes in the medical profession since the origination of the oath, however, whether the oath can serve any role in framing contemporary medical professionalism is questionable. For example, Friedrich Heubel (2015) argues that the "Charter on Medical Professionalism" represents a better means for defining contemporary medical professionalism than the Hippocratic oath. Moreover, as Fabrice Jotterand argues,

> the resources for a better understanding of medical professionalism lie not in the Hippocratic Oath, tradition, or ethos in and of themselves. Rather, it must be found in a philosophy of medicine that explores the values internal to medicine, thus providing a medical-moral philosophy so as to be able to resist the deformation of medical professionalism by bioethics, biopolitics, and governmental regulation. (2005, 108–109)

Although the Hippocratic oath cannot define contemporary medical professionalism, it can serve as a heuristic guide – according to Jotterand – toward that end. Ultimately, what is needed is not a return to Hippocratic medicine but rather a contemporary medical philosophy to define clinical medicine and its professionalism with respect to medical knowledge and practice.

With respect to medical philosophy, the Hippocratic oath and tradition teach that the virtues are important for performing medical duties and obligations, especially when faced with ethical challenges, and that they can provide the basis for a robust medical philosophy to guide contemporary medicine, when confronted with ethical challenges. Although virtue ethics waned historically, there has been a resurgence in its impact especially on medical practice and professionalism:

> As it stands, the Hippocratic Oath can no longer be viewed as *the* action-guiding inspiration of current medical practice. However, the last word has not yet been said...To only apply the rules—bottom-line ethics—is not a solid ethical foundation. What is needed is to focus on the integrity, consistency, and excellence of character in the physician-patient relationship. Virtue ethics is appealing for its ability to provide a normative basis to the values internal to medicine. (Ogunbanjo and van Bogaert 2009, 31)

Although the Hippocratic oath might not be "*the* action-guiding inspiration" for a contemporary medical ethics, it still provides motivation and inspiration to develop a robust medical philosophy, especially with respect to the Hippocratic virtues. The Hippocratic virtues can be divided into two major sets (Berry 1997). The first pertains to "the virtues of expertise and skill necessary to accomplish curing – the fundamental purpose of the healing art – including the virtues of diligence, carefulness, conscientiousness, and the like" (Berry 1997, 412). The other set involves "virtues necessary to caring for the patient – kindness, sympathy, loyalty, and the like" (Berry 1997, 412).

For contemporary medicine, Pellegrino (1995) has identified several virtues representing both sets of Hippocratic virtues, such as benevolence, compassion and caring, justice, and prudence, which could assist in constituting a comprehensive and normative foundation for medical ethics. But, he claims that although necessary, these virtues are insufficient for the task. In addition, Pellegrino embeds these virtues within a broader moral framework to generate a robust medical philosophy to guide contemporary medical practice. Specifically, he distinguishes four elements involved in a moral act – the agent performing the act, the act itself, the circumstances under which the act is performed, and the act's consequences. In this framework, he locates various modern moral theories. Thus, virtue theory is associated with the moral agent; deontology with the act; "particularizing theories," such as situational ethics, with the circumstances surrounding the act; and teleological theories with the act's consequences. He argues that such a framework bodes well for medicine, especially in terms with the telos or goal of providing quality medical care. He concludes, however:

> Today's challenge is not how to demonstrate the superiority of one normative theory over the other, but rather how to relate each to the other in a matrix that does justice to each and assigns to each its proper normative force. (Pellegrino 1995, 273)

Unfortunately, this challenge still confronts contemporary medicine.

In conclusion, Hippocrates and the Hippocratic tradition have served as an ideal or symbol for both the practicing physician and for the institution of medicine as a profession, especially in terms of motivating medicine to provide the patient with the best medical care possible (Cantor 2002; Scarborough 2002; Tullis 2004). As John Fabre notes:

> The very nature of medicine is such that, unless it is firmly based on idealistic foundations, on notions of altruism, love, and so on, it can rapidly degenerate into a squalid business.

> The Hippocratic doctors recognized this and they were not shy, as we are today, to preach idealism. (1998, 162)

Indeed, the way toward resolving many of the issues facing modern medicine, from ethical to political to practical, would benefit from the Hippocratic ideal of a healthcare system that truly cares for the patient. In other words, the Hippocratic spirit continues to animate efforts to develop medicine as authentically human. As George Daikos concludes concerning this spirit, "The humane spirit is one of our great Hippocratic heritages" (2003, 188).

Definitions of Key Terms

Hippocratic corpus	A collection of over 60 treatises on various medical topics that were written by Hippocrates' followers and assembled in Alexandria.
Hippocratic oath	A vow consisting of assurances made concerning the behavior of physicians toward benefiting and not harming one another and their patients.
Hippocratic question	The question concerning which treatises, if any, within the Hippocratic corpus were written by Hippocrates.
Hippocratic tradition	The impact of Hippocrates and his followers on the development of medical knowledge and practice throughout history.
Holism	The notion that the properties of the whole are greater than the arithmetic sum of the properties of the parts making up the whole.
Humoralism	A theory of health and disease based on the balance of the four humors, blood, phlegm, yellow bile, and black bile.
Vitalism	The notion that life is the result of an external force that animates the body.

Summary Points

- Hippocrates lived and practiced medicine during the golden age of ancient Greece.
- The Hippocratic corpus is a collection of medical treatises written by Hippocrates' followers and compiled in Alexandria during the Hellenistic period.
- Hippocratic medicine involved a shift from the religious and superstitious approach to health and disease to the rational and natural explanation of them.
- Hippocratic vitalism pertains to the pneuma that circulates throughout the body, thereby animating it.
- Hippocratic humoralism is the medical theory to describe health and disease in terms of the balance of four humors: blood, phlegm, yellow bile, and black bile (Fig. 1).

- The Hippocratic triangle of medical practice comprises the interaction of the patient, the disease, and the physician (Fig. 2).
- The Hippocratic oath consists of two main parts describing the obligation of physicians to the welfare of both the medical profession and its patients.
- The Hippocratic spirit represents the inspiration and motivation for developing a competent and caring healthcare system throughout history.

References

Antoniou SA, Antoniou GA, Granderath FA, Mavroforou A, Giannoukas AD, Antoniou AI (2010) Reflections of the Hippocratic oath in modern medicine. World J Surg 34:3075–3079

Astyrakaki E, Papaioannou A, Askitopoulou H (2010) References to anesthesia, pain, and analgesia in the Hippocratic Collection. Anesth Analg 110:188–194

Balzer W, Eleftheriadis A (1991) A reconstruction of the Hippocratic humoral theory of health. J Gen Phil Sci 22:207–227

Berry RM (1997) The genetic revolution and the physician's duty of confidentiality: the role of the old Hippocratic virtues in the regulation of the new genetic intimacy. J Leg Med 18:401–441

Boylan M (2005) Hippocrates (c. 450-c. 380 B.C.E.). In: Internet encyclopedia of philosophy. Available via DIALOG. http://www.iep.utm.edu/hippocra/. Accessed 05 July 2015

Bujalkova M, Straka S, Jureckova A (2001) Hippocrates' humoral pathology in nowaday's reflections. Bratisl Lek Listy 102:489–492

Cantor D (ed) (2002) Reinventing Hippocrates. Ashgate, Burlington

Cassell EJ (2004) The nature of suffering and the goals of medicine, 2nd edn. Oxford University Press, New York

Chang A, Lad EM, Lad SP (2007) Hippocrates' influence on the origins of neurosurgery. Neurosurg Focus 23:E9

Cheng TO (2000) Hippocrates and cardiology. Am Heart J 141:173–183

Cheng TO (2001) The true father of circulation: Harvey or Hippocrates? Ann Thorac Surg 71:1399–1404

Corpus Medicorum Graecorum (2015) Available via DIALOG. http://cmg.bbaw.de/. Accessed 07 July 2015

Coulehan J (2006) You say self-interest, I say altruism. In: Wear D, Aultman JM (eds) Professionalism in medicine. Springer, New York, pp 103–127

Cumston CG (1904) The phenomenism of Hippocrates. Med Lib Hist J 2:307–317

Daikos GK (2003) The Hippocratic spirit. Hormones (Athens) 2:186–188

Davey LM (2001) The oath of Hippocrates: an historical review. Neurosurgery 49:554–566

Duffin J (2005) Lovers and livers: disease concepts in history. University of Toronto Press, Toronto

Edelstein L (1967) Ancient medicine. Johns Hopkins University Press, Baltimore

Eknoyan G (1988) Origins of nephrology: Hippocrates, the father of clinical nephrology. Am J Nephrol 8:498–507

Fabre JW (1997) The Hippocratic doctor: ancient lessons for the modern world. Royal Society of Medicine Press, London

Fabre J (1998) Modern medicine and the Hippocratic doctors of ancient Greece. J R Soc Med 91:161–163

Frixione E (2012) Pneuma – fire interactions in Hippocratic physiology. J Hist Med Allied Sci 68:505–528

Gillett G (2004) Clinical medicine and the quest for certainty. Soc Sci Med 58:727–738

Gordon JS (1982) Holistic medicine: advances and shortcomings. West J Med 136:546–551

Hankinson RJ (1998) Magic, religion and science: divine and human in the Hippocratic corpus. Apeiron 31:1–34

Heidel AW (1941) Hippocratic medicine. Its spirit and method. Columbia University Press, New York
Helidonis ES, Prokopakis EP (2001) The contribution of Hippocratic oath in third millennium medical practice. Am J Otolaryngol 22:303–305
Heubel F (2015) The "soul of professionalism" in the Hippocratic oath and today. Med Health Care Philos 18:185–194
Hurwitz B, Richardson R (1997) Swearing to care: the resurgence in medical oaths. BMJ 315:1671–1674
Jecker NS (1991) Knowing when to stop: the limits of medicine. Hastings Cent Rep 21:5–8
Joly R (1983) Hippocrates and the school of Cos. In: Ruse M (ed) Nature animated. Springer, New York, pp 29–47
Jotterand F (2005) The Hippocratic oath and contemporary medicine: dialectic between past ideals and present reality? J Med Phil 30:107–128
Jouanna J (1999) Hippocrates (trans: DeBevoise MB). Johns Hopkins University Press, Baltimore
Jouanna J (2012) Greek medicine from Hippocrates to Galen (trans: Allies N). Brill, Leiden
Keränen L (2001) The Hippocratic oath as epideictic rhetoric: reanimating medicine's past for its future. J Med Hum 22:55–68
King H (2001) Greek and Roman medicine. Bristol Classical Press, London
Koenig H, King D, Carson VB (eds) (2012) Handbook of religion and health, 2nd edn. Oxford University Press, New York
Kontopoulou TD, Marketos SG (2002) Homeostasis: the ancient Greek origin of a modern scientific principle. Hormones 1:124–125
Kravitz R (1984) Why the Hippocratic oath is worth preserving. Pharos Alpha Omega Alpha Honor Med Soc 47:37–39
Langholf V (1990) Medical theories in Hippocrates. De Gruyter, Berlin
Lasagna L (1964) Would Hippocrates rewrite his oath? N Y Times Mag 11:40–43
Levine EB (1971) Hippocrates. Twayne, New York
Lloyd GER (1975) The Hippocratic question. Classical Quart (ns) 25:171–192
Lloyd G (2007) Pneuma between body and soul. J Roy Soc Anthro Inst (ns) 13:S135–S146
Longrigg J (1993) Greek rational medicine: philosophy and medicine from Alcmaeon to the Alexandrians. Routledge, New York
Longrigg J (1998) Greek medicine from the hero to the Hellenistic age: a source book. Routledge, New York
Lonie IM (1981) The Hippocratic treatises "On generation", "On the nature of the child", "Diseases IV": a commentary. Walter de Gruyter, Berlin
Mansfeld J (1983) The historical Hippocrates and the origins of scientific medicine. In: Ruse M (ed) Nature animated. Springer, New York, pp 49–76
Markel H (2014) The Hippocratic oath as an example of professional conduct. In: De Angelis CD (ed) Patient care and professionalism. Oxford University Press, New York, pp 19–30
Marketos S (1993) Medicine is an aspect of civilization: lessons from the Hippocratic medicine. Microsurgery 14:4–5
Marketos SG, Skiadas P (1999a) Hippocrates: the father of spine surgery. Spine 24:1381–1387
Marketos SG, Skiadas P (1999b) The modern Hippocratic tradition: some messages for contemporary medicine. Spine 24:1159–1163
Marketos SG, Diamandopoulos AA, Bartsocas CS, Poulakou-Rebelakou E, Koutras DA (1996) The Hippocratic oath. Lancet 347:101–102
Martin DB (2004) Inventing superstition: from the Hippocratic to the Christians. Harvard University Press, Cambridge, MA
Meffert JJ (2009) "I swear!" Physician oaths and their current relevance. Clin Dermatol 27:411–415
Miles SH (2004) The Hippocratic oath and the ethics of medicine. Oxford University Press, New York
Michell CB (2010) The Christian Hippocratic tradition in medicine. Ethics Med 26:69–70
Morgenstern J (2008) The medical oath: honorable tradition or ancient ritual? UWOMJ 78:P27–P29

Myers CS (1900) IV. – vitalism: a brief historical and critical review (I.). Mind 9:218–233
Normandin S, Wolfe CT (eds) (2013) Vitalism and the scientific image in post-enlightenment life science, 1800–2010. Springer, New York
Nutton V (1993) Humoralism. In: Bynum W, Porter R (eds) Companion encyclopedia of the history of medicine. Routledge, London, pp 281–291
Nutton V (2013) Ancient medicine, 2nd edn. Routledge, New York
Ogunbanjo GA, van Bogaert KD (2009) The Hippocratic oath: revisited. SA Fam Pract 51:30–31
Orr RD, Pang N, Pellegrino ED, Siegler M (1997) Use of the Hippocratic oath: a review of twentieth century practice and a content analysis of oaths administered in medical schools in the US and Canada in 1993. J Clin Ethics 8:377–388
Pellegrino ED (1995) Toward a virtue-based normative ethics for the health professions. Kennedy Inst Ethics J 5:253–277
Pellegrino ED (2000) Toward an expanded medical ethics: the Hippocratic ethic revisited. In: Veatch RM (ed) Cross cultural perspectives in medical ethics, 2nd edn. Jones and Bartlett, Boston, pp 41–53
Pinault JR (1992) Hippocratic lives and legends. Brill, Leiden
Pitman V (2006) The nature of the whole: Holism in ancient Greek and Indian medicine. Motilal Banarsidass, Delhi
Prioreschi P (1995) The Hippocratic oath: a code for physicians, not a Pythagorean manifesto. Med Hypoth 44:447–462
Robin ED, McCauley RF (1995) Cultural lag and the Hippocratic oath. Lancet 345:1422–1424
Rosalki J (1993) The Hippocratic contract. J Med Ethics 19:154–156
Scarborough J (1997) Hippocrates of Cos. Anc Greek Authors 176:199–219
Scarborough J (2002) Hippocrates and the Hippocratic ideal in modern medicine: a review essay. Int J Class Tradit 9:287–297
Schiefsky MJ (2005) Hippocrates on ancient medicine. Brill, Leiden
Smith WD (1979) The Hippocratic tradition. Cornell University Press, Ithaca
Smith WD (ed) (1990) Hippocrates pseudepigraphic writings. Brill, Leiden
Smuts JC (1926) Holism and evolution. Little and Ives, New York
Solmsen F (1950) Tissues and the soul: philosophical contributions to physiology. Philos Rev 59:435–468
Sullivan R (1996) Thales to Galen: a brief journey through rational medical philosophy in ancient Greece. Part I: pre-Hippocratic medicine. Proc R Coll Physicians Edinb 26:135–142
Sulmasy DP (1999) What is an oath and why should a physician swear one? Theor Med Bioeth 20:329–346
Sykiotis GP, Kalliolias GD, Papavassiliou AG (2006) Hippocrates and genomic medicine. Arch Med Res 37:181–183
Temkin O (1991) Hippocrates in a world of pagans and Christians. Johns Hopkins University Press, Baltimore
Temkin O (2002) "On second thought" and other essays in the history of medicine and science. Johns Hopkins University Press, Baltimore
Tullis R (2004) Hippocratic oaths. Medicine and its discontents. Atlantic Books, London
Veatch RM (1984) The Hippocratic ethic is dead. New Physician 48:41–42, 48
Wagley PF (1987) The Hippocratic oath. Hum Med 3:110–114
Wheeler LR (1939) Vitalism: its history and validity. Witherby, London
Yapijakis C (2009) Hippocrates of Kos, the father of clinical medicine, and Asclepiades of Bithynia, the father of molecular medicine. In Vivo 23:507–514

Causation and Correlation in Medical Science: Theoretical Problems

51

Federica Russo

Contents

Abstract

Establishing causal relations is a core enterprise of the medical sciences. Understanding the etiology of diseases, and the treatments to reduce the burden of disease, is in fact an instantiation of the very many activities related to causal analysis and causal assessment in medical science. In medicine, correlations have a "Janus" character. On the one hand, we should beware of correlations as they do not *imply* causation – a well-established "mantra" in statistics and in the philosophy of causality. On the other hand, correlations are a very important and useful piece of *evidence* in order to establish causal relations – a line of argument that is currently debated in the philosophical and medical literature. Understanding the limits and potentialities of correlations in medicine is all the more important if we consider the emergence of a "data-intensive science" when the search for correlations in big data sets is becoming key in the medical sciences.

F. Russo (✉)
Department of Philosophy, Faculty of Humanities, University of Amsterdam, Amsterdam, The Netherlands
e-mail: f.russo@uva.nl

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_46

Introduction: Causation in the Medical Sciences

The medical sciences are interested in describing, explaining, and intervening on causes and effects of health and disease. This is a very broad characterization that allows us to be as inclusive as possible when discussing causal issues in medicine.

On the one hand, "medical sciences" is an umbrella term that includes various strands of research, of health care, and, possibly, of alternative or complementary approaches to medicine. Thus, under "medical sciences" we might include clinical medicine, basic health care, epidemiology, gender medicine, or any other approach that engages with the causes and effects of health and disease from a scientific point of view. On the other hand, "causation" is also an umbrella term that includes various ways in which causes and effects are involved in the phenomena of health and disease. Thus, for instance, causation in medicine includes questions about the biochemical mechanisms of disease causation, or about the social inequalities connected to health inequalities, or with the design and implementation of preventative interventions at the individual or population level.

"Causation in the medical sciences" thus refers to a conceptual and methodological complexity that extends far beyond the scope of this chapter. Here, specifically, the focus is on the perennial question about the relation between correlation and causation. More precisely, this question amounts to asking to what extent or under what conditions can one infer causation from correlation – a question that has preoccupied philosophers of causality since the "probabilistic turn" in the 1960s and 1970s. The seminal works of I. J. Good (1961a, b) and Patrick Suppes (1970) paved the way for a probabilistic characterization and analysis of causal relations that is still of relevance today in the natural, social, and medical sciences.

In the medical sciences, the question arises mainly with respect to those research methods and scientific practices that involve substantial use of statistics for the analyses of data. Paradigmatic examples include (various strands of) epidemiology, evidence-based medicine, or data-driven approaches.

In order to grasp what the problem with correlations is, it is vital to understand its methodological basis. Medical research done in labs (e.g., oncological biomedicine) or in randomized controlled trials generates and collects data that have to be analyzed using statistical models. These models establish correlations, and the question is whether – and under what conditions – these correlations can be interpreted as *causal* relations.

Most often, these correlations are *quantitatively* expressed in the language of probability theory and statistics, in which case we talk about *statistical generalizations*. Typical examples are epidemiological claims about the risks of developing a disease given a certain exposure, or of not developing a disease given some preventative interventions. It is no accident, in fact, that the very concept of "risk" in the medical sciences is related to that of "causation," albeit no consensus has been reached about the precise terms of their relationships. But sometimes correlations may take the form of *qualitative* statements about difference-making relations. Further research is needed to spell out the form and use of qualitative difference-making claims in, e.g., the experimental reasoning as put forward by

Claude Bernard (1856) or in the narratives of case reports (on narratives, see, for instance, Goyal 2013, and on the controversial status of case reports in medicine, see, e.g., Nissen and Wynn 2012). Thus the discussion is confined here to correlations expressed in the quantitative language of probability and statistics.

So, given the long-standing discussion in epidemiology and germane areas (e.g., econometrics, quantitative sociology, or demography), it is a legitimate question to ask: What good are correlations in medicine? While it is widely agreed that "correlation is not causation," in the last decade or so, the debate has been shifting from the question of *What is the extra X that makes correlations causal?* (which, admittedly, has a metaphysical flavor) to a question about the import of correlations as *evidence* to establish causal claims (in the medical sciences, but also elsewhere). This shift also means that the question about correlation and causation is now approached from an epistemological and methodological point view, from which metaphysical consequences can be addressed.

Correlation Is Not Causation

It is widely agreed that "correlation is not causation." That this amounts to *nearly* an accepted truth is clear from the way the debate developed in the medical sciences and in the philosophy of causality.

On the one hand, the medical sciences – and especially epidemiology – have been investigating what concept cashes out causation in this context. Mark Parascandola and Douglas L. Weed offer an overview of the various possibilities, showing that none of the available concepts (e.g., counterfactual, INUS, probabilistic, etc.) capture all cases of "medical causation" (Parascandola and Weed 2001). These authors admit that the probabilistic concept of causation is the one that fares better than others, and yet it faces challenges. One reason is that there are cases where causes are necessary and/or sufficient for their effects; another reason is that, admittedly, probabilities (or correlations) do not guarantee causation. Analyses like the one of Parascandola and Weed (2001) typically lead to "precautionary" (and even skeptical) stances about the plausibility and use of an explicit causal talk (see also Lipton and Ødegaard 2005).

On the other hand, the philosophy of causality, since the "probabilistic turn" initiated by authors like Good and Suppes, identified the potential of probabilistic analyses, as well as problems and challenges thereof. Simply put, probabilistic analyses of causality hold that causes make a difference in the probability (of the occurrence) of the effect: $P(E) \neq P(E|C)$. This inequality can be read in two ways. Some causes *raise* the probability of the effect – for instance, consumption of fat food raises the probability of cardiovascular diseases. Some causes, instead, *lower* the probability of the effect – for instance, regular exercise *prevents* cardiovascular diseases. It is worth noting that the requirement of "difference-making" was originally stated only in terms of probability *raising*, thus questioning the status of preventatives as proper "causes" (for a discussion, see Illari and Russo 2014,

Chap. 8). This, however, is not the main conceptual challenge concerning "correlation and causation."

Two conceptual challenges can be identified. The first has to do with the "generic versus single-case" problem. The second has to do with the "third variable" problem.

The first problem – generic versus single-case – has been debated at length in the philosophical literature, and it is also known as the "population versus individual" or "type versus token" problem (for a discussion about terminological distinctions and similarities, see Illari and Russo 2014, Chap. 5). Simply put, philosophers soon realized that probabilistic inequalities valid at the population level do not guarantee causal inference to the individual level. Let's illustrate with a simplified medical example. The medical sciences have established, with a significant degree of confidence, that smoking causes (in the sense of *raising the probability of*) lung cancer. Yet, it is well known that some cases of lung cancer are not due to smoking and that not all smokers develop lung cancer. Situations like this raise at least two theoretical questions. One question concerns the status of (statistical) generalizations that purportedly have *causal* meaning: What does it mean that smoking causes cancer *at the population level*? Another question, also related to the previous one, concerns the "metaphysical priority" of population- or individual-level causal relations. Are causal relations *generic* (or type-level) and from these we have to derive single-case (or token-level) ones? Or, instead, are causal relations *single-case* and generic claims are mere (statistical) aggregates of those? Federica Russo and Jon Williamson analyze the relation between generic and single-case causal relations in medicine and suggest that one level has no metaphysical priority over the other (Russo and Williamson 2011b). Rather, from an epistemological and methodological point of view, there is a mutual dependence between the two levels.

The second problem – third variable – has been studied at length in statistics (with applications to medicine, social science, or others). To understand what the problem of the third variable amounts to, it is useful to recall the main steps involved into establishing correlational claims through statistical modeling. Briefly and simply put, once data are collected, scientists organize them according to variables and then study the dependencies and independences between these variables using statistical models. Each of these steps involves conceptual, methodological, and practical challenges that will not be discussed here. The point at stake is that, given the correlation between variables X and Y, it is possible to find a *third variable* Z, which, when included in the model, questions the validity of the correlation between X and Y. This may happen for a number of reasons. Let's illustrate with the aid of toy examples.

Consider a case where variable "yellow fingers" is correlated with variable "lung cancer"; however, once we introduce the variable "cigarette smoking," the correlation disappears because "cigarette smoking" causes *both* "yellow fingers" and "lung cancer" (cases like this are also discussed as instantiations of the common cause principle and cigarette smoking is said to *screen off* "yellow fingers" from "lung cancer"). A different case concerns the correlation between variables "coffee drinking" and "cardiovascular disease." This correlation may, or may not,

disappear when we introduce variable "cigarette smoking." In fact, "cigarette smoking" may be a common cause of both "coffee drinking" and "cardiovascular disease" – in this case the correlation would disappear. But it is also possible that "coffee drinking" has its own effect on "cardiovascular disease." In this case we should study the effects on "cardiovascular disease" due to "coffee drinking," those due to "smoking," and those due to possible interactions between the two. In cases like this, we have to *control* for possible *confounding variables*. Available statistical techniques for control include conditioning on specific values of variables, stratification ex ante or ex post, etc.

Generally speaking, a main theoretical challenge with correlations is their validity, whether internal or external. Thomas D. Cook and Donald T. Campbell introduced the terms internal and external validity, in the area of quasi-experimental methods (Cook and Campbell 1979). Internal validity refers to the confidence with which we deem the correlation between variables *X* and *Y* causal, in the population of reference. External validity refers instead to the possibility of establishing the same correlation also *outside* the population of reference. Thus, for instance, in the medical sciences we might be interested in establishing the efficacy of a drug for a specific population of reference and also for *other* populations. The same holds for the efficacy of public health interventions.

Another important aspect of correlations in the medical sciences concerns the fact that correlational claims are *generic*. One reason why we are interested in establishing (and validating) generic claims is that they contribute to building medical knowledge (population level) *and* to serve as a basis for diagnosis and prognosis (individual level). The challenge here concerns the kind of information that correlations provide for the purpose of establishing causal claims. It is worth noting that, thus formulated, the focus is shifted from the question *Why should we beware of correlations?* to the question *What good are correlations in the medical sciences?*. This latter question is examined next.

Correlation as Evidence

In the philosophy of causality, part of the debate has been devoted to explicating the very concepts of "cause" or "causality." Several accounts have been proposed, for instance, analyzing the concept of cause/causality in terms of counterfactuals, necessary and sufficient components, invariance relations, probabilities, etc. (for systematic presentation of such attempts, see Illari and Russo 2014). As mentioned above, the applicability of these concepts to the medical sciences, and to epidemiology in particular, has also been examined by Parascandola and Weed (2001). A different line of argument, however, has been introduced in the debate since the paper by Russo and Williamson (2007).

Russo and Williamson argue that the *concept* of causality should not be confused with the *evidence* needed to establish causal claims. Most accounts of causality can be reinterpreted as offering an account of the type of evidence to support causal claims. The thesis, now customarily referred to in the literature as "Russo-Williamson thesis"

(RWT), states that, typically, causal claims in medicine are established on the basis of *evidence* of difference-making and of mechanisms. This emphasis on the *evidence* needed to establish a causal claim makes RWT epistemological and methodological in character. A corollary of RWT is *evidential pluralism*, namely, the position according to which causal claims are established on the basis of multifarious evidence. Thus correlations are an important *evidential component* to establish causal claim, but causality is *not reduced* in any way to correlations (nor to mechanisms). RWT sparked lively debates and a new stream of research. Let us examine the core of the thesis, some of the prospective developments, and objections.

To begin with, it is worth clarifying the status of the thesis. RWT is an *epistemological* thesis about how to establish causal knowledge (in medicine). In particular, it is a thesis about the evidence that supports causal claims. RTW is *not* a metaphysical thesis about the *nature* of causality. In particular, the thesis does not state that causality is *constituted by* difference-making (correlations) and mechanisms. RWT can of course be discussed in its metaphysical implications, but that is an orthogonal issue. In some papers, Russo and Williamson (2011a) couple evidential pluralism with the epistemic theory of causality (for details, see Williamson 2005): Difference-making and mechanisms are evidential components, and the concept of causality is provided by the epistemic theory, according to which causality is the ultimate belief of an omniscient agent. Here, an epistemology for causal relations (RWT) is combined with a metaphysical theory about causation (the epistemic theory). However, not everyone embracing RWT also endorses this metaphysical position (see, e.g., Gillies 2011; Clarke et al. 2014).

Let's now go into the details of the thesis. Phyllis Illari (2011) points out that RTW should not be read as saying that there are different *types* of evidence (difference-making and mechanisms), but that difference-making and mechanisms capture the *object* of evidence, i.e., what we have (or need) evidence *of*. The difference is subtle but fundamental. In the first case, we are interpreting "difference-making" and "mechanisms" rigidly, as if these were fixed categories, and causal claims were established by ticking both boxes. In the second case, instead, "difference-making" and "mechanisms" refer to the type of information that we examine in establishing causal claims. Under this reading, evidential components become highly intertwined and interdependent, which is actually the case in the scientific practice. For instance, suppose a scientist is observing the modes of transmission of a bacterium, say, *Vibrio cholerae*. These observations may provide evidence of the mechanisms underlying the transmission; the same observations may also provide evidence that the bacterium makes a difference to the occurrence of the disease. Two things are worth noting. First, read this way, RWT does not imply that we must have full or complete knowledge of disease mechanisms – a point also made by Donald Gillies (2011). Second, difference-making can be quantitatively expressed in terms of (statistical) correlations but can also be expressed using qualitative statements, for instance, about counterfactuals.

It is now possible to explain more clearly the import and meaning of difference-making and of mechanisms. Evidence of difference-making is useful

in order to establish *change-relating* relations or to make predictions – to do so we need to know *that C* causes *E*. Evidence of mechanisms is useful in order to explain disease or to design intervention to reduce the burden of disease – to do so we need to know *how C* causes *E*. Thus, the term "evidence of production" better grasps what is at stake, as mechanisms are just one way in which causes produce effects – processes, information transfer, and the action of capacities are other ways in which we can grasp how causes produce effects (for a discussion see Illari and Russo 2014, Chap. 6).

Evidence of difference-making must be also considered as complementary to evidence of production. In fact, while evidence of production helps with *confounding* (one variant of the "third variable" problem mentioned above), evidence of difference-making helps with *masking*. Masking is the problem of establishing which mechanism "wins," when competing mechanisms are simultaneously active. For instance, exercising makes you burn calories and thus lose weight; but, at the same time, exercising makes you hungry and eat more. It is difficult to say which out of the two mechanisms will "win." So confounding and masking are in fact the two sides of the same coin. Evidence of production helps us decide what variables to include, exclude, or control in the statistical model. Evidence of difference-making helps us disentangle the different effects when multiple causal paths are simultaneously at work.

Rethinking correlations as an evidential component for establishing causal claims is also interesting in the light of Bradford Hill's viewpoints on causal inference (Hill 1965). In this famous paper, Hill formulated nine aspects to consider when making a judgment about a correlation between two variables. Howard Frumkin (2006) summarizes them thus:

1. **Strength of association**. The stronger the relationship between the independent variable and the dependent variable, the less likely it is that the relationship is due to an extraneous variable.
2. **Temporality**. It is logically necessary for a cause to precede an effect in time.
3. **Consistency**. Multiple observations, of an association, with different people under different circumstances and with different measurement instruments increase the credibility of a finding.
4. **Theoretical plausibility**. It is easier to accept an association as causal when there is a rational and theoretical basis for such a conclusion.
5. **Coherence**. A cause-and-effect interpretation for an association is clearest when it does not conflict with what is known about the variables under study and when there are no plausible competing theories or rival hypotheses. In other words, the association must be coherent with other knowledge.
6. **Specificity in the causes**. In the ideal situation, the effect has only one cause. In other words, showing that an outcome is best predicted by one primary factor adds credibility to a causal claim.
7. **Dose-response relationship**. There should be a direct relationship between the risk factor (i.e., the independent variable) and people's status on the disease variable (i.e., the dependent variable).

8. **Experimental evidence**. Any related research that is based on experiments will make a causal inference more plausible.
9. **Analogy**. Sometimes a commonly accepted phenomenon in one area can be applied to another area.

Viewpoints 1, 3, 7, and 8 are about difference-making, while viewpoints 2, 4, 5, 8, and 9 are about production or mechanisms. This is interesting because if the scientific community by and large accepts Hill's viewpoints, then RWT-like arguments are a good candidate for a philosophical conceptualization of the importance of correlations and of their complementarity to considerations about the mechanisms of disease causation. But the interest in Hill's viewpoints is not confined to RWT-like arguments. In fact, inferential approaches such as the one developed by Julian Reiss (2015) also appeal to these different aspects of causal relations. As Reiss himself puts it:

> *The* [*inferentialist theory of causality*] *maintains that the meaning of causal claims is given by their inferential connections with other claims. In particular, causal claims are inferentially related to evidential claims—the claims from which a causal claim can be inferred—as well as to claims about future events, explanatory claims, claims attributing responsibility, and counterfactual claims* (*claims predicting 'what would happen if'*)*—the claims that can be inferred from a causal claim.*

This opens up new spaces for philosophical investigations in order to understand the role and use of correlations in causal inference.

So far, the discussion about evidential pluralism has not made clear whether this position is normative or descriptive – a worry expressed by, e.g., Alex Broadbent (2011). On the one hand, a descriptive reading of evidential pluralism would simply testify the use of multifarious evidence in the medical sciences (and elsewhere). Of course, no description is totally neutral, and the account should explain how (historical accounts of) scientific practices are analyzed. On the other hand, a normative reading of evidential pluralism would prescribe current and future scientific practices to adhere to it. This is an attractive option, but one that should be handled with care. In fact, there is no simple way in which philosophy can tell science what to do, in a simple top-down way. The debate on the role and use of correlation should seen as an opportunity to foster a dialogue between philosophy and medicine, philosophers and medical scientists.

Stakes are high for two reasons. One reason is that the philosophy underlying evidential pluralism is in much need of input coming from the medical sciences. Another reason is that it is controversial whether evidential pluralism fits different scientific practices in the medical sciences. This is related to the problem, mentioned in the opening of this contribution, of defining the medical sciences. Some accounts aim at being as inclusive as possible, and they have to provide an account of evidential pluralism (and, for the matter, of the meaning, role, and use of correlations) in practices as diverse as randomized controlled trials, case reports, cohort studies, diagnosis, etc. In turn, this is related to questions about methodological pluralism. In fact, under those accounts embracing the view of "medical sciences" as an umbrella term for different scientific practices, one should also accept, albeit implicitly, that

causal relations are established using different methods, depending on the context. Thus, for instance, it is one thing to establish the efficacy of a drug or a treatment, for which a randomized controlled trial is perfectly appropriate, and it is another thing to establish what disease is causing such and such symptoms in a particular patient, for which *other* methods are appropriate.

Correlation and Data-Intensive Science

The problem of inferring causation from correlation should also be discussed in the context of *data-intensive science*. Medicine and epidemiology are increasingly using bigger and bigger data sets. Examples of research projects where big data sets are customarily created, analyzed, and used abound. One such example is the "EPIC" cohort. EPIC (European Prospective Investigation into Cancer and Nutrition) is a project jointly coordinated by IARC (International Agency for Research on Cancer) and Imperial College London. Initially, the project investigated the relation between (different types of) cancer and nutrition; later, the study also investigated several chronic diseases such as diabetes and included also genetic and environmental factors. Only between 1992 and 2000, the study examined data about some 520,000 individuals. Another study is the European consortium working on the "EXPOsOMICS" project. This project aims at developing novel methods to study environmental exposures, such as air pollution and water contamination, on selected diseases. A peculiarity of this project is the use of "omics technologies" that allow scientists to study changes in our bodies at the molecular level. The project also uses data from the EPIC cohort as well as many others. In spite of the hope of being able to establish meaningful (and even causal) correlations in big data, there are a number of delicate issues that the scientific and philosophical communities are currently debating.

One aspect relates to the "size" of these data sets: Is it really the *size* making the novelty, or is it something else? It might be argued that the use of emerging technologies, such as the omics technologies, is what allows us to produce data sets of unprecedented size. But, in turn, the use of technology for the production and analysis of data raises several methodological and epistemological issues. Some concern the very conceptualization of data (simply put, in spite of the name, data are not *given*, but rather constructed – for one account see, e.g., Leonelli 2015), while others concern the techniques for data analysis (statistics and data mining – see, e.g., discussion in a special issue (Merelli et al. 2014)).

In sum, providing an understanding of the role and use of correlations in causal inference raises important questions about the nature of causation itself but also about evidence and methods. These topics add up to the well-known, and much discussed, issues related to common cause structures or confounding that occupied much of the debate in statistics and in the philosophy of causality so far. The recent emphasis on data-intensive science, while opening up opportunities for studying correlations on even larger data sets, urges philosophical analyses about their conceptual underpinnings.

Definition of Key Terms

Correlation	The relation between two variables indicating that they are statistically dependent.
Causation	The relation between variables or events, indicated the relations of dependence and production between them.
Evidence	The information, input, or observation used to assess and support scientific claims about causation, explanation, or prediction.
Evidential pluralism	The position according to which evidence is multifarious.
Evidence of production	Information gathered from lab experiments, statistical studies, or other types of studies indicating how a cause produces its effect(s).
Evidence of difference-making	Information gathered from lab experiments, statistical studies, or other types of studies indicating that a case makes a difference to the occurrence of its effect(s).

Summary Points

- Large part of contemporary medicine is concerned with establishing causal relations with the aid of statistics.
- Studies using statistics have to consider carefully that "correlation is not causation," just as any other discipline that relies on quantitative analyses of data.
- Statistical tools used in medicine raise problems that are akin to those raised in other disciplines, for instance, the problem of confounding and control and the choice of variables, of the models, and of data in the first place.
- Despite all these warnings, correlations remain very useful to establish causal relations, as they are *evidence* for causal relations.
- Conceiving of correlations as evidence is part of a larger view of causation, according to which causal relations are established, based on various sources of evidence.
- Understanding the status of correlations for causal assessment in medicine is vital; data-driven approaches – where the search of correlation is a pillar – are more and more widespread.

References

Bernard C (1856) An introduction to the study of experimental medicine, 1927th edn. Macmillan, New York

Brendan C, Gillies D, Illari P, Russo F, Williamson J (2014) Mechanisms and the evidence hierarchy. Topoi 33:339–360. doi:10.1007/s11245-013-9220-9, Online first

Broadbent A (2011) Inferring causation in epidemiology: mechanisms, black boxes, and contrasts. In: Illari PMK, Russo F, Williamson J (eds) Causality in the sciences. Oxford University Press, Oxford, pp 45–69

Cook TD, Campbell DT (1979) Quasi-experimentation. Design and analysis issues for field settings. Rand MacNally, Chicago

Frumkin Howard (2006) Causation in medicine. http://www.aoec.org/CEEM/methods/emory2.html

Gillies D (2011) The Russo-Williamson thesis and the question of whether smoking causes heart disease. In: Illari PMK, Russo F, Williamson J (eds) Causality in the sciences. Oxford University Press, Oxford, pp 110–125

Good IJ (1961a) A causal calculus I. Br J Philos Sci 11:305–318

Good IJ (1961b) A causal calculus II. Br J Philos Sci 12:43–51

Goyal R (2013) Narration in medicine. In: Pier J, Schmid W, Hühn P, Meister JC (eds) The living handbook of narratology. Hamburg University, Hamburg, http://www.lhn.uni-hamburg.de/article/narration-medicine

Hill BA (1965) The environment of disease: association or causation? Proc R Soc Med 58:295–300

Illari PMK (2011) Mechanistic evidence: disambiguating the Russo-Williamson thesis. Int Stud Philos Sci 25(2):139–157

Illari P, Russo F (2014) Causality: philosophical theory meets scientific practice. Oxford University Press, Oxford

Leonelli S (2015) What counts as scientific data? A relational framework. Philos Sci 82:1–12

Lipton R, Ødegaard T (2005) Causal thinking and causal language in epidemiology. Epidemiol Perspect Innovat 2:8, http://www.epi-perspectives.com/content/2/1/8

Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D (2014) High-performance computing and big data in omics-based medicine. BioMed Res Int 825649. doi:10.1155/2014/825649

Nissen T, Wynn R (2012) The recent history of the clinical case report: a narrative review. RSM Short Rep 3(12):87. doi:10.1258/shorts.2012.012046

Parascandola M, Weed DL (2001) Causation in epidemiology. J Epidemiol Community Health 55:905–912

Reiss J (2015) Causation, evidence, and inference. Routledge, New York

Russo F, Williamson J (2007) Interpreting causality in the health sciences. Int Stud Philos Sci 21(2):157–170

Russo F, Williamson J (2011a) Epistemic causality and evidence-based medicine. Hist Philos Life Sci 33:563–582, http://philsci-archive.pitt.edu/id/eprint/8351

Russo F, Williamson J (2011b) Generic vs. single-case causal knowledge. The case of autopsy. Eur J Philos Sci 1(1):47–69

Suppes P (1970) A probabilistic theory of causality. North-Holland, Amsterdam

Williamson J (2005) Bayesian nets and causality: philosophical and computational foundations. Oxford University Press, Oxford

Evidence-Based Medicine in Theory and Practice: Epistemological and Normative Issues

52

Wendy Rogers and Katrina Hutchison

Contents

Abstract

Evidence-based medicine (EBM) emerged during the 1990s, with the aim of improving clinical practice by increasing the extent to which clinical care was informed by medical research, particularly randomized controlled trials (RCTs) and systematic reviews of RCTs. This chapter gives an account of EBM, followed by examination of epistemological and ethical justifications and critiques of EBM. EBM relies upon epistemological claims about the ability of RCTs to eliminate certain forms of bias and to establish whether or not there is a causal

W. Rogers (✉)
Department of Philosophy and Department of Clinical Medicine, Macquarie University, Sydney, NSW, Australia
e-mail: wendy.rogers@mq.edu.au

K. Hutchison
Department of Philosophy, Macquarie University, Sydney, NSW, Australia
e-mail: katrina.hutchison@mq.edu.au

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_40

relationship between an intervention and an outcome. However, epistemological critiques of EBM include reservations about whether EBM can "prove" causation, concerns about the rejection of mechanistic models of causation, challenges associated with applying the results of RCTs to individual patients, and lack of evidence regarding whether EBM has in fact benefitted patients and healthcare systems. The ethical justifications for EBM include its promise of better patient outcomes through better informed clinicians and the idea that public health policy based on EBM can support equity and minimize waste of resources. Ethical critiques of EBM note that despite its potential for reducing particular forms of bias, the research upon which EBM is based is often industry funded, creating conflicts of interest that are associated with new sources of bias. These include bias in the conduct of trials, the publication of results, and the choice of interventions for investigation. EBM also poses challenges for patient and clinician autonomy, especially where evidence-based clinical practice guidelines are enforced through targets or audits. In the face of these concerns, EBM is under pressure to reestablish its credibility. The chapter ends by identifying three current initiatives that seek to reinstate the aims of EBM to better inform healthcare decisions.

Introduction

Evidence-based medicine (EBM) is a formalized approach to using the results of research trials to inform the care of patients. It has been hugely influential in medical practice and medical education and upon health services more broadly. This chapter explains what EBM is and provides a brief account of the development of EBM since its introduction in the 1990s, before describing the epistemological and ethical foundations of EBM and current critiques of these. The chapter ends by noting suggestions for the future of EBM.

The term "evidence-based medicine" (EBM) was coined in 1980 to describe the appraisal and use of research results in the care of individual patients, as first proposed by the EBM Working Group at McMaster University. EBM sought to change the way that clinicians think about medical knowledge. The EBM Working Group (1992) described this change as a "paradigm shift" in what should count as evidence strong enough to inform medical practice. This shift heralded a move away from decisions based upon what were considered to be unsystematic clinical observations, reliance on mechanistic reasoning and pathophysiological principles, and deference to the views of experts. In contrast, EBM advocated decisions based upon the statistical analysis of the results of research trials.

Over time, EBM has had a major impact upon healthcare practice and policy, as a method for identifying and appraising the results of research studies and of synthesizing this information to guide clinical decision making. In one of the earliest papers by the EBM Working Group, EBM is described as a dramatic change "which involves using the medical literature more effectively in guiding medical practice" (EBM Working Group 1992, p. 2420).

More formal definitions of EBM emerged during the 1990s. One of the most significant of these was published in a 1995 editorial marking the launch of the first dedicated EBM journal:

> [E]vidence based medicine is rooted in five linked ideas: firstly, clinical decisions should be based on the best available scientific evidence; secondly, the clinical problem - rather than habits or protocols - should determine the type of evidence to be sought; thirdly, identifying the best evidence means using epidemiological and biostatistical ways of thinking; fourthly, conclusions derived from identifying and critically appraising evidence are useful only if put into action in managing patients or making health care decisions; and, finally, performance should be constantly evaluated. (Davidoff et al. 1995, p. 1085)

This comprehensive definition of EBM identifies the key notion of using the best available scientific evidence to address the clinical problems of individual patients. By specifying that the "best evidence" is derived from "epidemiological and biostatistical ways of thinking," this definition introduces the normative claim that some forms of evidence are to be preferred over others (Djulbegovic et al. 2009). The kind of evidence that is most highly valued within EBM is that produced from randomized controlled trials (RCTs) about the efficacy and safety of healthcare interventions (Sehon and Stanley 2003). There is a second normative claim in this definition, which is that clinical decisions should be based upon the best evidence as specified (Djulbegovic et al. 2009). Thus, EBM is defined, famously by Sackett et al., as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (1996, p. 71).

The new approach of EBM was underpinned by four assumptions. The first of these assumptions makes claims about the unreliability of clinical experience and intuition compared with knowledge obtained from the systematic and unbiased collection of observations, such as occurs in high-quality research. The second assumption is that another traditional source of medical knowledge, derived from pathophysiological principles, is likewise unreliable. While understandings of basic disease mechanisms are useful to guide clinical practice, it was claimed that relying upon pathophysiological principles may lead to adverse events or inaccurate estimates about the efficacy of interventions. The third assumption underpinning EBM is that in order to critically appraise and correctly interpret research literature (i.e., identify and use the best evidence), it is necessary for clinicians to understand certain biostatistical "rules of evidence," understood in terms of the reliability of statistically identified associations in research. The final assumption is that EBM will lead to "superior patient care" (EBM Working Group 1992, p. 2421). Thus, EBM marked a change away from clinical experience, mechanistic based reasoning, and the uncritical or haphazard use of research results toward independent practice based upon the critical appraisal of research results and the use of probabilistic evidence about the efficacy of interventions.

There have been four separate models of EBM since the late 1980s (Charles et al. 2011; Wyer and Silva 2009). The first consisted of formally applying clinical research evidence to medical practice (EBM Working Group 1992). The second model, from the mid-1990s, advocated for decision making based upon patient

preferences, research evidence, and clinical expertise in equal measure while explicitly recognizing the challenge of integrating research evidence with clinical expertise. Third, the prescriptive model of the early 2000s incorporated patient preferences with research evidence, together with information about the patient's clinical status; in this model, clinical expertise was seen as the overarching mechanism to combine these three elements. Just how this was to be achieved remained unclear. The most recent model of the mid- to late 2000s is called a model for evidence-based clinical care, applicable to healthcare practices beyond medicine. This model sees the inclusion of a fourth element, healthcare resources, to be considered along with patient values, research evidence, and clinical status. As with the previous model, clinical expertise must draw upon all of these elements in order to reach a considered decision.

Thus, EBM has evolved, from an intuitively attractive and almost unassailable initial proposal to use the best and latest research evidence to inform clinical decisions, into a much more specific and complex model prescribing the exercise of clinical expertise to reach a decision based upon patient values, the clinical status of the patient, and the availability of resources, as well as research evidence.

There are two main reasons as to why EBM emerged when it did. The first relates to growing recognition of the gap between research evidence and clinical practice, which resulted in "expensive, ineffective or harmful decision making" (Rosenberg and Donaldson 1995, p. 1122). This gap led to significant variations in practice, marked by the slow uptake of effective interventions, such as streptokinase for myocardial infarction, and the equally slow abandonment of harmful practices such as the use of anti-arrhythmic prophylaxis following myocardial infarctions (Djulbegovic et al. 2009; Howick et al. 2013). Clearly, it was problematic for practitioners to be putting their patients' lives at risk by being out of touch with new research evidence. The second reason may explain why clinicians had trouble keeping up with the literature: this period saw an explosion in the numbers of published papers. The increase in published research has been attributed, at least in part, to the 1962 Kefauver-Harris Amendments to the United States Federal Food, Drug and Cosmetic Act, which required that firms had to provide evidence of the effectiveness of their products. In addition, the tools of biomedical informatics facilitated literature searching, and the widespread use of computers allowed such searching to take place in the clinic rather than the library.

EBM promised a more reliable way of ascertaining the effectiveness of medical interventions than was possible using "the former paradigm" (EBM Working group 1992, p. 2421). EBM methods have the potential to reduce or eliminate bias and provide statistically significant evidence of efficacy even in the absence of an understanding of causal mechanisms. As well as the self-evident benefit of discriminating between ineffective or harmful interventions and safe and effective ones (Goodman 2003), EBM has a number of other advantages. EBM integrates research with clinical care, teaches clinicians how to critically appraise clinical trials, informs better use of resources by evaluating the clinical effectiveness of interventions, is broadly democratic in that most people can learn the skills of critical appraisal, and may foster better communication with patients (Rosenberg and Donald 1995).

EBM has had an enormous impact upon healthcare (Greenhalgh et al. 2014). First, EBM has led to a focus on research methodology leading to higher standards for research trials and publications. Well-known examples of these include the CONSORT Statement which is an evidence-based set of recommendations for reporting RCTs (Schulz et al. 2010), the GRADE approach for assessing the quality of evidence and the strength of recommendations (Guyatt et al. 2008), and the PRISMA Statement for reporting systematic reviews and meta-analyses (Moher et al. 2009). These and other EBM standards have been widely accepted and are used as benchmarks for assessing the quality of research. Second, EBM has led to the development of national and international organizations, such as the Cochrane Collaboration, undertaking systematic reviews, or those developing and updating evidence-based clinical practice guidelines (CPGs). Clinical practice guidelines have turned out to be the dominant mechanism by which research evidence is synthesized into a format that can be used by practitioners, as it is unfeasible for individual practitioners to perform their own systematic reviews. Third, EBM has dramatically increased the information literacy of clinicians (Wyer and Silva 2009). Finally, the methods of EBM have enabled clinicians and others to map the rapidly changing knowledge base, which is a prerequisite for knowledge translation.

The Epistemology of EBM

The central epistemological claim underlying EBM concerns what counts as good evidence for clinical decisions. Sackett et al. claimed that evidence from clinical experience alone (the "old paradigm") is biased, and therefore unreliable, and hence that systematic approaches to evidence should be preferred (1996). Several sources of bias were identified. First, doctors are likely to remember patients with good outcomes and hence consider their treatment effective. But the good outcomes may be unrelated to the treatment. For example, compliant patients, who return for follow-up, are more likely to get positive outcomes even if the treatment is ineffective. Second, most symptoms and signs tend to regress toward the mean over time irrespective of any intervention, but this can make any intervention administered in the interim seem effective. Third, efficacy may be overestimated in clinical care because of the placebo effect in the patient and the effect of the desire for success in both patient and doctor, biases which can only be eliminated by blinding within RCTs (Sackett 1989). Finally, even the most rigorous causal-inductive or mechanistic reasoning can be fallible (Djulbegovic et al. 2009; Howick et al. 2013).

In contrast, EBM uses a hierarchy of evidence, based on claims about reliability of knowledge obtained from different research methods. Evidence is graded into levels based upon certain methodological features of the research, specified in the hierarchy of evidence. Three central claims underpin the hierarchy. First, randomized controlled trials (RCTs) or systematic reviews of RCTs provide stronger evidence than observational studies. Second, comparative clinical trials (including RCTs and observational studies) offer stronger evidence than reasoning from pathophysiological principles (also described as mechanistic reasoning).

Third, comparative clinical studies offer stronger evidence than expert clinical opinions (Howick 2011). Thus, epidemiological evidence – whether in the form of RCTs or observational studies – is privileged over pathophysiological or mechanistic evidence, and both are privileged over the expertise or intuition of individual clinicians. The hierarchy has evolved over time, but its essential features remain largely unchanged: RCTs, or systematic reviews of RCTs, are at the top of the hierarchy as they are considered to be the most reliable and unbiased form of evidence. An early four-level hierarchy was published in 1979 (see Table 1: Canadian Task Force 1979, p. 1195). As with those that followed, the highest level of evidence ("best") derives from RCTs, while expert opinion is ranked fourth.

The most complex hierarchy evolved in the 2000s, developed by the Oxford Centre for Evidence-Based Medicine (OCEBM). This refers to the use of particular techniques for collating and synthesizing research results from multiple sources (known as systematic reviews) and specifies particular methodological features of trials that render them of greater or lesser quality (see Table 2 which is an adapted and simplified version of this hierarchy taken from the cited source).

Table 1 Levels of evidence (1979)

Level of evidence	Type of evidence
I	At least one properly randomized controlled trial
II-1	Well-designed cohort or case–control study, preferably from more than one center or research group
II-2	Evidence from studies comparing groups of patients between times or between places who did and did not receive the intervention under study
III	Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees

Canadian Task Force (1979, p. 1195: Table created from text in original)

Table 2 Summary of OCEBM 2001 levels of evidence

Level of evidence	Type of evidence
1	1a. Systematic reviews of RCTs (with homogeneity)
	1b. High-quality RCTs (with narrow confidence intervals)
	1c. All or no trials
2	2a. Systematic reviews of cohort studies (with homogeneity)
	2b. Individual cohort study (including low-quality RCTs, e.g., <80 % follow-up)
	2c. "Outcomes" research; ecological studies
3	3a. Systematic reviews (with homogeneity) of case–control studies
	3b. Individual case–control study
4	Case series (and poor-quality cohort and case–control studies)
5	Expert opinion without explicit critical appraisal or based on physiology, bench research, or "first principles"

Adapted from the version of the 2001 OCEBM hierarchy (as cited in BJU 2010)

Table 3 Summary of OCEBM 2011 evidence hierarchy for treatment interventions

Level of evidence	Type of evidence
1	Systematic review of RCTs or *n*-of-1 trials
2	RCT or observational study with dramatic effect
3	Non-randomized controlled cohort/follow-up study
4	Case series, case–control studies or historically controlled studies
5	Mechanism-based reasoning

Adapted from OCEBM (2011)

The OCEBM has now simplified their hierarchy (see Table 3), which, apart from the addition of systematic reviews as level 1, otherwise closely resembles the original 1979 hierarchy.

The randomized controlled trial (RCT) is at the heart of EBM. RCTs aim to produce valid results by ruling out more confounding factors than other research methods. Observational studies are susceptible to at least three kinds of bias, including self-selection bias whereby patients who choose, or are chosen, to participate in the study differ in important ways from patients who are not chosen; allocation bias, in which those recruiting research participants systematically favor those with certain characteristics, such as likely compliance, which affect the outcomes; and performance bias which occurs when patients know they are taking an experimental intervention, and this knowledge affects their behavior and outcomes. RCTs seek to overcome these biases and any other potential confounding factors. Randomization prevents self-selection and allocation bias by allocating participants randomly to each arm of the trial, so that any unknown variables are likely to be distributed equally between the groups. A second feature of RCTs, blinding, addresses performance bias in patients as well as confounding factors generated by clinicians whose views about the experimental intervention may affect their assessments of the outcomes. Blinding involves concealing the allocation (e.g., to the active or control arm of the trial) from both patient and researcher/treating clinician. A third feature of RCTs is that they compare the intervention in question with a control which may be standard therapy or a placebo (usually justified only if no alternative effective therapy exists). Thus, in an ideal RCT, an adequate number of patients are randomly allocated to a blinded treatment, the effects of which are assessed by a clinician/researcher who is unaware as to whether the patient is receiving the intervention or the control. In an ideal RCT, in which bias has been reduced to the extent possible and which has an appropriate sample size, any significant effects noted in the trial can be attributed to the intervention itself rather than bias, chance, or any other confounding factors.

Epistemological Critiques of EBM

EBM has been subject to a number of epistemological criticisms. In this section several of the most prominent are discussed. These include concerns about the extent to which the epidemiological methods privileged by EBM can demonstrate causal

relationships and related concerns about the discounting of pathophysiological evidence. Epidemiological research offers a very different type of causal understanding compared with pathophysiological research. Critics, however, point out that both have disadvantages; thus it is erroneous to exclude pathophysiological reasoning as a source of medical evidence. A further concern arises about how to apply the findings of RCTs, which generate probabilistic information about selected populations, to individual patients who share some (but not other) features with the trial population. Finally, there are concerns about judging the effectiveness of EBM. As it is not possible to randomize health systems either to use or not use EBM, the decision to practice EBM within a health system cannot be based on what EBM itself regards as the highest level of evidence.

One of the forms of evidence to which EBM ascribes a low value is pathophysiological evidence. Understanding the physiological underpinnings of disease should help to inform the development of effective treatments. When this works, it can lead to immediate, significant improvements in outcomes, as in the case of volumetric treatment of blood loss for hemorrhagic shock (Hardaway 2004). In contrast, however, this reasoning can lead to adverse outcomes when the mechanisms underlying mortality and morbidity are not fully understood. An oft-cited example is the increased mortality associated with prescription of anti-arrhythmia drugs following myocardial infarction. Despite preventing arrhythmias, which are mechanistically linked to mortality, these drugs did not reduce mortality but conversely increased it (Howick 2011, pp. 4–5).

One problem with mechanistic evidence is that it focuses on pathophysiological effects/pathways rather than patient outcomes. Thus, efficacy may be measured in terms of reduction of arrhythmias, or changes to other biological indicators, rather than in terms of significant patient end points (such as reduced mortality, alleviation of pain, or increased function) for which these are proxies. In contrast, because the biostatistical methods privileged by EBM tend to compare patient outcomes for two or more groups of patients, without necessarily referring to underlying mechanisms, they are able to identify effective or ineffective treatments even where the underlying causal mechanisms are unknown or not completely understood. However, this advantage of EBM is forfeited when RCTs have intermediate or surrogate end points, such as levels of HbA1C to reflect glycemic control in diabetes, rather than clinically relevant end points such as deaths from heart attacks.

The epistemological advantage of RCTs is their capacity to identify a causal relationship between a particular treatment and a patient outcome. This causal relationship is statistical and probabilistic. It is based upon statistically significant differences in outcomes occurring between groups that are otherwise identical aside from the intervention they receive. In such cases the difference in outcomes is attributed to the intervention because confounding factors are ruled out by the study design. However, the claim that all confounding factors are ruled out in high-quality RCTs has been challenged by a number of critics. While RCTs can rule out known confounders (such as age, sex, and known comorbidities) by equally distributing them across the arms of the trial or by performing baseline comparisons, it remains possible that unknown confounders are not equally distributed (Worrall 2007, 2011).

While it is correct that biostatistical methods, including well-designed RCTs, cannot infallibly demonstrate causal relationships between an intervention and an outcome, these concerns do not undermine the evidence hierarchies recommended by proponents of EBM, as the findings of well-designed RCTs are less likely to be confounded than those of observational studies or other non-randomized studies subject to self-selection, allocation, or preference bias. However, other forms of bias may affect the reliability of EBM, especially when the design and conduct of research are affected by conflicts of interest (see section on Ethical Critiques of EBM).

RCTs make claims about causation based upon tests of significance, which are used to determine whether any differences in outcomes between the control and treatment arms of a trial are due to chance or to the efficacy of the treatment. These tests report the likelihood of the result being due to chance (null hypothesis) as a probability. Conventionally a probability (*P* value) of 0.05 is used to indicate that a finding is significant rather than occurring by chance. However, statistical tests of significance may not always track important or relevant causal relationships. As Bradford Hill notes:

> [T]here are innumerable situations in which they [significance tests] are totally unnecessary – because the difference is grotesquely obvious, because it is negligible, or because, whether it be significant or not, it is too small to be of any practical importance. What is worse, the glitter of the *t* table diverts attention from the inadequacies of the fare. (1965, p. 299)

In addition, while a *P* value of 0.05 equates to a 95 % certainty that the findings are not an accident, on average, one in every twenty trials with a *P* value of 0.05 is likely to have a finding due to chance. This has prompted the claim that formal tests of significance cannot answer questions about causation (Bradford Hill 1965, p. 299).

The reliability of statistical analysis is a particular problem when it comes to subgroup analysis (Assmann et al. 2000). In a well-designed RCT with sufficiently large participant groups, known confounders are distributed equally between arms or can be adjusted for. However, this may not apply to subgroups within a trial, where the analysis is more likely to be affected by both known confounders (which may be unequally distributed within subgroups even if distributed equally in the trial at large) and unknown confounders which exert greater influence within the smaller sample size of subgroups. Members of subgroups are also unlikely to be allocated evenly across the arms of the trial, unless a modified randomization strategy involving blocking, stratification, or other techniques to balance the arms is used (Pocock and Simon 1975), while methods to control for known confounders are not always used (Assmann et al. 2000).

Even if epidemiological research is able to establish causal relationships or to demonstrate that such relationships are likely enough to safely proceed on the assumption that they pertain, the nature of these causal claims is philosophically puzzling. The measures of causal strength identified by epidemiologists are, mathematically speaking, calculations of the degree of association between two variables. In the absence of an extra ingredient, the *causal import* of such measures is unclear.

This is known as the "causal interpretation problem" (Broadbent 2013, p. 30). Researchers can establish (with a reasonable degree of certainty) that there is a causal relationship between the treatment and the outcome. However, the mechanism of causation cannot be explained by epidemiological studies, including the best designed RCTs, because these studies look at population level measures rather than the physiological effects of the intervention on individual patients. Thus, epidemiological causation is like a "black box" (Howick 2011, p. 124). In contrast, pathophysiological reasoning looks inside the black box at the mechanisms which determine how and why a treatment works.

Applying the results from RCTs to individual patients is also problematic. While RCTs are ideal for establishing that a treatment "works somewhere," they cannot establish whether a treatment "will work for us" in specific settings (Cartwright 2011, p. 1401). The situations where EBM is used are not necessarily similar in the relevant ways to the contexts where the RCTs were performed. There is no straightforward way of moving from the general probabilistic findings supported by RCTs to the particular knowledge required in clinical contexts, due to the complexity of causal connections and the challenge associated with working out how different factors are causally interacting to bring about the positive outcomes for some of the participants in the study (Cartwright 2011, p. 1401).

This problem has two parts: first, whether the target population (e.g., patients in a particular clinical context) is relevantly similar to the RCT populations and, second, whether the practical implementation of the intervention is relevantly similar to that of the RCT (Cartwright 2010). In order to resolve these problems, it is important to understand the mechanisms or causal capacities, which underlie and explain the regular connections between treatment and outcome that pertain in the RCT. Understanding the physiological underpinnings of treatment can provide a basis for identifying which patients will benefit from the application of a treatment that has been shown to work by RCTs and which patients for whom the treatment will not work (Cartwright 2011). However, the current conception of EBM excludes consideration of the causal model that informed the RCT, and information on the implementation conditions of the trial(s) may be incomplete.

A final epistemological concern regarding EBM is that there is no reliable evidence that EBM works. That is to say, it is not possible to randomize clinicians to either practice or not practice EBM and then compare patient outcomes in the two groups, nor is it possible to randomize health systems to implement or not EBM policies (Hayes 2002; Cohen et al. 2004). This criticism might seem unfair, but it is far from obvious what the overall impact of EBM is. Given the extent of change in healthcare and health research in the past 20 years, no meaningful historical comparison can be made. Furthermore, the costs associated with implementing EBM are significant, including the redesign of medical training and degree programs, the production of evidence-based guidelines, and provision of access to evidence for clinicians. Therefore, in order to represent value to health services, EBM should be significantly better than other alternatives.

Ethical Justifications for EBM

The ethical justifications for EBM are straightforward: EBM is grounded in widely shared assumptions about the value of health and the need to use the most effective means possible to protect health. Insofar as EBM is the most effective means, using it will lead to better health outcomes for patients (Gupta 2003). Thus, EBM is consistent with the ethical principles of beneficence and non-maleficence. In addition, at least some models of EBM incorporate patient values and preferences into decision making, thereby respecting patient autonomy. Finally, EBM has the potential to support equity in access to effective interventions and to minimize waste of health resources through the abandoning of ineffective or harmful interventions.

EBM is beneficence-based because it aims to ensure that the knowledge by which decisions are informed is the best (most reliable) possible. EBM thereby offers a scientific foundation for the implicit promise that the doctor does indeed know best about the effectiveness of possible treatment options. By using EBM, doctors are able to explain and justify their recommendations and to offer objective reasons for recommending one treatment rather than another. The implicit ethical claim of EBM is that it will lead to better patient outcomes than by using clinical expertise alone. This claim is supported by examples in the literature, such as that of the slow uptake of antenatal steroids for reducing the severity of lung disease in premature babies. By 1981, there was sufficient research evidence to demonstrate that the use of steroids significantly reduced infant mortality; however, this information had not been systematically collected or promulgated. As a result, steroids were not routinely used and research continued until 1995, leading to the preventable deaths of thousands of babies (Howick 2011, p. 163).

The use of EBM to identify and discard harmful or ineffective treatments meets the ethical requirement of non-maleficence. Using techniques of systematic review and meta-analysis, it is possible to discriminate between effective, ineffective, and harmful treatments. This information is essential for informing healthcare at the level of the individual patient and also for policy makers who may then decommission treatments that are harmful or ineffective. Treatments that had initial plausibility, such as the prophylactic use of anti-arrhythmic drugs in patients post-myocardial infarction or ligation of the internal mammary artery for angina, were later found to be harmful through the use of RCTs. EBM provides a systematic way of collecting and reviewing evidence, thereby increasing the likelihood that harmful or ineffective interventions will be identified and withdrawn.

Since the mid-1990s, models of EBM have included patients' values or preferences alongside research evidence. EBM proposes a transparent and open approach to decision making, in which evidence is used to inform the patient's choice about preferred treatment options. To be autonomous, decisions should be informed to the extent possible. Thus, EBM supports patient autonomy and informed consent, by its transparent approach to evidence. Yet just how patient preferences should be incorporated along with evidence and other relevant information into the decision process is unclear. One approach has been the development and use of patient decision aids

to bring evidence into the consultation in ways that support patient autonomy (Edwards and Elwyn 2001).

Finally, EBM has implications for justice. EBM is committed to the rigorous evaluation of research evidence and applying the findings to all relevant patients. This approach has the potential to foster equity in access to medical treatment by mandating the same effective treatment for patients, irrespective of irrelevant features, such as race. Such an approach can reduce discrimination when efficacious treatments are given equally to all relevant patients. And there are examples of EBM reducing discrimination. An EBM guideline on hemodialysis, for example, significantly increased access to dialysis for African-American men, who, prior to its introduction, had a 60 % greater likelihood of receiving inadequate hemodialysis compared with whites. After the guideline was introduced, there was a 92 % increase in the proportion of African-American patients receiving adequate hemodialysis (Owen et al. 2002). As well as more equitable access at the level of individual patients, EBM has been used, for example, in the UK, to mandate the fair distribution of effective interventions through evidence-informed policy and allocation decisions at the population level. This can lead to more transparent and fairer purchasing decisions and address inequities in access caused by variable provision of interventions across geographical regions.

Ethical Critiques of EBM

Despite the clear ethical foundations of EBM, it has been subject to sustained ethical critique. Concerns fall into a number of areas. First, as discussed above, there is no strong evidence that EBM leads to better health outcomes than alternative approaches to medical decision making, making it uncertain as to whether or not EBM is beneficent. This concern is amplified by a series of related worries about the effects of EBM upon research and the reliability of research results. Somewhat ironically given bias-reducing claims about RCTs, several kinds of bias have been identified in the research upon which EBM is based, including in the conduct of trials, the publication of results, and the choice of interventions for investigation. There are also ethical questions about the effects of EBM on the treatment of participants in clinical trials and about the broader impact of EBM on the research agenda. A second set of concerns relates to the use of EBM in practice where it may be used to mandate or withhold treatment, often through the use of guidelines. Patient (and practitioner) autonomy may be marginalized if there are incentives or penalties linked to compliance with EBM guidelines. Finally, there are concerns about the broader societal effects of EBM, such as on health equity, and in entrenching particular kinds of medical authority at the expense of other forms of expertise.

Although EBM aims to promote the use of research methods that minimize bias, a number of biases have been identified that affect the reliability of research and thus may lead to EBM incorrectly identifying interventions as more effective and/or less harmful than they really are (Gupta 2003; Every-Palmer and Howick 2014). Some of

these biases relate to the funding of clinical trials, while others arise from the nature of the underlying research questions. Conflicts of interest arising from the commercial funding of much of the research underpinning EBM are a major source of bias. There is now strong evidence that research funded by commercial sources is three to four times more likely to return positive findings (i.e., show that an intervention is effective) than research funded by noncommercial sources such as governments (Lexchin et al. 2003; De Vries and Lemmens 2006). This is problematic as it is estimated that the private for-profit sector funds 51 % of global research annually (Burke and Matlin 2008), with the suggestion that between two thirds and three quarters of published randomized controlled trials are industry funded (Every-Palmer and Howick 2014).

Bias may arise from manipulation of the study design to produce positive results, for example, by choosing a placebo or suboptimal dose of competitor drug as the comparator or by selecting participants with characteristics that favor the drug under investigation, rather than who reflect the target population for treatment. Outcomes may be selected to favor the trial drug, or statistical methods may be used to minimize or mask adverse events (Rogers and Ballantyne 2009). The trial may be too short to provide a meaningful estimate of efficacy for treatment of chronic conditions. Concealing adverse events has led to considerable preventable morbidity and mortality, resulting in a number of high-profile lawsuits and financial penalties. For example, Merck allegedly concealed evidence about the cardiac side effects of their blockbuster antiarthritic drug rofecoxib, leading to tens of thousands of excess cardiovascular events (Topol 2004), while DePuy Orthopedics (a subsidiary of Johnson & Johnson) used similar misinformation tactics in response to concerns raised about the safety of their metal-on-metal hip replacement prior to its eventual withdrawal from the market (Johnson and Rogers 2014).

A second kind of bias relates to the selective publication of research results. Funders own the results of research that they have sponsored, and they are under no obligation to publish these, especially if the findings are negative for the product under investigation. There are a number of publication practices that subvert the fair and transparent communication of research results. First, ghostwriting involves employees or subcontractors of the pharmaceutical industry drafting articles which are then published under the name of established academics (De Vries and Lemmens 2006). Links between the article and the funder may be difficult to identify, despite requirements by academic journals to disclose sources of funding and conflicts of interest. Negative results are suppressed, while positive results are published in results are published in strategic campaigns aimed at flooding the literature and creating a more optimistic record of research outcomes than warranted by the actual findings. For example, a campaign for the antidepressant sertraline involved 55 papers published between 1998 and 2001. These became a significant part of the evidence base about sertraline, drowning out other more critical results (Healy and Cattell 2003). Obtaining the raw data from commercial companies is extremely difficult, leading to initiatives which advocate for the registration of all clinical trials and the publication of all results, irrespective of funding source (AllTrials 2014).

A third kind of bias arises regarding the type of intervention under investigation, known as a technical bias (Gupta 2003). This favors familiar research methods and hence pushes research toward phenomena that we know how to investigate. For example, RCTs are ideally suited to pharmacological interventions, where it is possible to securely blind both participant and researcher through the production of identical-looking drugs (including placebos) for intervention and control groups. However, RCTs are far more difficult for complex interventions, especially those that involve physical modalities. Thus, the evidence base for surgery is much weaker than that for drugs, as it is difficult to mount surgical RCTs and, where the intervention is a change in surgical technique, rather than a new device, it may be difficult to secure funding as there is little potential for the commercialization of new surgical techniques. Likewise, there are far fewer RCTs of alternative and complementary therapies, and it is not clear whether at least some of the outcomes considered important to practitioners or patients are amenable to quantified assessment. Thus, technical bias leads to the skewing of the evidence base toward familiar interventions that are easy to investigate and quantify and away from complex interventions or those with qualitative outcomes. Commercial interests do not cause technical bias, but do amplify its impact.

Apart from introducing and amplifying bias and hence unreliability in research, commercial funding has other adverse effects, one of which is pressure to contain costs. This may lead either to outsourcing of research to contract research organizations (CROs) rather than academic teams within Western countries or to performing research in developing nations. CROs are under market pressures to be economical and produce results. This has led to concerns about the treatment of participants in research in the USA who make a living from serial enrollment in phase 1 pharmaceutical trials. Those involved as "guinea pigs" are often impoverished or indigent and lack protections if there are side effects or long-term consequences from their involvement in research (Elliott and Abadie 2008). Similar concerns about the exploitation of vulnerable populations arise when research is performed in resource-poor settings where participants may otherwise lack access to any healthcare and where the interventions under investigation address the diseases of affluent nations rather than those of most pressing concern within the nations hosting the research (Petryna 2007).

Concern about participants extends to vulnerable groups commonly excluded from research. This leads to an evidence base that is skewed toward the treatment of those represented among trial participants, who, in the latter part of the twentieth century, were predominantly white males (Rogers 2004). Women, ethnic minorities, and other groups perceived to be vulnerable, such as children and prisoners, were excluded from research for reasons including explicit protectionist policies of exclusion, practical considerations of research efficiency and cost, and false assumptions about the irrelevance of sex, gender, and racial differences (Rogers and Ballantyne 2009). Despite regulatory efforts to encourage the inclusion of women and minorities in research, distortions continue, with the persistent underrepresentation of women over the age of 65 in research and the overrepresentation of men in studies of heart disease and colorectal and lung cancer trials (Hutchins

et al. 1999; Murthy et al. 2004). These practices lead to research results that are unable to answer questions about the safety and efficacy of clinical interventions for underserved groups. Issues associated with the exclusion of vulnerable groups from research are only partly explained by commercial interests. While companies may want to avoid the costs or risks associated with including members of these groups in trials, wider social norms and patterns of access to healthcare also play a significant role.

One final issue regarding the commercial funding of research concerns the research agenda, which, broadly understood, determines what research is conducted and therefore what results will be available to inform EBM. The research agenda reflects a mix of the interests of industry and the governments of predominantly high-income countries, with an ever-increasing role played by commercial research sponsors. This has led to a focus on patentable treatments that address the needs of affluent markets. The problem here is twofold: an emphasis on patentable treatments (e.g., drugs) at the expense of other potentially effective but less-profitable interventions such as non-patentable behavioral and environmental solutions to ill health and a focus on conditions prevalent in rich populations (Trouiller et al. 2002). For example, of 460 trials investigating treatment for osteoarthritis of the knee, 380 (82.6 %) evaluated drugs despite inadequate or absent evidence about the effectiveness of other kinds of interventions (Tallon et al. 2000).

These ethical concerns arise from placing RCTs at the top of the evidence hierarchy. By so doing, EBM has, inadvertently, encouraged commercial interests in research, leading to adverse effects such as distorting the research agenda, limiting methodological diversity, impoverishing the range of interventions under investigation, and introducing biases into clinical research. Taken together, these undermine both the capacity of EBM to provide answers for important and relevant questions about the clinical care of patients and the reliability of evidence about the effectiveness of interventions. In turn, these features call into question the extent to which EBM is indeed beneficent.

A second set of ethical concerns arises in the use of EBM. Clinicians tend to rely upon easily accessible forms of evidence such as formal summaries or clinical practice guidelines (CPGs), rather than performing their own systematic reviews. This raises a number of issues including the applicability of the evidence to the patient in question and the weight accorded to the evidence compared with other factors that might affect decision making such as patient preferences, availability of resources, practitioner responsibilities, or relevant policies.

First, as noted above, RCTs can provide strong evidence of efficacy, but the epistemological strength of a trial may be inversely related to its applicability. In reducing the number of variables in trials, participants are often limited to those with few or no comorbidities. In contrast, many patients in clinical settings have comorbidities, and thus, it can be unclear to what extent evidence from tightly controlled trials applies to them. This leads to a lack of evidence about which interventions are effective for populations routinely excluded from research, such as members of vulnerable groups or those with the poorest health due to comorbidities, leaving them with fewer treatment options than members of

populations more likely to be included in trials (Rogers 2004). This injustice can be exacerbated when proof of efficacy is required to ensure access to treatment (Hope 1995).

Second, while it is widely accepted that patients should be able to exert their autonomy in making decisions about which healthcare interventions to accept and which to reject, patient autonomy may be compromised by rigid adherence to EBM. Use of evidence summaries or CPGs can lead to limited opportunities for exercising autonomy if patients are simply offered the choice of accepting or rejecting the evidence-mandated option. This point is particularly salient given that patient perspectives are largely excluded from the production of evidence through the use of researcher-defined populations, interventions, and end points. Interventions shown to be effective in RCTs may be unacceptable to patients because of the nature of side effects, cost, or inconvenience. Regarding guidelines, professionals and/or politicians choose the topics for guideline development; professional and/or economic interests dominate guideline recommendations; and guidelines are often used to direct rather than inform individual patient care. Thus, while at least some accounts of EBM propose using evidence to inform patient choices, this can be hard to achieve in practice when the evidence itself has been produced with little attention to what might matter for patients, and there are pressures to accept options that are statistically associated with better outcomes in RCTs.

As with patients, at least some practitioners feel that their autonomy and clinical judgment are undermined by overly directive CPGs or evidence summaries, sometimes referred to as "cookbook" medicine. Clinicians argue that following a guideline devalues or discounts their own knowledge of and expertise with their patients in favor of impersonally developed evidence-based recommendations. This is coupled with the fact that there is little information on how to integrate practitioner knowledge with EBM recommendations; and, in at least some practice settings, there may be penalties for not adhering to CPGs, which are used as tools to assess the quality of care. These concerns are amplified if practitioners do not trust the CPGs, either because they are seen as instruments of rationing rather than evidence about most effective care or because of commercial sponsorship of the guideline itself or the underlying research.

A third set of ethical issues relates to the broader societal effects of EBM. As EBM has become more widespread, it has been seen as a tool for rationalizing the provision of healthcare, appealing to the common sense notion that only healthcare known to be effective should be offered to patients. However, as noted above, not all potential recipients of healthcare are represented equally among research participants, and not all types of interventions are amenable to investigation through an RCT; thus, certain patient populations and interventions are likely to be neglected. In theory, this should not be a problem: EBM advocates the use of the best available evidence, and if this is from a cohort study rather than an RCT, then it is nonetheless the best evidence. But in practice, governments and insurers appeal to RCT evidence in decisions about the provision of some interventions rather than others and may justify their decisions on the grounds of justice in the allocation of resources (Gupta 2003). This takes on historic dimensions when new interventions, for which there is

an evidence base, command funding previously allocated to older interventions that lack evidence as judged by EBM standards. A second societal effect of EBM relates to its impact upon medical authority. Those who perform evidence syntheses and who formulate guidelines acquire the power to direct healthcare spending and commission research in various ways, thereby further entrenching medical authority and in the process shifting it away from clinicians and toward epidemiologists (and commercial funders). This concentration of power into the hands of the few with a commitment to the assumptions of EBM further marginalizes others with potentially valuable contributions to setting the research agenda or determining research priorities (Gupta 2003).

These ethical critiques of EBM challenge claim about the value of EBM in patient care and draw attention to the unintentional consequences of favoring RCTs above other research methods. While some are contingent (it would be possible to have research that avoided the biases introduced as a result of commercial pressures), others identify central and ongoing tensions in the philosophy of EBM surrounding how to use evidence in clinical decisions.

The Future of EBM

Critical analysis of the challenges facing EBM has stimulated recent campaigns for an EBM "renaissance" and for changes to the ways trials are funded and reported. Meanwhile, the rise of personalized medicine poses both challenges and opportunities for EBM.

The EBM renaissance group was developed after a meeting between EBM critics and proponents in Oxford in late 2013, leading to a call for a return to "real" EBM. "Real" EBM designates a version of EBM in which the care of the patient is the highest priority and the use of evidence is individualized to meet patient needs (Greenhalgh et al. 2014). In order to achieve this, the focus of clinician training must shift away from following templates, rules, and guidelines toward enhancing the higher-level intuitive skills that are markers of true expertise; and publications (in journals and by guideline groups) must be better attuned to the needs of those who will be reading and using them. Finally, research must change to become more independent and free from conflicts of interest and broader in the scope of methods recognized as providing high-quality evidence.

Demand for a broader research agenda incorporating different research questions and methodologies challenges the current form of EBM, in which one dominant methodology, the RCT, bears the overwhelming burden of producing high-quality medical knowledge. Although the "renaissance group" retains the idea of gold standard systematic reviews (Greenhalgh et al. 2014), some of their recommendations for a broader research agenda seriously challenge current understandings of EBM, due to the considerable influence of evidence hierarchies on the funding and publication of research.

In recognition of the biases introduced by commercially funded research, especially selective publication of results, the AllTrials campaign calls for the prospective

registration of all trials together with publication of a brief summary of the results within 12 months of completion of the trial and publication of full details about the trial's methods and results. AllTrials campaign documents and publications identify ways in which existing measures are failing, such as journals continuing to publish unregistered trials and the non-enforcement of FDA requirements for all trials located in the USA to be registered (AllTrials 2013; Goldacre 2013). AllTrials advocates for additional initiatives to improve the transparency of clinical trials and availability of data, including research contracts that do not allow companies to veto publication of the results, research ethics committees requiring publication of results as a condition of ethics approval, and treating the withholding of trial results as medical misconduct (AllTrials 2013, p. 5). All of these measures would, if implemented, go some way to addressing current shortcomings in the production of evidence.

The rise of personalized medicine poses challenges for EBM (Hamburg and Collins 2010). The notion of tailoring medical treatments to individual patients contrasts significantly with the epidemiological methods of EBM. Methods for deriving data from trials that might be useful for personalized medicine include stratified randomization, for example, randomization of cancer patients according to molecular information about their tumors. However, such stratification can reduce the effectiveness of blinding (Pocock and Simon 1975). The more strata that are introduced, the more challenging this issue becomes. A simpler form of "personalized" medical research is known as the *n*-of-1 trial. These are crossover trials with only one participant, in which the active and control arms are run sequentially to generate comparative data for individual patients. These trials can retain many of the advantages of well-designed RCTs (including the option of a placebo or active control and double blinding with some interventions), but rather than generating statistical data about a population, they generate personal data about the effectiveness and side effects of a treatment in an individual. In contexts where population level data is needed, multiple *n*-of-1 trials can be combined; in the future, combining the results from many such trials could be facilitated by sophisticated patient record-keeping systems (Lillie et al. 2011).

Conclusion

Since the early 1990s, EBM has changed the way that clinicians engage with the findings of medical research and apply these findings in the care of individual patients. EBM has influenced the priorities of researchers, policy makers, funders, and commercial entities involved in health research. A fundamental tenet of EBM is that certain forms of evidence, specifically that derived from well-designed RCTs and systematic reviews of RCTs, are the most reliable and should be privileged in the decision making of clinicians and policy makers. The evidence hierarchy has created incentives for researchers and funders to prefer RCTs over other research methods and encouraged clinicians and policy makers to eschew evidence deemed less reliable based on pathophysiological reasoning and expert opinion. Benefits of EBM include the generation of systems for publishing, retrieving, and summarizing

data generated from health research, as well as improved uptake of at least some research findings about effective and ineffective treatment. Although there are both epistemological and ethical justifications for EBM, significant critiques have arisen from both these perspectives. While it is unlikely that the quest for better evidence about healthcare interventions will be abandoned, the exploitation of EBM by commercial interests, and the impact of EBM in narrowing the research agenda, may lead to a more or less radical restructuring of the EBM hierarchy and the way that research is performed and evaluated.

Definitions of Key Terms

Clinical practice guidelines (CPGs)	Are tools to support clinical decision making by providing recommendations based upon syntheses of evidence.
Confounders	Are any factors that influence the outcomes of a research trial, other than the intervention under investigation.
Efficacy	Is a measure of the clinically beneficial outcome resulting from an intervention as measured in a clinical trial.
Epidemiology	Is the study of the determinants and distribution of diseases in human populations, often using group comparisons.
Evidence hierarchy	Is a method of ranking evidence derived from different sources, based on the view that the most reliable evidence is generated by research methods, such as randomized controlled trials, that minimize bias.
Evidence-based medicine (EBM)	Refers to the use of the best available scientific evidence to inform the clinical care of patients.
Mechanistic reasoning	Is inferring the likely effects of a therapy based upon an understanding of the relevant physiological mechanisms.
Meta-analysis	Is a statistical method for combining the results of multiple clinical trials in order to identify relevant outcomes that are not reliably discernable within individual trials.
Pathophysiological principles	Are generalized rules based upon an understanding of the relevant physiological processes. They may inform mechanistic reasoning.
Randomized controlled trials (RCTs)	Are a type of experimental study designed to minimize bias by randomly allocating participants to either active (receiving the experimental intervention) or control (receiving the standard therapy or placebo) arms of the trial.
Rules of evidence	Are agreements about the reliability of statistically identified associations in research.

Statistical significance	Refers to the likelihood of an association being due to the efficacy of the treatment rather than chance. Conventionally, results are considered statistically significant if the probability of them occurring by chance is less than 5 %.
Systematic reviews	Are a summary of evidence on a particular clinical question, based upon all of the available research evidence and using pre-specified methods aimed at reducing bias.

Summary Points

- Evidence-based medicine is a highly influential approach to using the results of research, the "best evidence," to inform the clinical care of patients.
- EBM differs from previous approaches to medical evidence by relying upon statistical analyses to determine efficacy, rather than reasoning based upon pathophysiological principles or causal mechanisms.
- The major epistemological innovation of EBM is to change thinking about the reliability of different forms of evidence used to inform practice. In EBM hierarchies, results from randomized controlled trials are ranked as the most reliable form of evidence, while clinical experience is considered the least reliable form of medical evidence.
- The ethical foundation of EBM is that it promises better healthcare by reliably distinguishing effective from ineffective or harmful interventions.
- Weaknesses of EBM include the inability of RCTs to rule out all potential confounders, the probabilistic nature of the causal relationship between interventions and outcomes demonstrated by RCTs, the undervaluing of mechanistic reasoning, and the difficulty of applying the results of trials to individual patients.
- Ethical critiques of EBM identify the effects of conflicts of interest caused by the commercial funding of much of the research used to inform EBM, which can introduce bias; the effect of EBM on the research agenda, skewing it toward particular research methods and interventions; and effects on the autonomy of decision making in the clinical encounter.
- EBM may be strengthened by increasing the range of interventions for which evidence is sought, and the methods so used, and by freeing itself from the effects of conflicts of interest.

References

AllTrials (2013) Missing trial data – briefing notes. http://www.alltrials.net//wp-content/uploads/2013/01/Missing-trials-briefing-note.pdf. Accessed 20 Aug 2014

AllTrials (2014) http://www.alltrials.net/. Accessed 20 Aug 2014
Assmann SF, Pocock SJ, Enos LE et al (2000) Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 355:1064–1069
Bradford Hill A (1965) The environment and disease: association or causation? Proc R Soc Med 58 (5):295–300
Broadbent A (2013) Philosophy of epidemiology. Palgrave Macmillan, New York
Burke M, Matlin S (eds) (2008) Monitoring financial flows for health research 2008. Geneva: Global Forum for Health Research. http://www.globalforumhealth.org/filesupld/MFF08/MonitoringFinancialFlows2008.pdf
Canadian Task Force on the Periodic Health Examination (1979) The periodic health examination. Can Med Assoc J 121:1193–1254
Cartwright N (2010) What are randomised controlled trials good for? Philos Stud 147(1):59–70
Cartwright N (2011) A philosopher's view of the long road from RCTs to effectiveness. Lancet 377:1400–1401
Charles C, Gafni A, Freeman E (2011) The evidence-based medicine model of clinical practice: scientific teaching or belief-based preaching? J Eval Clin Pract 17(4):597–605
Cohen AM, Stavri PZ, Hersh WR (2004) A categorisation and analysis of the criticisms of evidence-based medicine. Int J Med Inform 73:35–43
Davidoff F, Haynes B, Sackett D et al (1995) Evidence-based medicine. BMJ 310 (6987):1085–1086
De Vries R, Lemmens T (2006) The social and cultural shaping of medical evidence: case studies from pharmaceutical research and obstetric science. Soc Sci Med 62(11):2694–2706
Djulbegovic B, Guyatt GH, Ashcroft RE (2009) Epistemologic inquiries in evidence-based medicine. Cancer Control 16(2):158–168
Edwards A, Elwyn E (eds) (2001) Evidence-based patient choice: inevitable or impossible? Oxford University Press, Oxford
Elliott C, Abadie R (2008) Exploiting a research underclass in phase 1 clinical trials. N Engl J Med 358(22):2316–2317
Every-Palmer S, Howick J (2014) How evidence-based medicine is failing due to biased trials and selective publication. J Eval Clin Pract. doi:10.1111/jep.12147
Evidence-Based Medicine Working Group (1992) Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA 268(17):2420–2425
Goldacre B (2013) Are clinical trial data shared sufficiently today? No. BMJ 347:f1880
Goodman KW (2003) Ethics and evidence-based medicine: fallibility and responsibility in clinical science. Cambridge University Press, Cambridge
Greenhalgh T, Howick J, Maskrey N (2014) Evidence based medicine: a movement in crisis? BMJ 348:g3725
Gupta M (2003) A critical appraisal of evidence-based medicine: some ethical considerations. J Eval Clin Pract 9(2):111–121
Guyatt G, Oxman A, Vist G et al (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336:924–926
Hamburg MA, Collins FS (2010) The path to personalised medicine. N Engl J Med 363(4):301–304
Hardaway RM (2004) Wound shock: a history of its study and treatment by military surgeons. Mil Med 169(4):265–269
Hayes RB (2002) What kind of evidence is it that evidence-based medicine advocates want health care providers and consumers to pay attention to? BMC Health Serv Res 2(3). doi:10.1186/1472-6963-2-3
Healy D, Cattell D (2003) Interface between authorship, industry and science in the domain of therapeutics. Br J Psychiatry 183:22–27
Hope T (1995) Evidence based medicine and ethics. J Med Ethics 21(5):259–260
Howick J (2011) The philosophy of evidence based medicine. Wiley Blackwell, Oxford
Howick J, Glasziou P, Aronson JK (2013) Problems with using mechanisms to solve the problem of extrapolation. Theor Med Bioeth 34(4):275–291

Hutchins LF, Unger JM, Crowley JJ et al (1999) Under representation of patients 65 years of age or older in cancer-treatment trials. N Engl J Med 341(27):2061–2067
Johnson J, Rogers W (2014) Joint issues – conflicts of interest, the ASR hip and suggestions for managing surgical conflicts of interest. BMC Med Ethics 15:63. doi:10.1186/1472-6939-15-63
Lexchin J, Bero LA, Djulbegovic B et al (2003) Pharmaceutical industry sponsorship and research outcome and quality. BMJ 326:1167–1170
Lillie EO, Patay B, Diamant J et al (2011) The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? Pers Med 8(2):161–173
Moher D, Liberati A, Tetzlaff J et al (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339:332–336
Murthy VH, Krumholz HM, Gross CP (2004) Participation in cancer clinical trials: race-, sex-, and age-based disparities. JAMA 291(22):2720–2726
OCEBM Levels of Evidence Working Group (2001/2010) Levels of evidence. Brit J Urol 105:155
OCEBM Levels of Evidence Working Group (2011) The Oxford 2011 levels of evidence. Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653. Accessed 11 Aug 2014
Owen W, Szczech L, Frankenfield D (2002) Healthcare system interventions for inequality in quality: corrective action through evidence based medicine. J Natl Med Assoc 94:83S–91S
Petryna A (2007) Clinical trials offshored: on private sector science and public health. BioSocieties 2:21–40
Pocock SJ, Simon R (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31(1):103–115
Rogers WA (2004) Evidence-based medicine and justice: a framework for looking at the impact of EBM on vulnerable or disadvantaged groups. J Med Ethics 30:141–145
Rogers WA, Ballantyne AJ (2009) Justice in health research: what is the role of evidence-based medicine? Perspect Biol Med 52(2):188–202
Rosenberg W, Donald A (1995) Evidence based medicine: an approach to clinical problem-solving. BMJ 310(6987):1122–1126
Sackett DL (1989) Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 95(2 Suppl):2S–4S
Sackett DL, Rosenberg WMC, Gray JAM et al (1996) Evidence based medicine: what it is and what it isn't. BMJ 312(7023):71–72
Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 152(11):726–732
Sehon SR, Stanley DE (2003) A philosophical analysis of the evidence-based medicine debate. BMC Health Serv Res 3:1–10
Tallon D, Chard J, Dieppe P (2000) Relation between agendas of the research community and the research consumer. Lancet 355:2037–2040
Topol EJ (2004) Failing the public health – rofecoxib, merck, and the FDA. N Engl J Med 351 (17):1707–1709
Trouiller P, Olliaro P, Torreele E et al (2002) Drug development for neglected diseases: a deficient market and a public-health policy failure. Lancet 359(9324):2188–2194
Worrall J (2007) Why there's no cause to randomize. Br J Philos Sci 58:451–488
Worrall J (2011) Causality in medicine: getting back to the hill top. Prev Med 53:235–238
Wyer PC, Silva SA (2009) Where is the wisdom? I – a conceptual history of evidence-based medicine. J Eval Clin Pract 15(6):891–898

Randomized Trials and Observational Studies: The Current Philosophical Controversy

53

Jeremy Howick and Alexander Mebius

Contents

Abstract

The supposed superiority of randomized over non-randomized studies is used to justify claims about therapeutic effectiveness of medical interventions and also inclusion criteria for many systematic reviews of therapeutic interventions.

J. Howick (✉)
Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
e-mail: jeremy.howick@phc.ox.ac.uk

A. Mebius
Royal Institute of Technology (KTH), Stockholm, Sweden

Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK
e-mail: alenik@kth.se; alexander.mebius@phc.ox.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_45

However, the view that randomized trials provide better evidence has been challenged by philosophers of science. In addition, empirical evidence for average *differences* between randomized trials and observational studies (which we would expect if one method were superior) has proven difficult to find. This chapter reviews the controversy surrounding the relative merits of randomized trials and observational studies. It is concluded that while (well-conducted) observational can often provide the same level of evidential support as randomized trials, merits of (well-conducted) randomized trials warrant claims about their superiority, especially where results from the two methods are contradictory.

Introduction

A current and widely accepted view in medicine is that randomized studies are superior to non-randomized studies to support claims about therapeutic effectiveness (Higgins and Green 2008; Straus et al. 2011). This view is also sometimes used to as justification for excluding observational studies from systematic reviews (Higgins and Green 2008) and for making judgments about risk of bias in studies (Guyatt et al. 2008; OCEBM 2011). However, philosophers of science have criticized this view. For example, in the most widely cited critique of evidence-based medicine (EBM), John Worrall claims that randomization adds little epistemological value (Worrall 2002). Yet, more recently a systematic review concluded that there were no average differences between randomized trials and observational studies (which is contrary to what one would expect if one method were superior) (Anglemyer et al. 2014). The lack of average differences calls into question whether one method can be superior.

The aim of this chapter is to review current philosophical controversy. The following section begins with a brief description of the differences between randomized trials and observational studies and a warning about how to compare the two designs fairly (i.e., we must compare randomized trials and observational studies of similar *quality*). Section "A Note About Adequate Comparisons of Randomized Trials and Observational Studies" reviews some of the arguments that have been made against the view that randomized trials provide superior evidence, including some from Worrall (2002). The arguments for and against Worrall's view will not be reiterated here (but see Howick (2011) for an argument that Worrall sets up a straw man and La Caze et al. (2012) for an argument that Worrall's premise about the value of randomization is based on a misunderstanding). Section "Potential Versus Actual Benefits of Randomized Trials: The Elusive Search for Empirical Evidence That Randomized Trials and Observational Studies Provide Different Results" examines the implications of the recent systematic review that failed to detect an average difference between randomized trials and observational studies. Section "Internal Versus External Validity of Randomized Trials" reviews whether the external validity is more problematic for randomized trials than it is for observational studies. The chapter finishes with a brief review of some of the alleged practical advantages of observational studies (such as their ethical feasibility) and argues that many of these

are exaggerated. It concludes that well-conducted observational studies will often provide similar results as randomized trials. However, in the cases where there are differences between randomized trials and observational studies, accepted benefits of randomized trials suggest that we should side with the randomized trial. Moreover, since it is not possible to know in advance whether the results from observational studies and randomized trials will differ, new interventions should be introduced in the context of well-conducted randomized trials.

Randomized Trials and Observational Studies

Observational Studies

In controlled observational studies, investigators compare people who are subject to an intervention with those who are not. The investigators neither allocate patients to receive the intervention nor administer the intervention. Instead, they compare records of patients who have received an intervention and been treated in routine practice with similar patients who did not receive the intervention. The main problems with observational studies are that they suffer from (i) self-selection bias (sometimes called patient preference bias, or confounding by indication), (ii) adherence bias, (iii) allocation bias, and (iv) performance bias. We will explain each of these in turn.

In one typical observational study, Petitti et al. (1987) compared the records of 2,656 women who took hormone (estrogen) replacement therapy (HRT) with 3,437 who did not and followed them for 10 or more years to measure rates of coronary heart disease (CHD) and overall mortality. They found that HRT users were only half as likely to die as those who did not use HRT. Stampfer and Colditz (1991) conducted a systematic review of all the available studies of the effects of HRT in preventing CHD, all but one of which were observational, and found that women taking HRT appeared to be, on average, half as likely to die as women who did not take HRT. They concluded that HRT could substantially reduce the risk for coronary heart disease. As a result, many thousands of women were given HRT to help prevent coronary heart disease. However, later randomized studies on the effect of HRT found that far from preventing coronary heart disease, stroke, and cancer, it appeared to increase the risks of developing these conditions (in addition to others including dementia) for quite a number of women (Rossouw et al. 2002). Why this apparent anomaly?

The observational studies of the effects of HRT, like all observational studies, suffered from the problem that people who choose to take HRT are likely to be very different in many ways from people who choose not to do so ("self-selection bias"). There might, for example, be significant differences in age; behavior, such as whether or not they smoked, how often they ate vegetables, and how much alcohol they drank; their working conditions; or where they lived. These differences could all affect how likely people are to contract CHD or cancer *independent* of whether or not they take HRT. In other words, the differences in the results might not be caused

by the treatment, but are instead due to the better health of the women who chose to take HRT. Such differences between people in the experimental and control groups at the outset of a study and before the treatment is administered (the "baseline") are often referred to as "selection bias." Selection bias arising from patient choice is referred to as self-selection bias. Careful adjusting for baseline differences increases the quality of observational studies. At the same time, some differences will inevitably prove difficult to control for, because they are either unforeseen or information about them cannot be established. It can, for example, be difficult to obtain observational information about comorbidities, concomitant medication, and family history of disease.

Observational studies can also suffer from adherence bias. To see how, consider the following example. In a study of clofibrate versus placebo for treating coronary heart disease, researchers found that there was no difference in mortality between men treated with clofibrate and those in the placebo group (20 % in both groups). However, investigators found that patients who adhered to the treatment regime more strictly had a lower mortality (15 %) than those who did not (25 %) (Coronary Drug Project 1980). A systematic review confirms that adherence seems to be an independent factor that is directly correlated with positive outcomes (Boswell et al. 2012). This could be due to the fact that adherers are more hopeful or that adherence is typical among people who engage generally in more healthy behaviors. Since patients in an observational study choose to take the treatment, they could be more likely than an average patient in a trial to be an adherer, which could confound the study.

Another potential problem with observational studies is allocation bias, which arises because caregivers are in charge of deciding whether or not to prescribe a treatment. Caregivers could systematically favor certain sorts of patients. For example, if they thought the treatment was likely to be very effective, they might choose to give the treatment only to their sickest patients. Alternatively, if they were worried that the treatment had risky side effects that might be more serious for the sickest patients, they might choose to exclude this group.

Several confounding factors can also arise in observational studies because patients and caregivers know they are receiving treatment. Biases arising after the patient has received the treatment are often referred to as "performance biases." If patients believe they are taking a powerful therapy, whether or not the therapy is, in fact, powerful ("performance bias") and if patients know they are receiving the latest and best therapy, they might improve because of their beliefs and expectations and not because of the experimental therapy itself (Di Blasi et al. 2001). Similarly, investigator attitudes have been known to influence interpretation of rat (Rosenthal and Lawson 1964) and human behavior (Eisenach and Lindner 2004) and even to affect more "objective" measures such as blood cell counts (Berkson et al. 1939). Observational studies cannot generally deal with performance biases because (by definition) people taking the therapy know they are taking the therapy. These potentially serious issues warrant the worries with results from observational studies.

Randomized Trials

Randomized trials all involve comparing at least one experimental therapy with at least one control therapy. The control groups can either receive another treatment, a "placebo," or "no treatment." A placebo is a treatment capable of making people believe it is, or could be, the experimental treatment. It is often a sugar pill, although more sophisticated trials have also attempted to use a placebo that mimics the known side effects of the experimental treatment, to avoid patients becoming aware of whether they are in the experimental or control group (Howick 2011). "No treatment" controls are difficult to construct in practice. Participants are either left alone, in which case the investigators lose control over whether the "untreated group" choose to treat themselves with some other treatment, or they are closely monitored, although this is also known to have effects on the outcomes of the treatment or lack of it (Cocco 2009; McCarney et al. 2007).

Unlike in an observational study where patients choose whether to take the intervention themselves, participants in a randomized trial are randomly allocated to receive either an experimental intervention or a control. Simple random allocation is a process in which all participants have the same chance of being assigned to one of the study groups (Jadad 1998). Restricted randomization involves employing various strategies to ensure that the number of participants and various characteristics such as sex and age are similarly distributed between groups.

Strict randomization of participants to treatment and control groups reduces the risk of self-selection bias, adherence bias, and allocation bias, because neither participants nor caregivers can influence who receives the experimental intervention. However, unless the allocation sequence is concealed, randomization can be subverted, so in order for the potential benefits of randomization to be actualized, random allocation must be concealed. Violations of the assignment scheme are particularly dangerous when the investigators have a personal or financial interest in the new therapy appearing to be effective, because they can choose patients whom they think are more likely to benefit. Participants' knowledge of the group to which they are assigned can also corrupt the randomization process.

Randomized trials can also reduce the risk of performance bias. If the trial participants, caregivers (and perhaps also other groups involved in the trial), are blinded and do not know which participants receive the experimental intervention and which participants receive the control intervention, then performance biases can be ruled out. However, trials that are described as blinded in fact are rarely successfully blinded (Howick 2011). Participants who know they are receiving the "mere" control treatment could drop out of the trial or (which is bad for the validity of the trial but perhaps good for the participant) covertly seek other medication. This would tend to inflate the apparent benefits of the control treatment and reduce apparent benefits of the experimental treatment. Others may have read about potential side effects of the new treatment and therefore drop out of the experimental group.

Given that high-quality randomized trials can rule out more confounding factors than observational studies, it is unsurprising that even vociferous critics of the view

that randomized trials provide better evidence acknowledge (including Worrall 2002) acknowledge the superiority of randomized trials. To be sure, Worrall claims that the *only* (and he insists, small) benefit of randomized trials is their ability to reduce "selection bias." This, however, is an understatement, since randomized trials but not observational studies can rule out various sources of performance bias as well.

Comparing Randomized Trials and Observational Studies

A Note About Adequate Comparisons of Randomized Trials and Observational Studies

Comparing "high-quality" or "well-conducted" (more on what this might mean below) randomized trials with shoddily conducted observational studies would not provide a fair basis for comparing the relative merits of the two study designs, nor would a comparison of a carefully controlled observational study with a large effect with a small, biased randomized trial. When comparing randomized trials with observational studies, it is therefore important to compare "high-quality" randomized trials with "high-quality" observational studies. But what does it mean for a study to be "high quality"? Following Worrall (2002), Howick (2011) argues that the quality is related to the extent to which the *effect size* revealed in the study can be taken to account for likely *confounding factors*. A confounding factor (or "confounder") is one that (a) potentially affects the outcome, (b) is unequally distributed between experimental and control groups, and (c) is unrelated to the experimental intervention. Each confounder provides an alternative explanation for the results of the study. For example, age and smoking status are likely confounders in many studies because age and smoking are independently correlated with many important outcomes measured in clinical trials. In other words (on Howick's account), an observational study whose effect size is much larger than the combined effect of potential confounders should provide enough evidence to warrant the use of that treatment in clinical practice (Glasziou et al. 2007). This account has been criticized by Broadbent (Broadbent 2013) who notes that even if the effect size is large, an observational study does not rule out common causes. This is a legitimate criticism; hence, we should add that in addition to demonstrating a large enough effect size to rule out confounding, an observational study needs to demonstrate that a common cause is an unlikely explanation for the association.

For example, an observational study showing that high doses of vitamin C made common cold symptoms disappear within 5 days supports the hypothesis that vitamin C cures the common cold, but does *not* rule out the plausible alternative hypothesis that the common cold symptoms go away without any treatment within 5 days. Henceforth in this chapter when referring to randomized trials or observational studies, we refer to examples of each that are high quality.

Philosophers of science (Borgerson 2009) as well as medical researchers (Altman 2002) have criticized randomized trials on the basis that many randomized trials are

not high quality. For example, one particular randomized trial of treatments for sepsis suggested that using the monoclonal antibody to the endotoxin could cut mortality in half (Ziegler et al. 1991), but a subsequent trial, also randomized, but tenfold bigger, found that the same antibody could increase mortality (McCloskey et al. 1994). Many randomized trials are underpowered (Keen et al. 2005) and fail to successfully conceal or blind (Schulz et al. 1995; Wood et al. 2008; Schulz and Grimes 2002), which makes them susceptible to selection bias and allocation bias. Finally, they may also have effect sizes so small that statistically significant results can arise by chance (Sierevelt et al. 2007) or suffer from confounding from other sources that are not adequately explored (Smith and Ebrahim 2002; Barbui and Cipriani 2007). However, observational studies suffer from many of the same problems as randomized trials and perhaps more (Stroup et al. 2000). In order for the fact that many randomized trials are poorly conducted to count against the view that randomized trials provide better evidence than observational studies, one would have to show that randomized trials are *more likely* to suffer from bias than observational studies. The fact that some poorly conducted randomized trials should be interpreted with more suspicion than high-quality observational studies with large effects is well taken and also incorporated into common evidence-ranking schemes (Guyatt et al. 2008; OCEBM 2011).

Potential Versus Actual Benefits of Randomized Trials: The Elusive Search for Empirical Evidence That Randomized Trials and Observational Studies Provide Different Results

The above discussion illustrates that randomized trials have the potential to rule out numerous biases that threaten the validity of observational studies. However, empirical research supporting *actual* differences between the two study designs has proven hard to come by. A recent Cochrane Review summarized previous systematic reviews that compared results from observational and randomized trials (Anglemyer et al. 2014). The reviewers found that some randomized studies reported larger effect sizes than observational studies of the same treatment, while others had smaller, but often similar, effect sizes. On average Anglemyer et al. reported that there was no average difference between randomized controlled trials and observational studies. In their words: "there is little evidence for significant effect estimate difference between observational studies and RCTs, regardless of specific observational study design, heterogeneity, or inclusion of studies of pharmacological interventions" (Anglemyer et al. 2014, p. 2).

However, it is unclear whether Anglemyer et al.'s conclusion was acceptable. Specifically, one could challenge their decision to pool the results. The Cochrane Handbook cautions against pooling results if effect directions differ:

> A systematic review need not contain any meta-analyses. . .particularly if there is inconsistency in the direction of effect, it may be misleading to quote an average value for the intervention effect. (Higgins and Green 2008, p. 279)

In fact a similar Cochrane Review chose not to pool and drew opposite conclusions. Odgaard-Jensen et al. compared trials in which randomization was adequately described with trials in which randomization was inadequately described. Their results were similar to those in the Anglemyer et al. review: some adequately randomized studies reported larger effect sizes than inadequately randomized studies, while others reported smaller, but again often similar, effect sizes (Odgaard-Jensen et al. 2011). However, instead of reporting an average result, Odgaard-Jensen et al. concluded that "results of controlled trials with adequate and inadequate/unclear concealment of allocation sometimes differed. . .However, it is not generally possible to predict the magnitude, or even the direction, of possible selection biases and consequent distortions of treatment effects from studies with non-random allocation or controlled trials with inadequate or unclear allocation concealment" (Odgaard-Jensen et al. 2011, p. 10).

We will not go into detail here about whether the decision to pool or not was correct (see Howick and Mebius 2014, for a more complete discussion). What we can conclude from the results of the Odgaard-Jensen et al. and Anglemyer et al. reviews is that (adequate) randomized trials often provide similar results to inadequately or non-randomized studies. This is not surprising if we consider that if a treatment has a real (and moderate) effect, the effects will tend to show up in both well-conducted observational studies and well-conducted randomized trials. However, in some cases results from randomized trials and observational studies *do* differ, and in these cases we need to know which studies to trust. Given the ability of randomized trials to rule out biases (something that even skeptics admit), then ceteris paribus it is safe to side with the results from the randomized trial. In fact this is what happens in practice. For example, observational studies suggest that high doses of vitamin C reduce the risk of cardiovascular disease (Knekt et al. 1994), homeopathy reduces the risk of depression (Oberai et al. 2013), and metformin reduces the risk of cancer among patients with diabetes (DeCensi et al. 2010). However, these results have been contradicted by randomized trials and therefore have not been accepted (Sesso et al. 2008; Adler et al. 2013; Stevens et al. 2012). Our challenge to philosophers of science who criticize the view that randomized trials provide better evidence than observational studies is simple. Let them provide just one single example where results from well-conducted randomized trials are different from results from observational studies and where they would side with the observational study.

More recently, and in a story that was widely reported in the news, a 2006 Cochrane Review of randomized trials suggested that neuraminidase inhibitors ("Tamiflu") reduced the risk of, and could cure, swine flu (Jefferson et al. 2006). On the basis of that review, many countries stockpiled billions of dollars' worth of the drugs. However, the authors of the review suspected that the manufacturer had not published all the relevant trials. After a hard-won fight and appeal to the Freedom of Information Act, the authors of the original review obtained access to all the trials, whether published or not, and updated the review, where they found little evidence of benefit and strong evidence of harms from neuraminidase

inhibitors (Jefferson et al. 2014). At this stage the drug manufacturers suggested that the trials were unreliable because they tested the effects of the drugs in artificial (trial) conditions rather than "real world" (i.e., in an observational study) (Muthuri et al. 2014).

Internal Versus External Validity of Randomized Trials

Both Worrall (2002) and Cartwright (2007) argue that even if randomized trials have a higher degree of *internal validity* (the degree to which the study results of the study apply to the study population) than observational studies, they suffer from problems of external validity (the degree to which the study results apply to a "real world" or "target" population). Cartwright and Worrall are correct to draw our attention to the problem of external validity. Up to 90 % of potentially eligible participants are sometimes excluded from trials according to often poorly reported and even haphazard criteria (Mant 1999; Penston 2003; Zimmerman et al. 2002; Zetin and Hoepner 2007). For example, even the most effective antidepressants in adults have doubtful effects in children (Bylund and Reed 2007; Deupree et al. 2007). In another example taken from John Worrall (2007), the drug benoxaprofen (Oraflex™ in the USA and Opren™ in Europe) proved effective in trials in 18–65-year-olds, but killed a significant number of elderly patients when it was introduced into routine practice.

However, to infer from the problems with the external validity of randomized trials to any claim about the comparative benefits of randomized trials compared with observational studies is invalid. For one, if a study is not internally valid, then the issue of external validity is moot. Second, one would have to establish that observational studies have a *higher* degree of external validity than randomized trials to infer from the (alleged) relative lack of external validity of randomized trials. This assumption is taken for granted by Worrall and Cartwright and other philosophical critics of EBM, but they do not cite any evidence to support it. In fact scientists who actually do observational studies worry very much about the external validity of observational studies which themselves have inclusion criteria that can be very unrepresentative (Carlson and Morrison 2009). Doll and Hill's famous observational study of smokers, for example, was limited to doctors (who, at the time, were almost exclusively male). The people in this famous observational study were therefore very different from the general population. Moreover, many randomized trials ("pragmatic" trials) include almost all of the target population. For example, the GISSI-1 trial of thrombolysis for acute myocardial infarction recruited 90 % of patients admitted within 12 h of the event with a definite diagnosis and no contraindications (GISSI 1986): in other words, most of the people who would have been treated in practice. Third, neither Worrall nor Cartwright cites any empirical evidence that the alleged lack of representativeness of randomized trial populations is a real problem. Studies indicate that even if randomized trials appear to involve unrepresentative populations, the results generally apply to the target population (Vist et al. 2008).

Alleged (But Rarely Real) Relative Practical Advantages of Observational Studies

Numerous other alleged disadvantages of randomized trials are often used to challenge the view that randomized trials provide superior evidence. For example, it is often argued that randomized trials are sometimes unfeasible or unethical (McCulloch et al. 2002). This is true. For example, it would have required too large a sample size (and jumping through all but insurmountable ethical hurdles) to conduct a randomized trial that challenged Dr. Spock's advice to put babies to sleep on their stomachs. Hence, a number of very large observational studies were conducted that suggested (contrary to what Dr. Spock's mechanistic reasoning suggested) more babies who slept on their *backs* survived. However, it does not follow from the fact that randomized trials are sometimes unfeasible to the fact that they do not rule out more bias than observational studies *in cases where they are feasible*. Moreover, claims that randomized trials are unfeasible or unethical are often exaggerated. For instance, it is often claimed that randomized trials of surgical procedures are unfeasible (because surgeons have strong preferences) and unethical (because "control" or "sham" surgery usually involves incisions and anesthesia which are harmful). However, a recent systematic review identified 53 placebo-controlled trials of surgery, and in over half the "placebo" surgery was as good as the "real" surgery (Wartolowska et al. 2014). This result turns the ethical argument about randomized trials on its head: if the "placebo" surgery is as good as the "real" surgery, then it is arguably unethical to *not* conduct placebo-controlled randomized trials of surgical interventions. And given they have been done, they are also clearly feasible.

Another oft-heard argument is that industry interests influence results of randomized trials in various ways (Every-Palmer and Howick 2014). This is true. However, the very same industry influences also corrupt results from observational studies, mechanistic reasoning, and expert "consensus" statements (Jones et al. 2014). So again, one cannot infer from the fact that randomized trials are subject to influence from industry bias *by itself* to any claim about the relative merits of randomized trials. In fact one would suspect that it is easier for industry to introduce bias to observational studies and expert "consensus" statements. There is no need to go through the same regulatory processes to conduct an observational study, and these processes might provide some protection against nefarious influences of industry. Similarly, it is likely far cheaper to buy off a few experts at a consensus conference than to conduct a randomized trial.

Conclusion

Even critics of randomized trials admit that empirical studies show that well-conducted randomization and blinding can rule out bias, particularly selection bias (Worrall 2002). Given the benefits of (well-conducted) randomized trials over

(well-conducted) observational studies, where there is any conflict between the results of randomized and observational studies, it seems reasonable to side with the randomized study and assume that its results are more reliable (Howick 2011). However, if the results of the two different types of study are consistent or homogenous, as is often the case, there is no reason *not* to accept evidence from observational studies (Mebius 2014). It is probably unsurprising that the results are often similar, because many truly effective treatments will reveal significant effects in both types of study. However, we cannot predict in advance whether results from randomized trials and observational studies will differ. Our challenge to philosophical critics of the view that well-conducted randomized trials provide better evidence than well-conducted observational studies is simple. Let them provide a single example where results from a well-conducted randomized trial differ from results in a well-conducted observational study and where they believe the observational study results lie closer to the truth. Until this challenge is met, new treatments should be introduced in the context of well-conducted randomized trials, and existing treatments should be evaluated by measuring their effects within well-conducted randomized trials.

Definitions of Key Terms

Observational studies	In controlled observational studies, investigators compare people who are subject to an intervention with those who are not. The investigators neither allocate patients to receive the intervention nor administer the intervention. Instead, they observe what happens to people who choose (or are chosen by their healthcare practitioners) to take an intervention (or not).
Randomized trials	Participants in a randomized trial are randomly allocated to receive either an experimental intervention or a control.
Systematic review	Systematic reviews aim to gather all evidence that fits pre-specified eligibility criteria to address a specific research question. They aim to minimize bias by using explicit and systematic methods (Higgins and Green 2008).
Internal validity	Internal validity is a property that reflects the extent to which the causal conclusion of a study is justified for the study population.
External validity	External validity is the property of a study that renders its conclusions generalizable to populations outside the study.

Summary Points

- Observational studies tend to suffer from problems that are believed to increase the risk of producing inflated results in randomized trials; randomized trials are believed to overcome these problems.
- Randomized trials may have a lower degree of external validity than observational studies.
- The alleged superiority of randomized trials has not led to statistically significant average differences between results of randomized trials and results of observational studies.
- Because of the potential for randomized trials to rule out a greater degree of bias than observational studies, it follows that, ceteris paribus, we should side with randomized trials in cases where results from observational studies and randomized trials differ.
- New treatments should be introduced in the context of randomized trials.

References

Adler UC, Krüger S, Teut M et al (2013) Homeopathy for depression: a randomized, partially double-blind, placebo-controlled, four-armed study (DEP-HOM). PLoS One 8:e74537

Altman DG (2002) Poor-quality medical research: what can journals do? JAMA J Am Med Assoc 287:2765–2767

Anglemyer A, Horvath HT, Bero L (2014) Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev 4: MR000034

Barbui C, Cipriani A (2007) Publication bias in systematic reviews. Arch Gen Psychiatry 64:868

Berkson J, Magath T, Hurn M (1939) The error of estimate of the blood cell count as made with the hemocytometer. Am J Physiol 128:309–323

Borgerson K (2009) Valuing evidence: bias and the evidence hierarchy of evidence-based medicine. Perspect Biol Med 52:218–233

Boswell K, Cook C, Burch S, Eaddy M, Cantrell R (2012) Associating medication adherence with improved outcomes: a systematic literature review. Am J Pharm Benefits 4:e97–e108

Broadbent A (2013) Jeremy Howick: the philosophy of evidence-based medicine. Philos Sci 80:165–168

Bylund DB, Reed AL (2007) Childhood and adolescent depression: why do children and adults respond differently to antidepressant drugs? Neurochem Int 51:246–253

Carlson MD, Morrison RS (2009) Study design, precision, and validity in observational studies. J Palliat Med 12:77–82

Cartwright N (2007) Are RCTs the gold standard? Biosocieties 2:11–20

Cocco G (2009) Erectile dysfunction after therapy with metoprolol: the Hawthorne effect. Cardiology 112:174–177

Coronary Drug Project (1980) Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Engl J Med 303:1038–1041

DeCensi A, Puntoni M, Goodwin P et al (2010) Metformin and cancer risk in diabetic patients: a systematic review and meta-analysis. Cancer Prev Res 3:1451–1461

Deupree JD, Reed AL, Bylund DB (2007) Differential effects of the tricyclic antidepressant desipramine on the density of adrenergic receptors in juvenile and adult rats. J Pharmacol Exp Ther 321:770–776

Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J (2001) Influence of context effects on health outcomes: a systematic review. Lancet 357:757–762
Eisenach JC, Lindner MD (2004) Did experimenter bias conceal the efficacy of spinal opioids in previous studies with the spinal nerve ligation model of neuropathic pain? Anesthesiology 100:765–767
Every-Palmer S, Howick J (2014) How evidence-based medicine is failing due to biased trials and selective publication. J Eval Clin Pract 20:908–914
Glasziou P, Chalmers I, Rawlings M, McCulloch P (2007) When are randomised trials unnecessary? Picking signal from noise. Br Med J 334:349
Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI) (1986) Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet 22:397–402
Guyatt GH, Oxman AD, Vist GE et al (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 336:924–926
Higgins JJ, Green S (2008) The Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. Wiley Blackwell, Chichester
Howick J (2011) The philosophy of evidence-based medicine. Wiley Blackwell/BMJ Books, Chichester
Howick J, Mebius A (2014). In search of justification for the unpredictability paradox. Trials 15:480
Jefferson T, Demichell V, Di Pietrantonj C, Jones M, Rivetti D (2006). Neuraminidase inhibitors for preventing and treating influenza in healthy adults. Cochrane Database Syst Rev 19(3)
Jadad A (1998) Randomized controlled trials. BMJ Books, London
Jefferson T, Jones MA, Doshi P et al (2014) Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. Cochrane Database Syst Rev 4:CD008965
Jones B, Howick J, Hopewell J, Liew SM (2014) Response to 'Position statement on ethics, equipoise and research on charged particle therapy'. J Med Ethics 40:576–577
Keen HI, Pile K, Hill CL (2005) The prevalence of underpowered randomized clinical trials in rheumatology. J Rheumatol 32:2083–2088
Knekt P, Reunanen A, Jarvinen R et al (1994) Antioxidant vitamin intake and coronary mortality in a longitudinal population study. Am J Epidemiol 139:1180–1189
La Caze A, Djulbegovic B, Senn S (2012) What does randomization achieve? Evid Based Med 17:1–3
Mant D (1999) Can randomised trials inform clinical decisions about individual patients? Lancet 35:743–746
McCarney R, Warner J, Iliffe S et al (2007) The Hawthorne effect: a randomised, controlled trial. BMC Med Res Methodol 7:30
McCloskey RV, Straube RC, Sanders C, Smith SM, Smith CR (1994) Treatment of septic shock with human monoclonal antibody HA-1A. A randomized, double-blind, placebo-controlled trial. Ann Intern Med 121:1–5
McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D (2002) Randomised trials in surgery: problems and possible solutions. Br Med J 324:1448–1451
Mebius A (2014) Corroborating evidence-based medicine. J Eval Clin Pract 20:915–920
Muthuri SG, Venkatesan S, Myles PR et al (2014) Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. Lancet Respir Med 2:395–404
Oberai P, Balachandran I, Janardhanan N et al (2013) Homoeopathic management in depressive episodes: a prospective, unicentric, non-comparative, open-label observational study. Indian J Res Homoeopath 7:116–125
OCEBM Levels of Evidence Working Group (2011) The Oxford 2011 levels of evidence. Oxford Centre for Evidence-Based Medicine. Available at: http://www.cebm.net/index.aspx?o=5653. Accessed 17 June 2011
Odgaard-Jensen J, Vist GE, Timmer A et al (2011) Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 4:MR000012

Penston J (2003) Fact and fiction in medical research: the large-scale randomised trial. The London Press, London

Petitti DB, Perlman JA, Sidney S (1987) Noncontraceptive estrogens and mortality: long-term follow-up of women in the Walnut Creek Study. Obstet Gynecol 70:289–293

Rosenthal R, Lawson R (1964) A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. J Psychiatr Res 69:61–72

Rossouw JE, Anderson GL, Prentice RL et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA J Am Med Assoc 288:321–333

Schulz KF, Grimes DA (2002) Blinding in randomised trials: hiding who got what. Lancet 359:696–700

Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA J Am Med Assoc 273:408–412

Sesso HD, Buring JE, Christen WG et al (2008) Vitamins E and C in the prevention of cardiovascular disease in men: the physicians' health study II randomized controlled trial. JAMA J Am Med Assoc 300:2123–2133

Sierevelt IN, van Oldenrijk J, Poolman RW (2007) Is statistical significance clinically important? A guide to judge the clinical relevance of study findings. J Long Term Eff Med Implants 17:173–179

Smith GD, Ebrahim S (2002) Data dredging, bias, or confounding. Br Med J 325:1437–1438

Stampfer MJ, Colditz GA (1991) Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. Prev Med 20:47–63

Stevens RJ, Ali R, Bankhead CR et al (2012) Cancer outcomes and all-cause mortality in adults allocated to metformin: systematic review and collaborative meta-analysis of randomised clinical trials. Diabetologia 55:2593–2603

Straus SE, Glasziou P, Richardson WS, Haynes RB (2011) Evidence-based medicine: how to practice and teach EBM, 4th edn. Churchill Livingston, Edinburgh

Stroup DF, Berlin JA, Morton SC et al (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA J Am Med 283:2008–2012

Vist GE, Bryant D, Somerville L, Birminghem T, Oxman AD (2008) Outcomes of patients who participate in randomized controlled trials compared to similar patients receiving similar interventions who do not participate. Cochrane Database Syst Rev 3:MR000009

Wartolowska K, Judge A, Collins G et al (2014) Use of placebo controls in the evaluation of surgery: systematic review. Br Med J 348:g3253

Wood L, Egger M, Gluud LL et al (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. Br Med J 336:601–605

Worrall J (2002) What evidence in evidence-based medicine? Philos Sci 69:S316–S330

Worrall J (2007) Evidence in medicine. Philos Compass 2:981–1022

Zetin M, Hoepner CT (2007) Relevance of exclusion criteria in antidepressant clinical trials: a replication study. J Clin Psychopharmacol 27:295–301

Ziegler EJ, Fisher CJ Jr, Sprung CL et al (1991) Treatment of gram-negative bacteremia and septic shock with HA-1A human monoclonal antibody against endotoxin. A randomized, double-blind, placebo-controlled trial. N Engl J Med 324:429–436

Zimmerman M, Posternak MA, Chelminski I (2002) Symptom severity and exclusion from antidepressant efficacy trials. J Clin Psychopharmacol 22:610–614

Statistical Generalizations in Epidemiology: Philosophical Analysis

54

Federica Russo

Contents

Abstract

Epidemiology studies the variations in health in populations, according to a number of parameters. In this field, probability and statistics are used in order to provide a quantitative description and analysis of the variations in exposure and disease, as well as of the effects of possible preventatives. Thus, one goal of epidemiology is to establish statistical generalizations about health and disease in populations. Consequently, it is important to understand how statistical generalizations are established and what use one can make of them to establish medical knowledge or to design public health policies.

F. Russo (✉)
Department of Philosophy, Faculty of Humanities, University of Amsterdam, Amsterdam, The Netherlands
e-mail: f.russo@uva.nl

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_39

Introduction: Epidemiology and Statistics

Epidemiology studies the *variation* in health in populations. This means several things. For instance, epidemiology studies how a disease is spread in a population (or across populations), according to different levels of exposures. Epidemiology also studies the association between different factors (biological or socio-economic) and different diseases, or whether certain treatments, interventions, or preventative factors are associated with decrease in the burden of disease in a population. These different aspects of epidemiological investigation into health and disease are captured by the definition of Miquel Porta (quoted in Saracci 2010, 10):

> [*Epidemiology is*] *the study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control health problems.*

Two aspects are immediately worth noting. The first is that epidemiology is interested in *populations*, i.e., groups of individuals, not primarily in *individual patients*. The second is that any epidemiological result is relative to some population of reference or relative to a comparison between specific populations.

Epidemiology also has a "composite" nature, as it blends theoretical instruments of medical research and of probability and statistics (Saracci 2010). Probability and statistics, in particular, are tools borrowed from demography and social science, thus also revealing the peculiar place of epidemiology in the realm of the sciences: it is right at the frontier between the social and the biomedical sciences. This is not just due to the kind of methods used in epidemiology (mainly, probabilistic and statistical approaches) but also for the *object* of study. In fact, epidemiology studies variations in health, according to factors that are biological, socio-economic, demographic, etc. This is also reflected in the various subfields within epidemiology: some areas, like social epidemiology, prioritize the study of health variations according socio-economic factors, while others, like molecular epidemiology, prioritize the study of health variations according to molecular (e.g., genetic) factors.

It is implicit in the definition of epidemiology given above that epidemiology prioritizes a *quantitative* description and analysis of health, as it studies exposure and disease with the tools of probability theory and statistics. This raises immediately two questions that are also shared by other germane scientific disciplines such as demography. First, to what extent, or under what conditions, do statistical analyses allow us to establish causal relations? Second, is it part of the objectives of epidemiology to formulate recommendations for policy? The first question is a perennial question in the philosophy of causality and in scientific method. While in epidemiology this issue has some peculiarities (to be discussed later), it also shares features and questions arising elsewhere. The second question ultimately has to do with the purported descriptive or normative character of a discipline. It is a debated issue whether it is part of any epidemiological study to aim to formulate policy recommendations or whether these should be formulated outside epidemiological

studies. The formulation of policy recommendations is presented, in the definition of Porta reported above, as an integral part of epidemiology, although the relation between epidemiology and public health remains controversial (see, e.g., Jackson et al. 1999; Samet 2000).

In sum, it is a direct consequence of the meaning and definition of epidemiology to base empirical studies on the collection of data and on a statistical analysis of them. Epidemiology aims to establish *generic* claims about (variations in) health, and for that purpose probabilistic and statistical approaches are the preferred tools.

Statistical Generalizations

Statistical generalizations in epidemiology can be categorized according to their aim. *Descriptive* generalizations aim to provide a description of the variations in health and disease in a population. *Analytical* (or causal) generalizations also aim to identify those factors responsible for the observed variations. For instance, the European Centre for Disease Prevention and Control provides a report on development of measles and rubella in European countries for the period April 2014–March 2015 (ECDPC 2015). The reports include numbers about new reported cases, the countries most affected by the diseases, the percentage of the cases positively diagnosed by lab analyses, or about the percentage of vaccinated people. These types of report intend to provide a *description* of the health situation in a population. Although one might hypothesize that the revival of foci of infection might be due to a drop in vaccination, the *causal* character of such a generalization has to be established in further studies.

In order to understand the difference between descriptive and analytical generalizations, we need to introduce some technical terms used in epidemiology. *Prevalence* refers to the proportion of diseased individuals in a population, counted in a specific time and in a specific population. *Incidence* refers instead to the *new* cases reported in a given time lapse, and divided by the number of people who are at risk (but that do not have the disease). Prevalence and incidence are *descriptive* statistical concepts. Analytical generalizations also make use of the concept of *risk*, which is related (in way that is not always clear) to causality.

To begin with, epidemiology is interested in calculating, estimating, or analyzing different risks and odds. Risks and odds are associational measures that quantify the strength of association between two variables: a particular outcome (disease) and the presence of a factor (exposure).

For the purpose of this chapter, let us consider two variables E and D, denoting "exposure" and "disease," respectively, being binary or dichotomous, i.e., each has only two possible levels: exposed/unexposed and diseased/not diseased.

	D	
E	Diseased	Not diseased
Exposed	a	b
Unexposed	c	d

Here, $a + c$ is the marginal probability of disease, i.e., P(D), and $a + b$ is the marginal probability of exposure, i.e., P(E). Consequently, $b + d = \text{P}(\neg\text{D})$ and $c + d = \text{P}(\neg\text{E})$.

We can also organize observations in a contingency 2×2 table, having thus four cells.

	D	
E	Diseased	Not diseased
Exposed	n_{11}	n_{12}
	p_{11}	p_{12}
Unexposed	n_{21}	n_{22}
	p_{21}	p_{22}

The notation n_{ij} refers to the *number* of subjects observed in the corresponding cell, i.e., to the number of observations in the i-th row ($i = 1, 2$) and j-th column ($j = 1, 2$); the notation p_{ij} refers instead to the *proportion* of subjects observed in the corresponding cell, where $p_{ij} = n_{ij}/n$. With this data, we can compute relative risks, odds, and odds ratios and estimate probabilities.

The *relative risk* (RR) is defined as the ratio of risk in the exposed and unexposed group:

$$\frac{n_{11}/n}{n_{21}/n} = \frac{p_{11}}{p_{21}}$$

Thus, RR compares groups: the exposed and the unexposed. RR > 1.0 indicates that the risk of disease is increased when the risk factor (exposure) is present; RR < 1.0 indicates that the risk of disease is decreased when the risk factor is present, i.e., the factor is a protective factor or preventative.

The corresponding definition in terms of conditional probabilities is

$$\frac{P(D \mid E)}{P(D \mid \neg E)} = \frac{a/(a+b)}{c/(c+d)}$$

The *odds ratio* (OR) is another way to compare proportions in a 2×2 contingency table. OR is computed from odds, i.e., it is the ratio of the odds of disease in the exposed group and the odds of disease in the unexposed group:

$$\text{OR} = \frac{\text{Odds}_{ex}}{\text{Odds}_{unex}}$$

The odds of an outcome are equal to the probability that the outcome does occur, divided by the probability that the outcome does not occur. In a 2×2 contingency table, the probability of an outcome is equal to the number of times the outcome is observed divided by the total observations. Thus, we can write, for the odds of the exposure:

$$\text{Odds}_{ex} = \frac{n_{11}/(n_{11} + n_{12})}{n_{12}/(n_{11} + n_{12})} = \frac{n_{11}}{n_{12}}$$

where $n_{11}/(n_{11} + n_{12})$ is the probability that the disease occurs in the exposed group and $n_{12}/(n_{11} + n_{12})$ is the probability that the disease does not occur in the exposed group. We can express this in terms of conditional probabilities:

$$\text{Odds}_{ex} = \frac{P(D \mid E)}{P(\neg D \mid E)}$$

Similarly, for the odds of the unexposed:

$$\text{Odds}_{unex} = \frac{n_{21}/(n_{21} + n_{22})}{n_{22}/(n_{21} + n_{22})} = \frac{n_{21}}{n_{22}}$$

where $n_{21}/(n_{21} + n_{22})$ is the probability that the disease occurs in the unexposed group and $n_{22}/(n_{21} + n_{22})$ is the probability that disease does not occur in the unexposed group. We can express this again in terms of conditional probabilities:

$$\text{Odds}_{unex} = \frac{P(D|\neg E)}{P(\neg D|\neg E)}$$

OR can now be computed as

$$\frac{n_{11}/n_{21}}{n_{12}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

This is equivalent to

$$\frac{P(D|E)}{P(\neg D|E)} \times \frac{P(\neg D|\neg E)}{P(D|\neg E)}$$

It is also worth noting that there is a mathematical relation between odds and probabilities:

$$P = \frac{\text{Odds}}{1 + \text{Odds}}; \text{Odds} = \frac{P}{1 - P}$$

The interpretation of risks and odds raises at least two questions. One question concerns the fact that these measures make sense at the *generic* level, i.e., for groups of individuals, but *not* for individual patients. In other words, even if we can express risks and odds in terms of probabilities, this isn't convenient, as probabilities can be also taken to directly apply to the single case; however, in this context, this is a misleading interpretation. In fact, there isn't as yet a straightforward way to

determine the *individual* risk of developing a disease, knowing the risk for the population. For instance, inherited mutations of genes *BRCA1* and *BRCA2* are associated with a high risk of developing breast cancer, but this does not imply that I *will* develop cancer (even if these mutations are found in my body). One reason for this is due to what philosophers called "the reference class problem," which refers to the difficulty of assigning an individual patient to the (most) correct reference class and thus making secured inferences about their health (see also Statistical Generalizations and Philosophy of Science). One hope of personalized medicine is precisely to measure biological characteristics of individuals so that more precise diagnostic and prognostic inferences as well as "patient-tailored" risk prediction and treatment can be made (for a discussion, see, e.g., Hayes et al. 2014).

Another question concerns the possible *causal* import of these measures. Admittedly, risks and odds are *associational* measures and cannot be given a direct, straightforward causal interpretation. However, it would be a mistake to totally dismiss it either. Risks and odds should be seen as part of our evidence base in order to formulate causal claims about health and disease. It should be noted that whether *risk* is a causal notion has not been settled yet. On the one hand, it is difficult to see how the notion of risk is completely devoid of *any* causal connotation, as information of risks is routinely used to design preventive interventions in public health. On the other hand, risk is clearly not coextensive with the term "cause," as it is widely agreed that claims about risks do not imply causation.

The causal interpretation is, in a sense, the core issue about statistical generalizations in epidemiology. This emerges also when considering generalizations not necessarily expressed in terms of risks and odds, but formulated as results of statistical modeling. In fact, not all epidemiological studies involve only two binary variables (diseased/not diseased; exposed/unexposed). Many of them analyze large datasets with numerous variables, not just dichotomous. Here, general methodological caveats apply about the choice of variables, the use of background knowledge, the quality of data, etc. All these apply to statistical modeling in *any* discipline (useful discussions can be found in Freedman (2005) and Russo (2009), among others).

In the following, the controversies concerning two issues, notably, (i) statistical tests and (ii) confounding and control, will be highlighted.

Statistical tests. Establishing statistical generalizations involves performing tests, and a typical argument is that results of these tests have to be *significant*. This means, to begin with, that tests concern *hypotheses*. These hypotheses come from a given research question. For instance, do statins reduce cholesterol level? Or, do calcium supplements help prevent osteophorosis in women aged 50 +?. Hypothesis testing is meant to compare the hypothesis with observations sampled from the population. In a statistical test, we can identify the following elements:

- *Null hypothesis*: there is no association between the two variables. E.g.: no association between statins and lower cholesterol levels.

- *Alternative hypothesis*: there is an association between the two variables. E.g.: there is an association between calcium supplements and osteoporoses in women aged 50+.

Hypotheses in epidemiology may concern, for instance, the differences or similarities in frequency of disease across populations, places, or time. They may also concern the variation in frequency of disease in relation to some specific factor.

Typically, attention is given to the conditions to *reject* the null hypothesis; these concern the test statistic and the significance level. The significance level is chosen on the basis of the amount of type I or type II error one is prepared to accept and on the basis of the problem at hand. A type I error means that the null hypothesis is rejected when in fact it is true, and a type II means that the null hypothesis is accepted when instead the alternative is true. Common test statistics are the z-test, the F-test, and the X^2-test. The specificities of these tests will not be discussed in this contribution. The null hypothesis is accepted or rejected at a given significance level (the *p-value*), which is usually set at 5 % (for a very lucid and accessible presentation of hypothesis testing, see, e.g., Freedman et al. 1998, Chaps. 26–29).

The logic behind hypothesis testing may appear intuitively very simple. However, tests of significance hide difficulties that concern their *interpretation*. David Freedman and his coauthors offer an inventory of these difficulties (Freedman et al. 1998, Chap. 29; Freedman 2005, 60ff). Let us examine some of these. Firstly, the word "significance" might be misleading. In fact, in the statistical jargon, "significant" is not synonymous with "important" or "relevant" but with "probably true," i.e., not due to chance. Secondly, the p-value of a test depends on the sample size and presupposes the quite strong assumption that the sample is *representative* of the population. Moreover, the threshold for significance is rather relative. Textbooks usually recommend rejecting the null hypothesis at 5 % or at 1 % level. Yet these levels are arbitrary. Freedman and coauthors, for instance, make the point that there isn't a real difference between two p-values, say one set at 5.1 % and the other at 4.9 % (Freedman et al. 1998, Chap. 29). Therefore, they recommend reporting the test used and the exact p-value; otherwise, "statistically significant" is too vague a statement.

The meaning and use of p-values have been often discussed in the literature because they are susceptible of multiple interpretations. Lagiou et al. (2005), for instance, discuss the interpretation of p-values specifically in the context of epidemiological research and point to two major difficulties: (i) the p-value is interpretable only when *one* comparison or *one* test is performed and (ii) the p-value itself does *not* convey information about the *strength* of the association. This second point is worth explaining in detail, as it is closely connected with the causal interpretation of statistical generalizations. Statistical hypotheses concern, in one way or another, *correlations* between variables. But the p-value does not give the chance of the null hypothesis being *true*. A small p-value has to be interpreted as evidence against the null hypothesis, in particular as suggesting that something beside chance is operating to make the difference. As Freedman and coauthors explain very clearly, a test of significance does *not* shed light on the causes of

variations. Instead, significance tests merely test whether an observed variation is real (alternative hypothesis) or just chancy, that is, somehow an artifact of the dataset (see Freedman et al. 1998, Chap. 29).

The interpretation of probability is also worth mentioning. If we adopt a frequentist approach, what we test is not the probability of the hypothesis being true, but the probability of obtaining the observed sample *if* the hypothesis is true. This difference is subtle but fundamental. Under the frequentist interpretation, we cannot attach a probability value to a single case (for instance, a hypothesis). This is because probability expresses frequency of occurrence in finite or infinite sequences. Instead, if we adopt a Bayesian interpretation, we can attach a probability value to the single case and therefore have a meaningful way of expressing the probability of a particular hypothesis. For instance, if the hypothesis to be tested is about whether an unknown parameter θ lies in the interval (θ_1, θ_2) and confidence level for this test is 95 %, one may be tempted to interpret this as the probability of θ to lie in that interval. This interpretation, however, is not correct. Instead, this means that *if* we draw many samples of the same size and build the same interval around θ, *then* we can expect that 95 % of the confidence intervals will contain the unknown parameter. But this, notice, is not the same thing as asking what is the probability that a given parameter will lie in a given interval. For this reason Freedman and coauthors, discussing confidence intervals and the frequency interpretation, say that "chances are in the sampling procedure, not in the parameter" (Freedman et al. 1998, 347). Courgeau (2004) also provides a very lucid account of the meaning of hypothesis testing in a frequentist and in a Bayesian framework.

Confounding and control. Statisticians analyzing epidemiological (and other) data are well aware of the problem of confounding in establishing generalizations. Simply put, even if the statistical model attests to an association (or dependence or correlation) between two variables, this is no guarantee that the correlation corresponds to a causal relation. To begin with, correlations are symmetric. So, a priori, we cannot decide the direction in which causality is supposed to flow. But, in many occasions, we do have enough background knowledge – including temporal information about the occurrence of events – that allows us to *hypothesize* the direction of the causal relation. For instance, suppose we find a correlation, in a cohort study, about "lung cancer" and "yellow fingers," and suppose we know that individuals reported lung cancer events *after* yellow finger events. We might then be inclined to infer that having yellow fingers is a risk factor (or a cause?) of lung cancer. But once we include in the model a *third* variable, namely, "cigarette smoking," which is also temporally prior to lung cancer events, the correlation disappears. This is one (oversimplified and schematic) case where one variable (yellow fingers) confounds a causal relation (cigarette smoking → lung cancer). In this case, the solution is rather easy, as having yellow fingers is a mere "side effect" of cigarette smoking but has no proper causal role in this structure.

Some other cases are less easy to work out, even if "toy examples" rather than real epidemiological studies are analyzed. For instance, coffee drinking and heart disease are associated. This association, one might think, is explained away once

cigarette smoking is introduced, as cigarette smoking is supposedly the cause of heart disease. However, cigarette smoking is *also* associated with coffee drinking. This complicates the analysis. On the one hand, coffee drinking may still have its own affect on heart disease (maybe positive, maybe negative). So, to properly understand variations in the outcome (heart disease), we have to individually control cigarette smoking and coffee drinking. On the other hand, the correlation between cigarette smoking and coffee drinking may be in need of further examination, for instance, introducing another explanatory factor, say stress, that explains it away. It may turn out that stress too is associated with, or even causally responsible for, heart disease. More generally, confounding and control constitute a challenge for epidemiology because diseases have, in many cases, *multiple* causes, rather than just one. The shift from "monocausal" models to "multifactorial" models has been a major advancement for epidemiology, both conceptually and methodologically (for a discussion, see Broadbent 2013).

Examples like this may easily turn into a conundrum to solve, but the real message to convey is the following. In *real* science, most often than not, we do *not* know which variables are confounded, which variables should be controlled for, and which other variables should be measured and included in the model. It is precisely the task of statistical modeling to analyze the relations among variables and to build a cogent story about their causal or noncausal role. In practice, statistics has been quite successful in developing methods for controlling variables at the level of study design (also called ex ante stratification) or after data collection (ex post stratification). It is not the goal of this contribution to provide a thorough presentation of methods for control, but rather to point to some of the theoretical issues involved.

Statistical Generalizations and Philosophy of Science

Statistical generalizations also raise philosophical issues, notably about their *status*. In fact, philosophers of science have long been interested in laws of nature because laws tell us how the world is, allow prediction about what will happen under specified circumstances, and are part of our explanations of phenomena. It didn't take long for philosophers to realize that laws of nature apply to *some* portions of reality, but not others. The quantum world is one example, but social and health contexts are no less controversial. While the debate on what makes an empirical generalization a lawful statement is not settled, it is also widely agreed that thermodynamics has laws, but not epidemiology. Thus, no matter how precisely we state a statistical generalization about vaccination habits and measles outbreaks or about smoking habits and cancer development, these are *not* laws. In the following, it will be discussed what good are statistical generalizations (in epidemiology) even if they are not laws.

In philosophy of science, Woodward (2003) put forward the idea that we should investigate what confers explanatory power to statements that are not lawful. His

arguments mainly concern the kind of generalizations established in economics and social science, but it is easy to extend them to epidemiology. A note on terminology: it is necessary to adapt the statistical jargon used earlier to the philosophical argument presented next.

Woodward examines the status of empirical generalizations (i.e., statistical generalizations) and claims that these are *change-relating* relations. This characterization has important epistemological (and methodological) consequences, as also highlighted in Russo (2011). We can express empirical generalizations under the general form $Y = \beta X + \varepsilon$. It is worth noting that this "reduced" form is certainly general enough, even though it already encapsulates some hypotheses, for instance, that the relation between X and Y is linear. In many cases, this is certainly not the case. But for the present discussion, this is not central. A variational reading of this equation amounts to the following: variations in Y are due to variations in X. How much Y varies is quantified by the parameter β (and the errors ε indicate that the relation is stochastic rather than deterministic).

Two remarks are in order. First, a generalization is about *variations*. In epidemiology, we are interested in variations in the occurrence of disease, in exposure, in time, factors, etc. Second, implicit in that reading is also that there must be a variation *within* variables X and Y. If we study the relation between vaccination habits and measles outbreak, the dataset must contain observations about vaccination and *non*-vaccination and about occurrence and *non*-occurrence of measles. Some philosophers have expressed this idea emphasizing that causation is contrastive (see, e.g., Schaffer 2005; Northcott 2008). However, we are not to causation yet. All the equation $Y = \beta X + \varepsilon$ says is that there is a *joint variation* between X and Y.

Woodward (2003) famously explained that for change-relating relations to be causal, they also have to be *invariant*, notably invariant under interventions on the putative cause variable. Woodward is at pains to explain what that means using mainly examples from physics. The account, however, is meant to apply to socio-economic contexts too and, with some amendment, to epidemiology. Simply put, Woodward's account prescribes that *causal* generalizations are the ones that show invariant properties. This means that if we performed an intervention I on X and hold fixed any other possible factor influencing Y, Y should also vary. One peculiarity of this account concerns the meaning of I: interventions are manipulations on the putative cause variable X, such that they change *only* X, via X they change *only* Y, and they are uncorrelated with anything else in the model. Another peculiarity of this account is that it oscillates between providing a conceptual analysis of causation in terms of invariance under interventions and providing a methodology for testing whether empirical generalizations are (or are not) causal (for a discussion, see, e.g., Strevens 2007; Strevens 2008; Russo 2012). Some commentators pointed out that either way (i.e., whether the project belongs to the metaphysics or to the methodology of causation), it is ill suited to observational contexts, because the account hinges too heavily on *manipulations*. This is clearly the case of epidemiology, where the large majority of the empirical studies are observational, rather than experimental.

The account can regain generality and become applicable to non-experimental contexts using a "non-interventionist" notion of invariance. The hint comes from Woodward's analysis (Woodward 2003, p. 312), as he describes how invariance was tested in a 1959 study on smoking and lung cancer. At that time, the biochemical mechanisms of carcinogenesis were mainly unknown and empirical generalizations were established on the basis of large epidemiological studies, rather than on the results of lab experiments. Simply put, Woodward points out that, in the absence of interventions, we must check whether a correlation is stable (or invariant) across different subpopulations, for instance, men and women, different age groups, socio-economic status, different levels or types of smoking, etc. While Woodward calls this type of invariance *weak*, Russo (2014) argues that we should not create an opposition between strong and weak invariance, but rather understand how invariance tests are *implemented* in different modeling practices. After all, different implementations of invariance tests do share some common features, notably that they test the robustness and regularity of joint variations of variables and that they aim at establish *generic* claims. Both are important epistemological points.

Concerning the first, its relevance has to do with what has also been called in the literature the "contrastive" character of causation, and that has been mentioned earlier: we need things to change and vary in order to establish which changes and variations are causal. The second is relevant because it has to do with the *scope* of generalizations. We need generalizations to be *generic*, namely, valid for the population as a whole, because this is the way they can contribute to building medical knowledge and to design public health interventions. It is worth noting that this does not exclude that case reports, which are essentially about single cases rather than populations, be important. Their role and use for medical knowledge and policy are different and fall beyond the scope of this contribution.

The problem of the Population of Reference

It is also worth drawing the attention to an issue that is simultaneously of theoretical and methodological relevance: the choice of the population of reference. Statistical studies in epidemiology and social science are all highly sensitive to the choice of the population of reference. This is key to extract a representative sample, to collect data, and to interpret the results of statistical analyses. Several issues are at stake.

One is the *scope* of statistical generalizations. If we establish generalizations about dengue disease using data collected from some regions in Brazil, are the results also valid for Indian regions where the disease is present? One might raise the point and argue that, clearly, regions in Brazil and India must be different in some respect, thus undermining the possibility of exporting the generalizations from Brazilian to Indian contexts. Suppose now we are interested in studying psycho-social risks related to stress and burnout at workplace in Belgium. *What* Belgium are we referring to? Only within Brussels capital region, we should pay attention to the composition of the sample: Walloon, Flemish, non-European

immigrants, "eurocrats," etc., these are already *four* groups having distinct characteristics and yet composing *one* population; these four sub-populations might require quite different analyses or intragroup comparisons. Thus there is no single way in which we can define a population of reference. Sometimes it is from geographical parameters, some other times it is from ethnic or socio-economic factors or depending on exposure and occurrence of disease, or others.

The straightforward methodological consequence is that the choice of the population of reference has to be carefully pondered, using available background knowledge and, sometimes, preliminary analyses of data. In empirical studies, this choice is typically made *before* data are collected, but this does not exclude that the population of reference is refined in the course of empirical investigation, for instance, invariance or other tests may reveal further relevant sub-populations to be considered in the study.

The question of the population of reference is also directly related to the question of validity. In the methodological literature, validity received systematic discussions since the work of Cook and Campbell (1979), whose discussion refers to quasi-experimental models in social science. However, their considerations about validity are relevant to most statistical modeling practices, including in epidemiology, where discussions abound. In this context, validity refers to the confidence with which we draw conclusions from the study of correlations. *Statistical conclusion* validity refers to whether we gathered enough evidence and performed enough tests to infer that a given correlation is causal (or not causal). *Internal* validity refers to the confidence with which we can establish that the results apply to the chosen population of reference. *Construct* validity is about choosing the right "construct" for variables that cannot be measured directly, typically, "socio-economic status," or "education" but also self-rated health status or the like. Finally, *external* validity refers to the possibility of extending the results to *other* populations.

There is a vivid and vast debate raised by this taxonomy of validity; however, it will not be examined in detail here. It will suffice to mention that, in epidemiology, the debate often polarizes around an alleged dilemma: studies either have high internal validity or high external validity, but not both. This is important because, as the argument goes, we are not simply interested in establishing results at a "local" level but also to use them widely in public health interventions. Thus, if a vaccination program against dengue fever is successful for one population, we might want to try it out elsewhere. Conversely, if we can establish robust results about the obesity epidemic in children worldwide, it does not follow that we managed to identify factors that are specific to some population, rather than another. Broadly speaking, these are terms in which the debate is set.

It is worth noting, however, that validity is *not* an intrinsic property of studies, but of the *process* of carrying them out. The validity of results should be assessed with respect to the rigor used during the *whole* process of data collection, data analysis, interpretation of results, etc.

Questions about the choice of the population of reference and validity emerge when considering the hypothesis of "universal biological response,", which is

usually made in randomized controlled trials (Victora et al. 2004). This means that we assume that individuals respond to treatments in a way that is similar enough. In turn, this presupposes that our bodies function in very, very similar ways. Victora et al. (2004) question this hypothesis saying that, while it may have plausibility if short causal paths are considered, individual responses are instead highly heterogeneous when more complex causal paths are involved.

In addition to this line of argument, it is worth mentioning that individual responses may be different, even in short, simple causal pathways when *relevant* factors are considered. An interesting example in this respect is gender medicine, as it is trying, since some decades now, to spell out the mechanisms of health and disease involved in different genders. These may differ because of biological or socio-psycho-behavioral differences or because of the way male and female illnesses are understood. For instance, the phenomenon of wrong diagnoses of heart attacks in women has been widely documented; similarly, male breast cancer is poorly understood, and its mechanisms are largely extrapolated from studying females. Thus, gender ought not to be used *just* as a classificatory variable to use in a posteriori partitions of the population. Instead, the goal is to understand what is involved in different modes of being exposed, or of disease mechanism, or reacting to interventions.

An analogous argument holds for the use or "race" in epidemiological (and other social science) studies. Studying and understanding variations in exposure, disease, and interventions according to race may be important to capture *social and behavioral* factors. For instance, ethnic differences in hypertension and blood pressure have long been reported and documented. Surely, including data about race may also help explain differences in the biology of health and disease (see the hypothesis of universal biological response mentioned earlier), but clearly we should beware of not reviving value judgments from ethnic differences in health and disease.

Usefulness of Statistical Generalizations

Statistical generalizations, as discussed in previous sections, are sensitive to a number of methodological and philosophical caveats. At the methodological level, statistical generalizations are vulnerable to the problem of confounding, and, more generally, they do not automatically license causal inference. At the philosophical level, it is controversial to assign a clear status to statistical generalizations: they clearly aren't laws and yet they are essential to gain knowledge about health and disease. Thus, it is important to highlight the *usefulness* of statistical generalizations, in spite of all the caveats already discussed.

To begin with, statistical generalizations are useful to establish generic knowledge about health and disease. Epidemiology aims at providing a faithful description and explanation of the variation of health and disease *in populations*, and, for this reason, statistical *generalizations* are vital. They are vital because it is on the basis of *generalizations* that we can design public health interventions.

Because of their generic character, statistical generalizations are also useful in that they provide *evidence of correlation*. Evidence of correlation complements evidence of production, or of mechanisms, in establishing causal claims in the medical sciences.

Statistical generalizations are also useful to make inferences about single cases. Claims about single cases are not deductively derived from the corresponding generic claims in any simple or direct way. Indeed, the relation between the generic and single-case level is complex. On the one hand, generalizations are not mere aggregates of single cases, and, conversely, single cases are not mere instantiations of generic relations. Russo and Williamson (2011) describe this complexity for the case of autopsies, showing how each level participates in establishing claims at the other level. Kleinberg (2013) develops a variant of the "connecting principle," originally proposed by Sober (1986), and explains, from a statistical point of view, how generalizations can be fruitfully used in making inferences to the single case.

Finally, with the rapid development and use of techniques for the analysis of *big* datasets, it is important to reflect upon the value of correlational claims. In epidemiology, data-intensive science offers an opportunity to explore the determinants of health and disease with unprecedented variety, volume, and velocity (the so-called three Vs). Scientists and philosophers alike are nonetheless cautious in declaring the advent of a revolution that will put an end to the conundrum of how to infer causation from correlation (see, e.g., Mooney et al. 2015; Alyass et al. 2015).

Definitions of Key Terms

Epidemiology	The study of variations of health and disease in populations according to biological and socio-economic factors.
Statistical generalization	Scientific statements expressing in a quantitative way facts about health and disease in a population, for instance, about risks or about the effectiveness of a drug or about an intervention.
Medical knowledge	The body of knowledge about health and disease that scientists gather together through epidemiological, laboratory, and other forms of studies.
Population of reference	The specific population being the object of an epidemiological study.
Confounding	Phenomenon occurring when a variable interferes while studying the correlation between two other variables.
Control	Any statistical technique to avoid or minimize confounding, for instance, conditioning on relevant variables or stratification.

Summary Points

- Epidemiology studies variations in health and in exposure in populations.
- Epidemiology uses probability and statistics to establish generalizations about exposure and disease about populations.
- Statistical generalizations are generic claims. While they lack the typical features of laws (of nature), they should be sufficiently invariant (or robust) to be used to establish medical knowledge or to design public health policies.
- Statistical generalizations in epidemiology are always established with respect to some population and reference class.
- An important methodological aspect in establishing statistical generalizations concerns controlling for possible confounding factors.

References

Alyass A, Turcotte M, Meyre D (2015) From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med Genomics 8(1):33

Broadbent A (2013) Philosophy of epidemiology. Palgrave McMillan, New York

Cook TD, Campbell DT (1979) Quasi-experimentation design and analysis issues for field settings. Rand MacNally, Chicago

Courgeau D (2004) Probabilité, démographie et sciences sociales. Math Soc Sci 167:27–50

ECDPC (2015) Measles and rubella monitoring, April 2015 – reporting on April 2014 to March 2015 surveillance data and epidemic intelligence data to the end of April 2015. European Centre for Disease Prevention and Control. http://ecdc.europa.eu/en/publications/Publications/Measles-rubella-monitoring-second-quarter-2015.pdf

Freedman DA (2005) Statistical models: theory and practice. Cambridge University Press, Cambridge

Freedman D, Pisani R, Purves R (1998) Statistics, 1st edn. Norton, New York

Hayes DF, Markus HS, David Leslie R, Topol EJ (2014) Personalized medicine: risk prediction, targeted therapies and mobile health technology. BMC Med 12:37. doi:10.1186/1741-7015-12-37

Jackson LW, Lee NL, Samet JM (1999) Frequency of policy recommendations in epidemiologic publications. Am J Public Health 89(8):1206–11

Kleinberg S (2013) Causality, probability, and time. Cambridge University Press, New York

Lagiou P, Adam H-O, Trichopoulos D (2005) Causality in cancer epidemiology. Eur J Epidemiol 20:565–74

Mooney SJ, Westreich DJ, El-Sayed AM (2015) Commentary: epidemiology in the era of big data. Epidemiology 26(3):390–94. doi:10.1097/EDE.0000000000000274

Northcott R (2008) Causation and contrast classes. Phil Stud 139:111–23

Russo F (2009) Causality and causal modelling in the social sciences measuring variations. Methodos series. Springer, New York

Russo F (2011) Correlational data, causal hypotheses, and validity. J Gen Phil Sci 42(1):85–107

Russo F (2012) On empirical generalisations. In: Dieks D, Gonzalez WJ, Hartmann S, Stoeltzner M, Weber M (eds) Probabilities, laws, and structures. Springer, Berlin, pp 133–50

Russo F (2014) What invariance is and how to test for it. Int Stud Phil Sci 28(2):157–83. doi:10.1080/02698595.2014.932528

Russo F, Williamson J (2011) Generic vs. single-case causal knowledge. the case of autopsy. Eur J Phil Sci 1(1):47–69

Samet JM (2000) Epidemiology and policy: the pump handle meets the new millennium. Epidemiol Rev 22:145–154, http://epirev.oxfordjournals.org/content/22/1/145.full.pdf

Saracci R (2010) Epidemiology. A very short introduction. Oxford University Press, New York

Schaffer J (2005) Contrastive causation. Philos Rev 114(3):297–328

Sober E (1986) Causal factors, causal inference, causal explanation. Proc Aristotel Soc Suppl Vol 60:97–113

Strevens M (2007) Essay review of Woodward, making things happen. Philos Phenom Res 74:233–49

Strevens M (2008) Comments on Woodward, making things happen. Philos Phenom Res 77 (1):171–92

Victora CG, Habicht J-P, Bryce J (2004) Evidence-based public health: moving beyond randomized trials. Am J Public Health 94:400–405

Woodward J (2003) Making things happen: a theory of causal explanation. Oxford University Press, Oxford

Personalized Medicine: Conceptual, Ethical, and Empirical Challenges

55

Jan Schildmann and Jochen Vollmann

Contents

Abstract

The development of so-called personalized medicine (PM) has raised great hopes and expectations among researchers, patients, health-care providers, and politicians. This chapter explores firstly the terminology and conceptual premises of PM. In the second stage, there will be a brief review of the state of the art of PM and medical-technical challenges associated with this approach to medicine. The subsequent normative analysis will focus on two topics which have been given particular consideration in the philosophical and ethical debate around PM: (1) the relation between PM, autonomy, and responsibility of the individual and (2) the setting of priorities in light of the PM approach to research and practice.

J. Schildmann (✉) • J. Vollmann
Faculty of Medicine, Institute for Medical Ethics and History of Medicine, Ruhr-Universität Bochum, Bochum, Germany
e-mail: jan.schildmann@rub.de; jochen.vollmann@rub.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_71

Introduction

Modern medicine has access to extensive genetic information about humans. The human genome was decoded in the international Human Genome Project, and technical progress in the field of sequencing technologies enables inexpensive analyses of the complete genome of an individual. Clinical medicine seeks to utilize these insights from molecular genetic research to treat patients more effectively. Knowledge about the individual genes of a patient in the field of medical diagnostics and treatment is being used to develop tailored treatments. One example involves situations in which doctors are able to determine whether or not a cancer drug will be effective against a specific tumor by identifying specific genetic biomarkers in a patient prior to starting treatment. This development which is often called "personalized medicine," "individualized medicine," or "targeted treatment" has raised great hopes and expectations among researchers, patients, health-care providers, and politicians (Collins 2010). At the same time "personalized medicine" (PM) has raised fears among others with regard to risks to informational privacy and solidarity within publicly financed health-care systems (Kollek and Lemke 2008). In recent years PM and its implications have been at the center of numerous ethical, legal, and social analyses. Given that the term PM is often used in a broad and underdetermined sense, it comes as little surprise that considerable parts of the interdisciplinary debate rest on various understandings of PM. Against this background this chapter aims firstly to shed some light on the terminology and conceptual premises of PM. In the second stage, there will be a brief review of the state of the art of PM and medical-technical challenges associated with this approach to medicine. Such analysis is important for any empirically informed, applied ethical analysis which aims to avoid discussions of scenarios which have little to do with current or expected implications of PM in practice (Fischer et al. 2015). The subsequent normative analysis will focus on two topics which have been given particular consideration in the philosophical and ethical debate around PM, namely, (1) the relation between PM, autonomy, and responsibility and (2) the setting of priorities in light of the PM approach to research and practice (Vollmann 2013).

Definition and Conceptual Aspects

The term "person" is usually understood to include psychosocial and evaluative aspects of the human being. Hence PM may be understood as a form of medicine taking into account multiple dimensions of the patient as person. However, most articles use the term as a label for strategies which are limited to biological features of individuals and according to which subgroups of patients can be stratified for the purposes of prevention, diagnosis, and treatment of disease. In recent years several study groups have sought to define PM by different methodological approaches. According to a "precising definition" by Schleidgen et al. (2013), based on the use of the term in research literature, PM is an approach to medicine which "seeks to improve tailoring and timing of preventive and therapeutic measures by utilizing

biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics" or, in a slightly adapted version, as an approach which "seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics" (Schleidgen et al. 2013). The conceptual analysis of Langanke et al. (2012) points in a similar direction. These authors define "individualized medicine" as "research approaches and health care practices, if the biomarker-based prediction of (a) diseases and/or (b) the effectiveness of therapies by stratification is central" (Langanke et al. 2012). The considerable amount of work which has been invested into definitions of PM can be seen as a reaction to the notoriously vague usage of PM in the debate. This not only presents an obstacle to the discourse on the scientific as well as public level but also hinders the development of regulation and policy on issues which are related to PM. Furthermore, and relevantly from an ethical perspective, an underspecified use of the term PM may raise hopes and fears, which often enough reflect rather the goals of interest groups rather than an interest in a sincere discourse about facts and values relevant to the development of PM (Langanke et al. 2012; Schleidgen et al. 2013).

In contrast to the above definitions, the term "personalized medicine" alludes to a kind of medical care which focuses on the health situation and the particular needs of each individual person. This is incorrect and misleading in two ways. Firstly, the molecular genetic complexity of many illnesses makes the possibility of a treatment custom tailored to each individual person very improbable, while the extremely high efforts and costs of this approach do not appear feasible in the current health-care system. What the term connotes is, therefore, not *personalized* diagnosis and treatment, but at best diagnostic and therapeutic approaches which are targeted at specific patient subgroups – for example, groups which have the same tumor biomarkers (*stratified medicine*). Secondly, medical care focused on molecular genetic characteristics has nothing to do with medical care oriented to the individual patient. *Individualization* only takes place at the molecular genetic level, but not at the personal level between doctor and patient. In order to achieve a personal treatment, the "person" of the patient should be placed at the center of treatment, and this is exactly what so-called personalized medicine does not do (Hüsing 2010; Dabrock et al. 2012). A person is not only distinguished by biological traits but also by individual psychological and social characteristics and needs. Individuals have their own lifestyles, values, and preferences (Yurkiewicz 2010). Law and ethics emphasize the normative implications of the concept of personhood, as evident in ongoing debates about so-called personhood (Lampe 1998). As a consequence, the patient in the doctor-patient relationship is entitled to adequate education and information from the doctor and has the right to consent to or to refuse a treatment (Kohnen et al. 2013). The patient's self-determined decision must be respected, even if it goes against the doctor's advice and against a medical indication, precisely because we ascribe the person these rights (Vollmann 2008).

This ethical and anthropological understanding of the term "person" is expressed by many people in their wishes about modern medicine. Patients wish to be

perceived by their doctors and by medical institutions as individual persons with wishes and normative preferences. In the citizens' report "High-Tech Medicine – What Kind of Health Care Do We Want?" of the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF), citizens demand that medical and nursing staff should have better communication skills. Furthermore, alongside the specialist subjects, mental and interpersonal aspects in day-to-day patient care must play an equal role in medical and nursing education and training and in research. The importance of taking time for the patient should be rediscovered in modern medicine (BMBF 2011; Siegmund-Schultze 2011). This broader cultural understanding of the term "person" and the wishes of citizens for personal medical care are not considered in so-called PM. The term sounds appealing, but is misleading. The intention of the inappropriate use of the term "person," which is conveyed in numerous texts and images in advertising materials, is to achieve a positive image and wide acceptance in society. It is important to debunk this questionable advertising strategy because it abuses the concept of personhood, perceives patients primarily as carriers of molecular genetically determined traits, suggests a genetic determinism for medicine (Kerr and Cunningham-Burley 2000), and aims at setting specific priorities in research funding. The latter, in particular, requires a transparent and critical discussion, as well as democratic decision making.

State-of-the-Art and Empirical Challenges

PM, which will be understood in the following as approaches in medical research and clinical practice based on biological markers such as genetic mutations and which is used for prediction of diseases and/or the effectiveness of therapies, has gained considerable success in some fields of medicine. This is in particular the case for patient subgroups in oncology. However, it is also research in this medical field which demonstrates that the vision of a "targeted treatment" seems realistic only for a minority of patients in the near future. Among the reasons for this are the multitude of genetic variations associated with a disease, the interplay between environment and genetic makeup, and mechanisms of resistance as they can be observed in patients receiving targeted treatment (Browmann et al. 2014).

An important empirical challenge for PM is to translate the findings from genome-wide association studies (GWAS) into effective preventive, diagnostic, and therapeutic measures. In GWAS large volumes of genetic and clinical data are analyzed, for example, to identify associations between biomarker and certain diseases. Subsequently companion diagnostics, a combination of a test for a biological marker and a treatment for patients who carry the biomarker, are developed. One well-known example is the companion diagnostic of HER2 and the medication trastuzumab. The biomarker HER2 is prevalent in a proportion of women with breast cancer. Health research shows that while women who carry the HER2 biomarker have a worse prognosis, they benefit from trastuzumab. This antibody targets the HER2 receptor and improves health outcomes in this patient subgroup.

In recent years a high and still increasing number of biomarkers associated with diseases have been detected. However, little resources are invested in research to establish the validity and clinical utility of these markers (Ludwig 2012). The high number of biomarkers and substances targeting these biomarkers and the investments which are required to conduct prospective trials make it unlikely that all PM interventions can be assessed according to the established criteria of evidence-based clinical medicine. A further challenge is that the small number of patients with a specific marker will make it difficult to conduct trials with a large enough number of patients to be able to demonstrate statistically significant effects of a companion diagnostic. Against this background and based on the assumed possibility of translating biological concepts, identified in the context of PM, into clinical practice, it has been suggested that the usual cascade of clinical trials requested to prove the benefit of new substances may not be necessary to prove the benefit of PM interventions (Sleijfer et al. 2013). However, critiques point out that such demand is based on undue confidence in biological models and genetic determinism. According to this position, the clinical utility of companion diagnostics needs to be established according to the same standards of evidence-based medicine as this is true for other forms of medical treatment. Given the worldwide collaborations and other resources owned by pharmaceutical companies, it seems realistic that at least for a proportion of biomarkers and companion diagnostics, the standards of evidence-based medicine could be met if the industrial sector would be willing to set respective priorities (Browmann et al. 2014; Ludwig 2012; Schildmann et al. 2015).

The above sketch of the state-of-the-art and empirical challenges sheds light on the difficulties to prove the benefit of a particular intervention within PM. Moreover it is notable that the thresholds of evidence to determine benefit (and/or harm) are themselves the focus of scientific debate. What evidence is needed to accept PM interventions as being of more benefit than harm? Is deviation from established evidentiary standards justified in light of biological models which suggest that a certain treatment targets molecular markers which are associated with a certain disease? The answers to these questions differ, in part depending also on whether they come from the perspective of a clinician, a biomedical scientist, the pharmaceutical industry, or another interest group. Furthermore, empirical analysis can inform judgments on evidence. However, ultimately there will always be a normative component when making judgments about the evidentiary level required (Browmann et al. 2014; Strech and Tilburt 2008).

Autonomous Decision Making and Responsibility Within the Context of PM

The implications of PM with regard to patient autonomy and responsibility for health have been at the center of philosophical and ethical analysis. While a comprehensive review of the debate is beyond this chapter, the remit is to explore two frequently forwarded (and criticized) claims in this debate in more detail. The

first claim is that by generating information about biomarkers associated with certain diseases and information about the effectiveness of particular treatment, PM will enable patients to make more autonomous decisions. The second claim is that the increase in knowledge about health risks associated with genetic makeup leads to obligations on the side of citizens to take more responsibility for their health. While there are links between the two claims, they will be presented and analyzed separately.

Making autonomous decisions in health care is clearly dependent on the information which is available to the patient prior to decision making. Accordingly, the ethical and legal doctrine of informed consent requires that competent patients are informed about health-related information at stake and subsequently can make decisions free from undue influence (Beauchamp and Childress 2012). In line with the account of autonomy underlying the doctrines of informed consent, one can speak about an improvement of autonomous decision making if one provides patients with more detailed health-related information. A patient with cancer can be described, for example, as making a more autonomous decision if she not only is informed about the diagnosis but also about a biomarker which is relevant for the responsiveness to a specific treatment. However, it should be noted that this account of autonomous decision making hinges on a number of premises. The first is to accept the linkage of autonomy solely with mental capabilities. As pointed out by Wabel in his analysis of different concepts of autonomy within the context of health care, such an understanding of autonomy omits to take into account that our ability to make decisions is affected by our physical experiences (Wabel 2015). Given the often intense consequences of illness on our body, Wabel suggests "embodied autonomy" as an alternative concept which takes into account the interdependence of physical experiences and decision making (Wabel 2015). Even if the more limited view that autonomous decision making is mainly linked with cognitive competences is accepted, the claim that provision of more information correlates with more autonomous decisions is open to challenges. Given the multitude and complexity of information generated in the context of PM and empirical findings which indicate that even many physicians have difficulties in understanding the clinical implications of this information (Hessling and Schicktanz 2012; Wäscher et al. 2013), it is an open question whether more information leads to more autonomous decision making by patients. Thirdly, the quality of health information generated in the context of PM needs to be taken into account when considering PM as a means to improve autonomous decision making. As pointed out in the preceding section, there is at present considerable evidentiary uncertainty with regard to the validity of many biological markers identified and their clinical utility. The combination of a high volume of information and lack of knowledge regarding the quality of generated data poses a challenge to the facilitation of informed autonomous decision making.

Following the admittedly brief and incomprehensive review of arguments for and against the claim that PM leads to more autonomous decision making, this section shall be concluded with some remarks on the often made link between PM and the call for persons' responsibility for health. The foundation of this claim is the

view that the gain of knowledge about risks associated with biomarkers and the possibility to test for these biomarkers imply an obligation for the individual to acquire knowledge about such risks and to ensure a health-related behavior which is in line with any detected risks or predispositions. Such obligation may imply that one's care about one's own health will be taken into account when considering the premium for health insurance (Rohr and Schade 2000). While the argument will be explored here within the context of PM, it should be pointed out that it is not specific to the PM approach. After all most risk factors identified on a genetic level do not differ significantly from many other risk factors. This means that a claim to take responsibility for health against the background of a particular risk factor could be made with regard to a person who carries a genetic mutation posing her at risk to a specific disease as well as with regard to a person with risk factors such as smoking. However, and in line with the analysis of Langanke et al. (2013), the connection between an increase of health-related knowledge by PM and any claims for taking more responsibility for one's health hinges on presuppositions which often are not made explicit. First of all any demand for taking into account genetic risk dispositions with regard to health-related behavior requires sufficient evidence that a particular genetic (or other biological marker) causes a certain clinical manifestation. As pointed out above, such a clear link between biomarker and disease or other clinical manifestations is given in only a few cases. Furthermore, talk about responsibility of the individual for health within the context of any PM developments makes sense only if there is knowledge that a certain health-related behavior or other intervention affects the health of the patient (in a positive manner). It does not make sense, for example, to consider responsibility for health of the individual if it has been shown that a biomarker *causes* a certain disease regardless of the health-related behavior of the individual person. Finally, it will be a matter of public and also normative debate what can be appropriately expected from an individual with a certain biomarker-based risk constellation. As pointed out by Langanke et al. (2013), any demand for responsible health-care-related behavior of the individual in light of PM generated findings will require normative justifications. Even in cases of good evidence for biomarker-associated health risks and the availability of treatment, a society would need to make ethical decisions whether, and if so, on what grounds such knowledge would imply a demand for a specific health-care-related behavior.Responsibility

Priority Setting and Opportunity Costs

Personalized medicine is frequently used as a synonym for progress and the promise of modern medicine per se and often is presented in an uncritically positive way in research, business, and the media. Public research funding has declared personalized medicine to be a priority both at the national and European level (BMBF 2013; European Commission 2013), and large pharmaceutical and biotechnology companies invest substantial amounts of money in this research. Modern medicine is facing a new "revolution" due to new scientific insights and

the close cooperation of research, clinics, and industry (Browmann et al. 2014; Hüsing 2010).

The high investment costs in research based on molecular genetic criteria raise the question of opportunity costs. This type of research ultimately provides stratified medical care that benefits subgroups of patients. Investments in this field have been made for more than a decade and, due to many open research questions, will continue to be made in the future (Rauprich 2010). Given the limited resources in the health-care sector, prioritization is required already at the research level regarding the extent of public resources that will flow into particular areas of the health-care system. A research priority in one area limits the remaining research funds for other medical speciality areas. With regard to the promotion of and funding the high costs of personalized medicine, this difficult normative and political decision is further exacerbated as at present there are only a relatively small number of patients who may benefit from these measures (Browman et al. 2011). That is why some clinical physicians are concerned that other important clinical and health-care areas, which might be beneficial for many patients, will be neglected due to the prioritized promotion of personalized medicine (Ludwig 2012). Based on previous experience, high profits can be expected from expensive cancer drugs for small patient groups (so-called niche busters), and, therefore, this approach continues to appear lucrative for the pharmaceutical industry without taking into account the health needs of the majority of patients in our health-care system.

Whereas in oncology, a small portion of patients have benefited from the innovations of personalized medicine, they have until now brought no benefit for patients in other socially and medically important disease groups. An example is the common disease type 2 diabetes: no molecular genetic descriptions of subgroups, biomarkers, and so on are superior to the usual preventive, diagnostic, and treatment options, and they do not improve the health situation of the patients affected (Schulze 2011). For such complex, multifactorial diseases, it seems unlikely that new molecular genetic insights will contribute to significant advances. Rather sociomedical care approaches and intensive public health research are needed to enable and support at-risk and affected people to adopt healthy behaviors as individuals. However, this research is seriously underfunded in our health-care system. Another example is the increasing importance of mental illness as a public health concern in our society. Mental illness and its treatment and prevention are of great significance for the patients affected, health insurance companies, and pension fund insurance companies who bear the cost for rehabilitation and for the labor market. The current care of these patients in our health-care system is under criticism due to excessively long sick-leave times, excessive waiting times for psychiatric and psychotherapy treatment and/or inpatient rehabilitation measures, and too frequent early retirements due to mental disorders. Investments are, therefore, required in research to develop new concepts for social-psychiatric prevention and treatment, for example, enabling effective prevention and early intervention at the workplace and improving the cooperation between, for example, the company doctor, primary care physician, psychiatrist, and hospital. This raises the issue of whether a society should respond to the increasing importance of mental illness primarily with high investments in molecular genetic research for

"personalized treatment" or invest at least in equal measure in social-psychiatric and mental health research, which is allocated relatively little funding in current research policy.

Therefore, from a medical ethics perspective, the existing preference for molecular genetic medicine in personalized medicine in contrast to other research fields in the publicly funded health-care system needs to be critically examined. In essence, all prioritization decisions are ethical decisions in which competing values must be weighed (Rauprich 2010). In doing so, transparency must prevail regarding who decides about what facts, which criteria are used, and on which arguments decisions are based. Therefore, it is ethically unacceptable that influential individual interests de facto determine medical research priorities and resource allocation in the publically funded health-care system; but this is exactly what is currently happening under the innocuous label of "personalized medicine." Cost-benefit assessments of the individual treatments – now often discussed – are also insufficient, since, on the basis of empirical data, they only allow statements about the medical benefits and the costs of the treatment area under investigation. In practice, the selection of the treatment area for research already frequently represents a setting of priorities within the overall spectrum of possible health-promoting measures without prior reflection on the norms involved. What is required for our health care in the future are transparent and democratically legitimized superordinate medical and research policy prioritizations.

Definition of Key Terms

Personalized medicine (synonyms: "individualized medicine," "stratified medicine")	An approach to preventive, diagnostic, and therapeutic measures in health care by which patient groups are stratified on the basis of biological markers.
Evidence-based medicine	An approach to medicine which advocates clinical decision making in medicine based on the strongest available evidence in health research such as randomized controlled trials and systematic reviews of controlled trials.

Summary Points

- "Personalized medicine" is an approach to medicine which makes use of biological marker to stratify patients into subgroups with the aim to improve prevention, diagnosis, and treatment.
- The term is misleading in two ways. Firstly, the molecular genetic complexity of many illnesses makes treatment custom tailored to each individual person very

improbable. Secondly, medical care focused on molecular genetic characteristics has nothing to do with medical care oriented on the individual patient.
- The frequently made ethical claims regarding PM as means to improve autonomous decision making and as a basis for an increase in health-related responsibility on the side of citizens or patients hinge on conceptual and empirical premises and cannot be supported without considerable qualifications.
- The high investment in research and structures necessary for PM and associated opportunity cost raises questions of justification on the spending of resources for a multidimensional approach to health care.

References

Beauchamp T, Childress J (2012) Principles of biomedical ethics, 7th edn. Oxford University Press, Oxford

Browman G, Hébert PC, Coutts J (2011) Personalized medicine: a windfall for science, but what about patients? Can Med Assoc J 183:E1277

Browmann G, Virani A, Vollmann J et al (2014) Improving the quality of 'personalized medicine' research and practice: through an ethical lens. Pers Med 11:413–423

Bundesministerium für Bildung und Forschung (2011) Bürgerreport: Hightech-Medizin – Welche Gesundheit wollen wir? Bürger-Dialog. Bundesministerium für Bildung und Forschung, Berlin

Bundesministerium für Bildung und Forschung (2013) Aktionsplan Individualisierte Medizin Ein neuer Weg in Forschung und Gesundheitsversorgung. http://www.gesundheitsforschung-bmbf.de/_media/BMBF_MASTER_Aktionsplan_IndiMed.pdf

Collins FS (2010) The language of life. DNA and the revolution in personalized medicine. HarperCollins, New York

Dabrock P, Braun M, Ried J (2012) Individualisierte Medizin: Ethische und gesellschaftliche Herausforderungen. Forum 27:209–213

European Commission (2013) Use of "-omics" technologies in the development of personalised medicine. http://ec.europa.eu/health/files/latest_news/2013-10_personalised_medicine_en.pdf. Accessed 26 June 2015

Fischer T, Dörr M, Haring R et al (2015) Alarming symptoms of a paradigm shift? An approach to bridge the gap between hypothetical ethics and the current status of individualised medicine research. In: Vollmann J, Sandow V, Wäscher S, Schildmann J (eds) The ethics of personalised medicine: critical perspectives. Ashgate, Farnham, pp 25–40

Hessling A, Schicktanz S (2012) What German experts expect from individualized medicine: problems of uncertainty and future complication in physician-patient interaction. Clin Ethics 7:86–93

Hüsing B (2010) Individualisierte Medizin – Potenziale und Handlungsbedarf. Z Evid Fortbild Qual Gesundhwes 104:727–731

Kerr A, Cunningham-Burley S (2000) On ambivalence and risk: reflexive modernity and the new human genetics. Sociology 34:283–304

Kohnen T, Schildmann J, Vollmann J (2013) Patients' self-determination in "personalized medicine": the case of whole genome sequencing and tissue banking in oncology. In: Braun M, Dabrock P (eds) "Individualised Medicine" between hype und hope. LIT, Münster, pp 97–110

Kollek R, Lemke T (eds) (2008) Der medizinische Blick in die Zukunft. Gesellschaftliche Implikationen prädiktiver Gentests. Campus Verlag, Frankfurt a. M

Lampe E-J (1998) Persönlichkeit, Persönlichkeitssphäre, Persönlichkeitsrecht. In: Lampe E-J (ed) Persönlichkeit, Familie, Eigentum: Grundrechte aus der Sicht der Sozial- und Verhaltenswissenschaften. Westdeutscher Verlag, Opladen, pp 73–102

Langanke M, Lieb W, Erdmann P et al (2012) Was ist Individualisierte Medizin? Zur terminologischen Justierung eines schillernden Begriffs. ZME 58:295–314

Langanke M, Fischer T, Erdmann P et al (2013) Gesundheitliche Eigenverantwortung im Kontext Individualisierter Medizin. Ethik Med 25:243–250

Ludwig W-D (2012) Möglichkeiten und Grenzen der stratifizierenden Medizin am Beispiel von prädiktiven Biomarkern und "zielgerichteten" medikamentösen Therapien in der Onkologie. Z Evid Fortbild Qual Gesundhwes 122:11–22

Rauprich O (2010) Rationierung unter den Bedingungen der Endlichkeit im Gesundheitswesen. In: Thomas G, Höfner M, Schaede S (eds) Endliches Leben Interdisziplinäre Zugänge zum Phänomen der Krankheit. Mohr Siebeck, Tübingen, pp 229–256

Rohr M, Schade D (2000) Selbstbestimmung und Eigenverantwortung im Gesundheitswesen – Ergebnisse des Workshops zu Forschungsbedarf im Bereich Medizin und Gesundheit, Arbeitsbericht Nr. 176. Akademie für Technikfolgenabschätzung in Baden-Württemberg, Stuttgart

Schildmann J (2015) 'Personalised medicine': multidisciplinary perspectives and in 'interdisciplinary recommendations on a framework for future research and practice'. In: Vollmann J, Sandow V, Wäscher S, Schildmann J et al (eds) The ethics of personalised medicine: critical perspectives. Ashgate Publishing, Farnham

Schleidgen S, Klinger C, Bertram T et al (2013) What is personalized medicine: sharpening a vague term based on a systematic literature review. BMC Med Ethics 14:55. doi:10.1186/1472-6939-14-55

Schulze M (2011) Programme of the international symposium: predictive genetic testing, risk communication and risk perception: Value of genetic. Information for diabetes risk prediction. Robert Koch Institut, Berlin

Siegmund-Schultze N (2011) Versorgung von Krebspatienten: Menschliche Zuwendung aufwerten. Dtsch Arztebl 108:A-932

Sleijfer S, Bogaerts J, Siu LL (2013) Designing transformative clinical trials in the cancer genome era. J Clin Oncol 31:1834–1841

Strech D, Tilburt J (2008) Value judgments in the analysis and synthesis of evidence. J Clin Epidemiol 61:521–524

Vollmann J (2008) Patientenselbstbestimmung und Selbstbestimmungsfähigkeit: Beiträge zur klinischen Ethik. Kohlhammer, Stuttgart

Vollmann J (2013) Persönlicher – besser – kostengünstiger? Kritische medizinethische Anfragen an die "personalisierte Medizin". Ethik in der Medizin 25:233–241. Parts of this publication form part of this article. With kind permission of Springer Science+Business Media

Wabel T (2015) Patient as person in personalised medicine. Autonomy, responsibility and the body. In: Vollmann J, Sandow V, Wäscher S, Schildmann J (eds) The ethics of personalised medicine: critical perspectives. Ashgate, Farnham, pp 53–64

Wäscher S, Schildmann J, Brall C et al (2013) Personalisierte Medizin in der Onkologie. Erste Ergebnisse einer qualitativen Interviewstudie zu Wahrnehmungen und Bewertungen onkologisch tätiger Ärzte. Ethik Med 25:205–214

Yurkiewicz S (2010) The prospects for personalized medicine. Hastings Cent Rep 40:14–18

Synthetic Biology and Its Envisioned Significance for Modern Medicine

56

Matthias Braun, Jens Ried, and Peter Dabrock

Contents

Abstract

Synthetic Biology (SB) is one of the leading branches within the current bundle of emerging biotechnologies. Following the hypothesis that the further development of SB will be negotiated at the interface of science and society, this chapter points out the current developments and challenges within SB by addressing the scientific as well as the societal issues.

Introduction

Synthetic Biology (SB) aims at designing and constructing new biological parts, devices, and systems as well as redesigning and modulating existing natural components with a strict focus on engineering principles. Nevertheless, up to now, there is no universally accepted definition of SB. Thus, this umbrella term covers quite disparate areas of work, including the group of "modulated

M. Braun (✉) • J. Ried • P. Dabrock
Department of Theology, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
e-mail: matthias.braun@fau.de; jens.ried@fau.de; peter.dabrock@fau.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_42

components", which elaborate and build (new) biological systems as well as the synthesis of extensive DNA strands.

Among other applications within the fields of agriculture or energy, SB substantially contributes to a so-called modern or biomarker-based medicine approach. Concrete examples are the development of novel and low-cost diagnostics and biosensors by engineering entire biological systems or DNA as nanomaterial. This material, if added to a sample of blood, urine, or water, is able to signal the presence of particular markers or pathogens. However, it is hitherto still open whether the promises and different approaches will come to the stage of concrete clinical use. Furthermore, the frame of SB is closely connected to different societal expectations, challenges, and fears. This chapter is premised by the hypothesis that the further development of SB will be negotiated at the interface of science and society. Thus, a detailed knowledge about the ethical and societal challenges is as necessary as taking note of the promising scientific progress within the field. On this account the chapter will map the field by drawing the picture of the scientific approaches and progresses and subsequently linking them with the different ethical and societal challenges.

Mapping the Diverse Field of Synthetic Biology

Synthetic Biology (SB) is supposed to be one of the leading branches within the current bundle of emerging biotechnologies (Nuffield Council on Bioethics 2012). Within the frame of SB different scientific disciplines such as physical and chemical sciences, biology, computer sciences, engineering, and biotechnological approaches are combined. This chapter maps the field of SB by firstly sketching out the overall approach of SB, secondly briefly elaborating the different pathways within SB, and thirdly plotting the current field of medical applications (see Fig. 1).

The underlying conceptual approach of SB is to gain a more in-depth and accurate understanding of biological systems. Therefore SB addresses a well-defined understanding of the organizational principles of biological organisms. By using a methodological framework of prediction, analysis, modulation, as well as by building new biological components (Kamm and Bashir 2014) SB tries to conceptualize and finally create new modularized biological systems. Thus, SB is perhaps more precisely understood if it is seen as a platform of different interacting biotechnological tools and newly modularized and constructed reagents (Cole 2014).

Within such a platform approach two different perspectives can be identified. On the one hand, there is much research activity within the so-called top-down approach to SB. This approach tries to progressively simplify cells by removing genes that are thought to be not necessary to sustain the essential properties of cellular life such as self-maintenance and self-reproduction (Purnick and Weiss 2009). The overall aim is to engineer a minimal cell, which is able to represent an organism by only comprising the lowest number of genes necessary to maintain basic cellular functions. On the other hand a so-called bottom-up approach equally

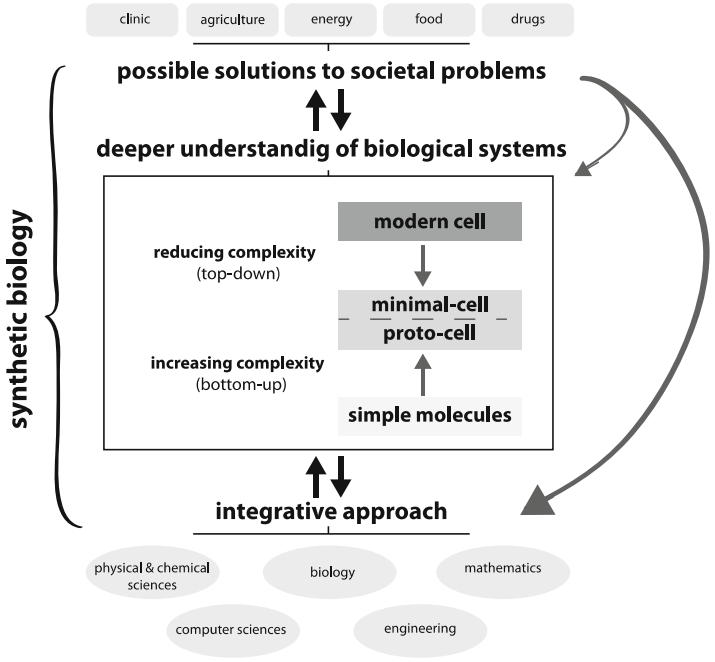


Fig. 1 Mapping the diverse field of synthetic biology. Own illustration according to Dzieciol and Mann 2012, Cole 2014, & Kamm and Bashir 2014

aims at building a certain kind of minimal cell but tries to sidestep the usage of complex cellular structures by starting with simple molecules or inorganic catalysts. This approach is also known as protocell biology (Dzieciol and Mann 2012). Protocell models, which are constructed by involving and combining simple membrane-bound and cell-like components, try to give an explanation how both a *prebiotic* – with regard to a more historical angle – and a *synthetic* cell – with regard to a more biotechnological perspective – can be designed and constructed. However, beyond these distinctions, there are many entanglements and conjunctions between these general approaches. Thus, one of the most popular, as well as critically discussed, experiments within SB, conducted by the group of Craig Venter, has been a mix of different methodological approaches and techniques (Gibson et al. 2010).

In the long term, the different disciplines and approaches within SB aim to offer a variety of diagnostic and therapeutic applications. With regard to a first scientific endeavor toward new achievements for a so-called modern medicine, SB focuses on the development of genetic circuits that link therapeutic activities to the detection of molecular disease signals. This first wave aims to pave the way to develop targeted therapeutics with increased efficacy and safety. Furthermore, first explorations indicate that synthetic control circuits may reduce the inherent tumorigenicity of stem cells (Schuldiner et al. 2003) and improve the efficiency of induced

pluripotent stem cell reprogramming (Maherali et al. 2008). Novel genetic circuits, which are capable of guiding the ex vivo construction of complex tissues, may be built in the foreseeable future as researchers are continuing to unravel the SB behind cell fate decisions (Kueh and Rothenberg 2012). Up to now, the transition of these systems to concrete medical applications has been constrained by the limited availability of devices that are able to connect synthetic circuits with information in living systems (Chen et al. 2012). While the first wave of synthetic systems focused on the development of genetic circuits that encode dynamic behavior, cellular computational operations, and biological communication channels, in the second scientific wave the current focus of research focuses on implementing SB components in diverse fields of application (Ruder et al. 2011). Within this second wave SB is starting to tackle relevant medical challenges and provides new types of diagnostic and therapeutic tools for treating significant human pathologies (Weber and Fussenegger 2012; Aurand et al. 2012) or to develop new ways to combat the increasing incidence of antibiotic-resistant bacterial infections (Krom et al. 2015). Particular attention is paid to making a contribution toward the treatment of cancer or infectious diseases, as well as to approaches in vaccine development, microbiome engineering, cell therapy, and also regenerative medicine (Ruder et al. 2011). Beyond the first achievements within a synthetic version of the antimalarial compound artemisinin (Carothers 2013) there are several projects within SB, which sustainably aim at contributing to the fight against different communicable diseases such as the human immunodeficiency virus (Hansen et al. 2013; Rerks-Ngarm et al. 2009) or to enhance the hepatitis C virus vaccine (Liang 2013) development. One example within this field is the development of a vaccine-based approach to prevent diarrheal disease (Vohra and Blakey 2013). Up to now, there is a lot of infrastructure required to provide basic sanitation. Therefore the use of synthetic oral vaccines might offer a more rapid solution to a serious global childhood health issue by reducing the need for highly trained staff as well as the requirement for a sustained cold storage chain.

In envisioning and partly fulfilling such wide-ranging approaches SB can be seen as a paradigmatic case of the so-called emerging biotechnologies. The common feature of these technologies is that they intertwine innovative and cutting-edge scientific approaches with the societal desire for new possible solutions for current unsolved medical challenges. Thereby the developments envisioned by SB fit well into an environment of science governance in which research directions are set by scientific priorities as well as by societal challenges. Against this background, increased public funding is spent within the field (Pei et al. 2012).

Ethical, Legal, and Societal Challenges Within Synthetic Biology

Within the past years there have been different agendas and approaches in order to identify possible societal challenges within SB (Deutsche Forschungsgemeinschaft et al. 2009; Presidential Commission for the study of Bioethical Issues 2010; National Research Council and National Academy of Engineering 2013; OECD

2014). These reports mainly follow – with different accentuations – the idea of the existence of four major challenges (Schmidt et al. 2009). In the present article, the important issue of dealing with big data biology is added as a fifth challenge.

First, safety and security problems are pointed out (Deutscher Ethikrat 2014). Within this topic a frequently discussed issue is the problem of a possible misuse of the results and products of SB (Douglas and Savulescu 2010). Research results, which originally aim to increase the amount of scientific knowledge, can also be used for alternative purposes. Insofar as such information, reagents, and new technological approaches have the potential to be used both for beneficial as well as for harmful purposes, the work involved is designated as "dual use research" (World Health Organization 2010). More precisely for this range of possible use and misuse the term *Dual Use Research of Concern* (DURC) has gained international customary usage (World Health Organization 2013; Deutscher Ethikrat 2014). In order to face this problem, different measures and strategies have been developed. Many scientific organizations have elaborated and implemented codes of conduct as a kind of self-regulative setting of standards (Wilholt 2012) in order to influence the actions of the respective researchers. The crucial point for a high "quality" of these regulative effects is whether the codes of conduct contain rules of law in a strict sense, or are functioning more as a voluntary self-commitment (Qi and Arkin 2014). However, up to now, the existing laws are rated to sufficiently cover the current research action (Bar-Yam et al. 2012). Additionally, there are several points – especially with regard to the top-down approach – which need particular and ongoing awareness, particularly concerning the possible ecological effects. The aforementioned critical points are, first, the differences of the physiology of natural and synthetic organisms; second, the hitherto unknown alteration of synthetic organisms in different habitats; third, the possible evolution and adaptation of the produced synthetic organisms; and fourth, the possibility of microbes to take up free DNA from the environment or to exchange their genetic material with other organisms (Dana et al. 2012). Currently there are different approaches and endeavors to provide a foundation for a safer use of synthetic biology products such as the idea to work, for example, on synthetic bacteria that are isolated from natural ecosystems by a reliance in synthetic metabolites (Mandell et al. 2015).

Second, especially with regard to the protocell approach, ethical issues from a possible blurring of cultural concepts and distinctions such as "living versus non-living matters" or "natural versus artificial" have become subject to different explorations (Dabrock et al. 2013a). Notions and metaphors such as "creating life" or "playing God" can be understood as society's attempts of finding expressions for the present significance and effect of the technological development (Pearson et al. 2011). Especially the metaphor "playing God" shows that it was neither originally nor solely the frame of SB where such metaphors were originally coined. In fact, such metaphors have rather been used throughout long periods of time and then again as a heuristic marker in the discourse on scientific or new medical and biotechnical procedures (Dabrock 2009; Coady 2009; Dworkin 2000). In the following statement, the theologian Paul Tillich points out the main issue: "The significant thing, however, is not the replacement of one metaphor by another but

the changed vision of reality, which such replacement expresses" (Tillich 1963, 15). Reformulated in a metaphor-theoretical way, the used metaphors are not only figures of speech but also forms that represent the individual and societal comprehension and constitution of reality. At this point the analyses of different metaphors used by science and society could indicate two different processes, which are caused by the emergence of new biotechnologies. On the one hand the capacity of biotechnologies may lead to profound transformations in the respective social, economic, or physical environments and therefore may have significant implications for the different shared ways of life. On the other hand, the generation of novel objects not found in nature may disturb and alter schemes of meaning and value and thereby gain potential for societal unease (Dabrock et al. 2013b).

Third, economic issues, especially in regard to questions of intellectual property (IP) and biocommercialization, have been discussed (Nuffield Council on Bioethics 2012). The conventional means through which medicinal products are developed and delivered to patients are IP-driven commercialization processes. The most common form of IP protection in biotechnology can be found in the form of patents and patent applications (Douglas and Stemerding 2013). Within such an IP-driven innovation process, two main implications with regard to the adaption of a SB approach for global health issues can be detected. The first is the possibly limited access to products that are marketed at prices that most people in "developing" or "under-developed" countries cannot afford. Second, there are only little incentives to develop drugs that will principally benefit people in those countries, since the potential users do not constitute an attractive market for pharmaceutical companies (Hollis 2013). Such a possible mismatch between access and availability need not, however, imply a break with current patent systems or intellectual property regimes. In fact, there are different approaches to develop further models, for example, alternative incentive strategies (van den Belt 2013). Open access has been seen as one of the considerable principles within such alternative strategies (van den Belt 2013): Scientists are allowed to benefit from using the developed and produced parts and components, which are available from the registry, for designing their own components and systems. In exchange, registry users are expected to share information and data on existing parts and new parts, thereby allowing the growth and improvement of this community resource (BioBricks Foundation 2013). However, the views among synthetic biologists on where to draw the line between public versus private ownership of parts and design principles differ significantly (Oye and Wellhausen 2009).

Fourth, the ethical and societal debate about dealing with emerging biotechnologies in general and SB in particular moves toward the question about who must and should be involved in making decisions pertaining to the stated questions (Pauwels 2009, 2013). Thus, it is not only at stake if the promises of SB will be fulfilled but likewise *how* and by *whom* they will and should be propelled. In other words, SB will become what scientists, innovators, regulators, funding agencies, civil society organizations, and others make of it. It could be used to foster public-value innovation or to stabilize and bolster existing power structures. The direction of this process will depend on which (group of) actors is involved and what kind of

applications will, depending on which reasons, be in the focus. Furthermore the debate about the necessity and possibilities of public participation in science fits well into a science policy environment in which research directions are set less by disciplinary priorities and more by the need to address societal challenges (Carrier and Nordmann 2011). For that reason public participation in science is not only another "nice to have" item on the agenda of assessing emerging biotechnologies but will be decisive for the question of the future trajectory of SB. Furthermore the public engagement within science could be seen as a kind of bottleneck: Steady and consistent participation of society in SB would be a strong accelerator for the development of SB. On the other hand, if society decides to minimize their participation, it will be hardly possible for SB to set the envisioned and promised aims (Jones 2014). Recently, the modes of public participation in science have become subject to change: It is now easier than ever for nonprofessionally trained people to participate in the governance, regulation, and translation of science, as well as in some of the core activities of science itself (Prainsack 2014). At this point SB could possibly take a leading position in pushing this very development: It has perhaps never been that easy to participate in, as well to contribute to, science as in the so-called do-it-yourself biology ("biohackers") or in the International Genetically Engineered Machine (iGEM) competition. However, it still remains open which concrete concept of citizen science will gain a broader acceptance. The angle ranges from *citizens as data collectors* to *citizens as ancillary scientists* to *citizens as partners* up to *citizens as full-valued scientists* (Prainsack 2014).

Fifth, perhaps one the most challenging issues about the further contribution of SB within the field of modern medicine is big data. The recent development of SB demonstrates that the lines between SB and a so-called systems medicine approach are becoming more and more blurred – if there have ever been strict distinctions (Altaf-Ul-Amin et al. 2014). Thus, the more bioinformational perspective of systems biology and the more biotechnological approach of SB are becoming widely intertwined. In order to scrutinize how the different molecules and synthetic components could fit together, a massive set of data and backups about small molecules, proteins, and genes is needed. Up to now, there are enduring challenges for handling, processing, and moving this complex information as well as of the simulation clusters (Schadt et al. 2010). At this point the most puzzling problems are, on the one hand, the heterogeneity of the biological data caused by a wide range of experiments, which reveal many different and nonstandardized types of information (Marx 2013a). On the other hand, huge biological data and analysis volumes have to be stored via cloud computing while scientists are aware that there are risks of biohacking (Marx 2013b). Beneath these more technical challenges a so-called big data biology is also supposed to create a radical shift in how society thinks about research (Boyd and Crawford 2012). Therefore, big data biology reframes key questions about the constitution of knowledge, the processes of research, how societies can and should engage with information, the understanding and the categorization of reality, as well as challenges for the understanding of privacy issues (Dabrock 2012). The entanglement of these five points offers a sufficient and well-suited perspective to map the recent as well as the upcoming challenges within SB.

In connection to these intertwined five key aspects, the further development of SB is not only a question of the ongoing as well as predictable scientific progress, but it will also be determined by the societal estimation and appraisal of SB (Nuffield Council on Bioethics 2012). Therefore, the ethical and societal assessment of SB is challenged not only in terms of one or two of the outlined aspects, such as questions of biosafety and biosecurity (Douglas and Savulescu 2010) or intellectual property issues. Rather, SB has to be understood as a technological field at the interface of science and society, which is triggered by scientific progress as well as societal concerns, expectations, and unease. These societal hopes, fears, and expectations are linked to the general label of SB, even if the concrete formation of societal unease may differ with regard to the different fields of application, as well as to the different perspectives concerning, for example, protocell or the minimal cell approach. Apart from that, different agents, such as scientists, civil society organizations, and political decision-makers, have variable expectations toward SB (Jones 2014). Furthermore, the appraisal of and the attitude toward SB seems to also be strongly linked to the different fields of application (European Commission 2014). Up to now, it still remains unclear whether SB will be associated with the so-called red (medical application) or green (environmental application) biotechnology.

Definitions of Key Terms

Synthetic Biology	SB is an umbrella term covering quite disparate areas of work, which aim to design and construct new biological parts, devices, and systems as well as to redesign and modulate existing natural components with a strict focus on engineering principles.
Emerging Biotechnologies	Although emerging biotechnologies vary widely in nature and purpose, they jointly aim at bringing together a broad field of knowledge, a specific frame of research, and a more or less envisioned application of the respective techniques as well as the possible development of future products.
Citizen Science	Citizen Science is a term for all those endeavors which aim to scrutinize, observe, as well as improve public participation with science.
Big Data Biology	In systematically combining biological approaches, big data sets, and predictive elements big data biology reframes key questions about the constitution of knowledge, the processes of research, how societies can and should engage with information, the understanding and the categorization of reality, as well as challenges for the understanding of privacy issues.

Dual Use Research of Concern	Research results, which originally aim to increase the fund of scientific knowledge, can also be used for alternative purposes. Insofar as such information, reagents, and new technological approaches have the potential to be used both for beneficial as well as for harmful purposes, the work involved is designated as "dual use research."

Summary Points

- Synthetic Biology (SB) is a diverse field of research with different agendas and approaches integrating different disciplines and methods. SB aims to design and construct new biological parts, devices, and systems as well as to redesign and modulate existing natural components with a strict focus on engineering principles.
- In the long term, the different disciplines and approaches within SB aim to offer a variety of diagnostic and therapeutic applications. Therefore one of the basic endeavors of SB is the development of genetic circuits that link therapeutic activities to the detection of molecular disease signals in order to make them prospectively usable for medical applications.
- Regarding the societal impact of SB, five major challenges can be detected: first, biosafety and biosecurity issues; second, ethical issues from a possible blurring of cultural concepts and distinctions such as "living versus non-living" or "natural versus artificial"; third, economic issues, especially regarding questions of intellectual property (IP) and biocommercialization; fourth, the question about who must and should be involved in making decisions pertaining to further developments; fifth, and perhaps as one the most challenging issues about the further contribution of SB within the field of modern medicine, the issues of a so-called big-data-biology.
- Entangled with these five key aspects the further development of SB and its possible contribution to the development of medical applications is not only a question of the ongoing as well as predictable scientific progress but also a question of the determining force of societal estimation and appraisal of SB.

Acknowledgments This work is part of the MaxSynBio Consortium, which is jointly funded by the German Federal Ministry of Education and Research and the Max Planck Society.

References

Altaf-Ul-Amin M, Afendi FM, Kiboi SK, Kanaya S (2014) Systems biology in the context of big data and networks. BioMed Res Int. doi:10.1155/2014/428570

Aurand TC, Russell MS, March JC (2012) Synthetic signaling networks for therapeutic applications. Curr Opin Biotechnol 23(5):773–779

Bar-Yam S, Byers-Corbin J, Casagrande R, Eichler F, Lin A, Oesterreicher M, Regardh P, Turlington RD, Oye KA (2012) The regulation of synthetic biology. A guide to United States

and European Union regulations, rules and guidelines. SynBERC and iGEM Version 9.1, 10 Jan 2012 [Online]. Available: http://synberc.org/sites/default/files/Concise%20Guide%20to%20Synbio%20Regulation%20OYE%20Jan%202012_0.pdf. Accessed 29 Oct 2014

Biobricks Foundation (2013) Frequently asked questions [Online]. Available: https://biobricks.org/bpa/faq/#top. Accessed 31 Oct 2014

Boyd D, Crawford K (2012) Critical questions for big data. Inf Commun Soc 15(5):662–679

Carothers JM (2013) Design-driven, multi-use research agendas to enable applied synthetic biology for global health. Syst Synth Biol 7(3):79–86

Carrier M, Nordmann A (2011) Science in the context of application: methodological change, conceptual transformation, cultural reorientation. In: Carrier M, Nordmann A (eds) Science in the context of application. Springer, Dordrecht/New York, pp 1–11

Chen YY, Galloway KE, Smolke CD (2012) Synthetic biology: advancing biological frontiers by building synthetic systems. Genome Biol 13:240–250

Coady CAJ (2009) Playing God. In: Bostrom N, Savulescu J (eds) Human enhancement. Oxford University Press, Oxford, pp 155–180

Cole JA (2014) Synthetic biology: old wine in new bottles with an emerging language that ranges from the sublime to the ridiculous? FEMS Microbiol Lett 351:113–115

Dabrock P (2009) Playing God? Synthetic biology as a theological and ethical challenge. Syst Synth Biol 3:47–54

Dabrock P (2012) Privacy, data protection, and responsible government. Key issues and challenges in biobanking. Public Health Genomics 15:227–312

Dabrock P, Braun M, Ried J (2013a) From functional differentiation to (re-)hybridisation. The challenges of bio-objects in synthetic biology. In: Greif H, Weiss M (eds) Ethics – society – politics. De Gruyter, Berlin, pp 347–379

Dabrock P, Braun M, Ried J, Sonnewald U (2013b) A primer to 'bio-objects': new challenges at the interface of science, technology and society. Syst Synth Biol 7(1–2):1–6

Dana GV, Kuiken T, Rejeski D, Snow AA (2012) Synthetic biology: four steps to avoid a synthetic-biology disaster. Nature 483:29

Deutsche Forschungsgemeinschaft, Acatech, Deutsche Akademie Der Naturforscher Leopoldina (2009) Synthetic Biology Online. Available: http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2009/stellungnahme_synthetische_biologie.pdf. Accessed 27 Oct 2014

Deutscher Ethikrat (2014) Biosecurity – freedom and responsibility of research [Online]. Available: http://www.ethikrat.org/files/opinion-biosecurity.pdf. Accessed 27 Oct 2014

Douglas T, Savulescu J (2010) Synthetic biology and the ethics of knowledge. J Med Ethics 36:687–693

Douglas CMW, Stemerding D (2013) Governing synthetic biology for global health through responsible research and innovation. Syst Synth Biol 7:139–150

Dworkin R (2000) Sovereign virtue. The theory and practice of equality. Harvard University Press, Cambridge

Dzieciol A, Mann S (2012) Designs for life: protocells models in the laboratory. Chem Soc Rev 41:79–85

European Commission (2014): Special Eurobarometer 419: public perceptions of science, research and innovation [Online]. Available: http://ec.europa.eu/public_opinion/archives/ebs/ebs_419_en.pdf. Accessed 27 Oct 2014

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi Z-Q, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329(5987):52–56

Hansen SG, Ford JC, Lewis MS, Ventura AB, Hughes CM, Coyne-Johnson L, Whizin N, Oswald K, Shoemaker R, Swanson T, Legasse AW, Chiuchiolo MJ, Parks CL, Axthelm MK, Nelson JA, Jarvis MA, Piatak MJ, Lifson JD, Picker LJ (2013) Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. Nature 473(7348):523–527

Hollis A (2013) Synthetic biology: ensuring the greatest global value. Syst Synth Biol 7:99–105
Jones RAL (2014) Reflecting on public engagement and science policy. Public Underst Sci 23 (1):27–31
Kamm RD, Bashir R (2014) Creating living cellular machines. Ann Biomed Eng 42(2):445–459
Krom RJ, Bhargava P, Lobritz MA, Collins JJ (2015) Engineered phagemids for nonlytic, targeted antibacterial therapies. Nano Lett 15(7):4808–4813
Kueh H, Rothenberg E (2012) Regulatory gene network circuits underlying T cell development from multipotent progenitors. Wiley Interdiscip Rev Syst Biol Med 4:79–102
Liang TJ (2013) Current progress in development of hepatitis C virus vaccines. Nat Med 19 (7):869–878
Maherali N, Ahfeldt T, Rigamonti A, Utikal J, Cowan C, Hochedlinger K (2008) A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell 3:340–345
Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddard BL, Church GM (2015) Biocontainment of genetically modified organisms by synthetic protein design. Nature 518(7537):55–60
Marx V (2013a) Biology: the big challenges of big data. Nature 498:255–260
Marx V (2013b) Genomics in the clouds. Nat Methods 10(10):941–954
National Research Council & National Academy Of Engineering (2013) Positioning synthetic biology to meet the challenges of the 21st century. Summary report of a six academies symposium series. The National Academies Press, Washington, DC
Nuffield Council on Bioethics (2012) Emerging biotechnologies: technology, choice and the public good [Online]. Available: http://nuffieldbioethics.org/wp-content/uploads/2014/07/Emerging_biotechnologies_full_report_web_0.pdf. Accessed 24 Oct 2014
OECD (2014) Emerging policy issues in synthetic biology [Online]. OECD Publishing. Available: http://dx.doi.org/10.1787/9789264208421-en. Accessed 30 Oct 2014
Oye KA, Wellhausen R (2009) The intellectual commons and property in synthetic biology. In: Schmidt M (ed) Synthetic biology. Springer, Berlin, pp 121–139
Pauwels E (2009) Review of quantitative and qualitative studies on U.S. public perceptions of synthetic biology. Syst Synth Biol 3:37–46
Pauwels E (2013) Public understanding of synthetic biology. BioScience 63(2):79–89
Pearson B, Snell S, Bye-Nagel K, Tonidandel S, Heyer LJ, Campbell AM (2011) Word selection affects perceptions of synthetic biology. J Biol Eng 5:9–11
Pei L, Gaisser S, Schmidt M (2012) Synthetic biology in the view of European public funding organisations. Public Underst Sci 21(2):149–162
Prainsack B (2014) Understanding participation: the 'citizen science' of genetics. In: Prainsack B, Schicktanz S (eds) Genetics as social practice. Ashgate, Farnham, pp 147–164
Presidential Commission for the Study of Bioethical Issues (2010) New directions the ethics of synthetic biology and emerging technologies Online. Available: http://bioethics.gov/sites/default/files/PCSBI-Synthetic-Biology-Report-12.16.10_0.pdf. Accessed 27 Oct 2014
Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol 10:410–422
Qi LS, Arkin AP (2014) A versatile framework for microbial engineering using synthetic non-coding RNAs. Nat Rev Microbiol 12:341–453
Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, Premsri N, Namwat C, De Souza M, Adams E, Benenson M, Gurunathan S, Tartaglia J, Mcneil JG, Francis DP, Stablein D, Birx DL, Chunsuttiwat S, Khamboonruang C, Thongcharoen P, Robb ML, Michael NL, Kunasol P, Kim JH (2009) Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med 361(23):2209–2220
Ruder WC, Lu T, Collins JJ (2011) Synthetic biology moving into the clinic. Science 333:1248–1252
Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP (2010) Computational solutions to large-scale data management and analysis. Nat Rev Genet 11:647–657

Schmidt M, Ganguli-Mitra A, Torgersen H, Kelle A, Deplazes A, Biller-Andorno N (2009) A priority paper for the societal and ethical aspects of synthetic biology. Syst Synth Biol 3:3–7

Schuldiner M, Itskovitz-Eldor J, Benvenisty N (2003) Selective ablation of human embryonic stem cells expressing a "suicide" gene. Stem Cells 21:257–265

Tillich P (1963) Systematic theology, volume three: life and the spirit history and the kingdom of God. University of Chicago Press, Chicago

van den Belt H (2013) Synthetic biology, patenting, health and global justice. Syst Synth Biol 7:87–98

Vohra P, Blakey G (2013) Easing the global burden of diarrhoeal disease – can synthetic biology help? Syst Synth Biol 7(3):73–78

Weber W, Fussenegger M (2012) Emerging biomedical applications of synthetic biology. Nat Rev Genet 13:21–35

Wilholt T (2012) Die Freiheit der Forschung. Begründungen und Begrenzungen. Suhrkamp, Frankfurt a.M

World Health Organization (2010) Responsible life sciences research for global health security. A guidance document [Online]. Geneva. Available: http://whqlibdoc.who.int/hq/2010/WHO_HSE_GAR_BDP_2010.2_eng.pdf. Accessed 29 Oct 2014

World Health Organization (2013) Informal consultation on dual use research of concern [Online]. Available: http://www.who.int/csr/durc/consultation/en. Accessed 29 Oct 2014

Complementary and Alternative Medicine (CAM) and Its Relationship to Western Medicine

57

Pekka Louhiala

Contents

Abstract

CAM is an acronym combining two terms, "complementary medicine" and "alternative medicine," both of which are recent. The definitions of CAM point out the diversity of phenomena behind the concept and list therapies currently belonging to the CAM field. A universal definition that would provide a demarcation line between CAM and the dominant system does not exist, and CAM is best understood as a residual category, defined by its exclusion from "official" or "medical school" medicine. Some CAM treatments are fundamentally incompatible with science, but some treatments, currently belonging to the CAM domain, will, sooner or later, be included in mainstream medicine, if their effectiveness can be demonstrated. CAM as a concept may be useful in

P. Louhiala (✉)
Department of Public Health, University of Helsinki, Finland
e-mail: pekka.louhiala@helsinki.fi

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_47

describing a phenomenon from a sociological or political point of view, but from the scientific perspective there is only one medicine.

That is why there are, and always will be, pseudo-healers, wise women, homeopaths, and allopaths. (Tolstoy 2001, p. 518)

Introduction

CAM is an acronym combining two terms, "complementary medicine" and "alternative medicine," both of which are only a few decades old. "Alternative medicine" first appeared in medical journals in 1975 and "complementary medicine" in 1985. "Integrative medicine" was introduced in an English language journal in 1995, although it had appeared in German in an article 2 years earlier and in French already in 1951 (Louhiala and Puustinen 2012).

"Alternative medicine" as a term dates back to the alternative lifestyle movement that originated in the United States in the late 1960s (Issit 2009). "Complementary medicine" was adopted in Britain with the political objective of raising the question of whether medicine could include some of the alternative healing practices in its tool kit. "Integrative medicine" was introduced in order to suggest a deeper relationship between alternative treatments and medicine (Louhiala and Puustinen 2012).

In general, the "alternative movement" was part of a societal trend toward the rejection of science as a method of determining truths. Within the movement, it was also often asserted that "scientific medicine" (or "conventional medicine") is only one of a vast array of options in health care. The movement was ideologically close to the view that science is not necessarily more valid than pseudoscience.

All of the commonly used terms in the debate concerning medicine and its "alternatives" are problematic in one way or another, and often they describe both the phenomenon in question and the motives of the person using the term (Louhiala 2010).

Firstly, the nature of "alternative" in "alternative medicine" is anything but clear. Advocates of the term usually fail to define what they claim to offer an alternative to and on what grounds. In their rhetoric, medicine is presented as a monolithic and closed system that needs an alternative.

Secondly, choosing an alternative means that, in general, the other option is rejected. If we want, for example, to travel from London to Paris, there are several alternatives. One is fast, one is cheap, one is environment-friendly, etc. They all do, however, take us from London to Paris. There are also many ways to explore these alternatives scientifically. If a genuine "alternative" medicine existed, it should produce results that are similar to those of ordinary medicine, and a causal correlation between the treatment and the result should be demonstrable.

Thirdly, there are similar problems with all the other terms, too. If someone introduces herself as a practitioner of *complementary* medicine, is she not implicitly

saying that she masters both ordinary medicine *and* some additional methods that do not belong to the toolbox of the majority of physicians? *Traditional* medicine can mean almost anything, and *official* or *school* medicine refers to medical education in a particular area and at a specific time. If *evidence-based* medicine is defined as "the conscientious, explicit and judicious use of current best evidence in making decisions about the medical care of individual patients," there is certainly not a physician alive who would not claim to practice it, as Mark Tonelli (1998) has remarked. And if *orthodoxy* in medicine were determined by the durability and degree of acceptance achieved by any particular medical idea, then humoral medicine would represent "orthodox Western medicine" *par excellence* (Bivins 2007). It certainly endured unchallenged far longer than biomedicine, which is less than 200 years old.

Some Historical Remarks

Various ideas and practices concerning health and illness have occurred throughout history, and they have often contradicted each other and offered alternative means with which to understand and alleviate illness and suffering. In the CAM rhetoric, it is not uncommon to claim that the practices marketed under these terms date back thousands of years (Larson 2007) or at least to the medical disputes of the eighteenth century. However, naming ancient ideas and practices as alternative, complementary, or integrative medicine is problematic, since none of the terms were in use prior to the 1970s (Louhiala and Puustinen 2012).

In order to have medical systems and practices that can properly be regarded as "alternative," one must have a recognized and at least relatively stable orthodoxy to which they oppose themselves. Such an orthodoxy emerged in the Western medical marketplace only in the nineteenth century, the "Paris School" being often identified as a starting point of modern scientific medicine (Bivins 2007).

The quacks of the eighteenth century did not present their medicines or therapies as "alternative" to those of orthodox physicians, apothecaries, or surgeons, but as better (Bivins 2007). They did not propose different medical systems nor different understandings of diseases, but argued that their remedies simply operated more effectively than their competitors. The common and dominating medical system was humoral medicine, which persisted in orthodox practice until the mid-nineteenth century. In fact, many "alternative" therapies of today have their roots in humoral medicine, although this is not the case with homeopathy.

In the early nineteenth century, there was no scientific medicine in the modern sense of the term. Closest to its idea came "allopathy," a term invented by Samuel Hahnemann, the German physician and founder of homeopathy. In Hahnemann's terminology, allopathy meant "treatment with opposites," while the basic principle of homeopathy was "like treats like" (*similia similibus curantur*). According to Roberta Bivins (2007), homeopathy made, along with mesmerism, strong claims to scientificity and was popular with the same educated consumers who also eagerly supported the natural sciences. Homeopathy's commercial and therapeutic successes forced major changes in ordinary medical practice. Sir John Forbes, a

prominent physician of his time, noted in 1858 that "the favourable practical results obtained by the homoeopathists – or to speak more accurately, the wonderful powers possessed by the natural restorative agencies of the living body, demonstrated under their imaginary treatment – have led to several other practical results of value to the practitioners of ordinary medicine" (cited in Bivins 2007, 99).

The early homeopaths used the rhetoric of opposition to – and oppression by – medical orthodoxy to draw attention to the flaws of allopathic practice, which allowed them to build a strong identity. However, at the same time, they left homeopathy open to being grouped with all the other self-proclaimed "alternatives," some of which deserved respect, while others did not (Bivins 2007).

The origin of the notion of alternative medicine can be traced back to the late 1960s in the United States where, especially among college students, strong critique arose against a bourgeois lifestyle and values in the wake of the Vietnam War and the threat of a nuclear holocaust. Some authors pinpointed the heyday of this cultural phenomenon to the summer of 1968, when tens of thousands of youth drifted to San Francisco to join a spontaneous gathering that was named the hippie movement by the American press and was referred to the New Age by the proponents of this subculture (Issit 2009).

What started as a hippie or New Age movement with ideals of peace, freedom, and "planetary consciousness" soon lost its momentum and split into various diverse expressions of discontent for the mainstream American way of life and its values. One common denominator was the need to find alternatives to current housing, farming, food consumption, family structure, child-rearing, schooling, etc. Criticism of medical theory and practice can be seen as a part of this general development. After all, medicine in those days was male dominated and an increasingly technologically based activity, both of which were associated with the political and military power structure of the time.

The attempt to seek ways to meet the need for healing practices that were free of medical dominance led to the adoption of various indigenous healing systems, some of which were imported as side products of Eastern religions, especially Buddhism and Hinduism. Spiritual teachers of these religions had been imported to the United States ever since the late 1960s. Along with their cosmological views, they produced ideas on health, illness, and healing that were based on their general world view. In 1973, this development was boosted by President Nixon's visit in the People's Republic of China, where acupuncture was introduced to the West (although the practice itself had been known in Europe for centuries).

After this general development, there suddenly emerged a growing demand for nonmedical healing practices among the affluent, younger generation both in the United States and in Europe in the early 1970s. The rest of humankind relied, as they still do, on local indigenous healers and medical help, when available and affordable. Since there was no official training available with which to gain competence in these newly commercialized healing practices, self-appointed practitioners and trainers appeared who offered courses and diplomas in Eastern and other practices that were more or less adapted to Western taste. This situation not only led to competition between practitioners but also to competition with the

medical establishment. This competition took place mainly in the media, which uncritically applied catchwords such as alternative, natural, soft, and holistic.

Current Definitions

In the light of the history of CAM and related terms, it is obvious that a universal definition that would provide a *demarcation line* between CAM and the "dominant system" cannot be reached.

The Committee on the Use of Complementary and Alternative Medicine of the American Public Board on Health Promotion and Disease Prevention (2005) defined CAM as

> . . . a broad domain of resources that encompasses health systems, modalities, and practices and their accompanying theories and beliefs, other than those intrinsic to the dominant health system of a particular society or culture in a given historical period. CAM includes such resources perceived by their users as associated with positive health outcomes. Boundaries within CAM and between the CAM domain and the domain of the dominant system are not always sharp or fixed.

The definition is an important description in pointing out the diversity of the phenomena behind the concept. This complexity also explains why the boundaries between the CAM domain and mainstream medicine are neither sharp nor constant.

Within the Cochrane Collaboration (a global independent network producing health information), an *operative* classification of CAM has been developed. It consists of a long list of therapies that the Cochrane Complementary Medicine Field classifies as complementary or alternative. The therapies are listed in alphabetical order, starting from açaí, acupressure, and acupuncture and ending with zinc supplements, Zishen Tongli Jianonang (a Chinese herbal medicine), and zone therapy (http://www.compmed.umm.edu/cochrane/CAM.asp). The authors of the list do not consider it to be exhaustive and point out that it is subject to expansion and elaboration over time. In fact, they question whether it is possible to arrive upon a definitive set of therapies that are universally agreed upon as CAM.

The above definitions aim to be neutral and descriptive, making no a priori claims about the ideological background or effectiveness of CAM therapies. The diverse nature of the therapies is acknowledged also among the representatives of CAM, but unsubstantiated claims about a shared ideology are often made in the literature. In an entry in the Encyclopedia of Applied Ethics, for example, it is claimed that CAM therapies "share a similar approach to treatment which differs fundamentally from that of orthodox medicine" (Whitelegg 1998). The author goes on to paint a black-and-white picture of the world as follows:

> . . .the biomedical perspective of practical exclusion of nonphysical factors as agents influencing either the cause or the progress of illness bases its treatment on rational and objective observation and evaluation with no interference from subjective influences. Complementary medicine, on the other hand, sees the patients and their problems as inextricably linked with their circumstances and their individual reactions to them and

> their lifestyles, attitudes, and environments, and will consider body, mind, and spirit in its treatment.

Reflecting aspects from all the definitions above, Wolpe (2002) suggested that CAM is best understood as a "residual category," which means that it is defined by its exclusion from "official" or "medical school" medicine. "Alternative medicine" was defined along these lines in 1998 in a large study on national trends on the use of alternative medicine in the United States (Eisenberg et al. 1998):

> Alternative medical therapies, functionally defined as interventions neither taught widely in medical schools nor generally available in US hospitals. . .

In a thorough and critical article, Stephen Barrett (1998) accepted the category "alternative medicine" but suggested that, "to avoid confusion, 'alternative' methods should be classified as genuine, experimental, or questionable." In his terminology,

> *genuine* alternatives are comparable methods that have met science-based criteria for safety and effectiveness. *Experimental* alternatives are unproven but have a plausible rationale and are undergoing responsible investigation.. . . *Questionable* alternatives are groundless and lack a scientifically plausible rationale.... The archetype is homeopathy.

Barrett's classification is meaningful if the category "alternative medicine" is taken for granted. Some authors argue that such a category is not needed, and it is simpler and more useful to distinguish between *mechanisms not fully understood* and *mechanisms obviously absurd*. Hrobjartsson and Brorson (2002), for example, have written:

> If a postulated mechanism is absurd according to standard scientific position, there is a tendency to ascribe a prior probability of zero to a hypothesis about therapeutic effects, for example in the case of homeopathy. . . Other complementary/alternative therapies, for example acupuncture, are also based on theories foreign to conventional science, but are not obviously absurd: physiological responses caused by the insertion of needles on certain spots are not necessarily incompatible with standard scientific thinking. Therefore, the prior probability of acupuncture to have clinical effects exceeds zero.. . .

Hrobjartsson and Brorson's view may, however, be too simple. The history of science provides plenty of examples of ideas that were originally dismissed as absurd and persisted as anomalies, only for new research to eventually provide sufficient support a mechanism to be proposed. The "standard scientific position" may be wrong and exclude the possibility of the maverick thinking that leads to paradigm shifts in science.

A political implication of the categorization by Hrobjartsson and Brorson would be that public money should not be invested in research on methods that are based on obviously absurd mechanisms. Although a preliminary clinical trial may not need to be expensive, it nevertheless implies an allocation of intellectual and economic resources. If the probability of getting positive results is very low, the enrollment of patients and the allocation of resources raise both ethical and socioeconomic problems.

CAM describes thus a *political* or *sociological* category, but it is also an example of a *buzzword*, which, according to Merriam-Webster Dictionary (2015), is "an important-sounding usually technical word or phrase often of little meaning used chiefly to impress laymen." The advocates of various forms of CAM are adept at

using also other buzzwords and slogans like "natural," "soft," or "holistic," the meaning of which is vague (Louhiala and Puustinen 2012).

Natural, Soft, and Holistic

The three main concepts with which the advocates of alternative medicine have justified their products and treatments have been "natural," "soft," and "holistic." "Official medicine," on the other hand, has been considered "unnatural," "hard," and "fragmented."

To name and treat illnesses is a cultural phenomenon. In that sense, there are no treatments available in nature and all treatments are unnatural. As *Pneumococci* or HI viruses multiply in a patient's body, it is a fully natural phenomenon. When trying to interfere with their flourishing, we act against nature, no matter whether we use antibiotics, herbal remedies, or prayers. The often used claim by the proponents of CAM, that their methods act through strengthening the body rather than through killing the germs directly, does not change that fact. It only leaves the work to be done by the body rather than by antibiotic pills.

The problem with the term "soft" in this context is that it supposedly refers to the treatment used and not to the therapist in charge. In the light of the definitions of CAM, it is obvious that "softness" is by no means not a common factor between different CAM modalities. From the point of view of the patients, individual therapists of CAM as well as conventional medicine can be soft or hard in their practice.

The term "holistic" appears at least as often as "soft" in the rhetoric of CAM. The term is, again, offered as an antithesis to conventional medicine's alleged lack of a holistic feature. It seems to be an empty slogan that does not describe essential and common features of a multitude of treatments in the category CAM.

CAM Meets EBM

Evidence-based medicine (EBM) originated from the concern that numerous ineffective treatments had been adopted by mainstream medicine, and the randomized controlled trial (RCT) was viewed as the most reliable method by which to identify treatments that actually work.

The ideas behind EBM are old, but the concept was introduced to the wider medical community in 1992 as "a new approach to teaching the practice of medicine" and "a new paradigm for medical practice" (Evidence-Based Medicine Working Group 1992). According to the authors, EBM "de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making and stresses the examination of evidence from clinical research." In particular, the 1992 paper instructed clinicians to search for studies with the question "Was the assignment of patients to treatments randomized?" The article was a bold program statement that divided the medical world into the old-fashioned pre-EBM and the revolutionary new EBM types of medicine.

A definition of EBM was formulated by the pioneers 4 years later: "Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett et al. 1996). Despite its obvious vagueness, this has remained the most widely cited definition of EBM.

The story of the concept has been a success, although during all these years, it has not been clear what the phenomenon behind the three letters actually is. From the very beginning, critical voices were also heard, and already in 1998, a paper titled "The Rise and Fall of EBM" was published (Charlton and Miles 1998).

Many years and several definitions later, it is obvious that Timmermans and Mauck (2005) were right when they wrote that "The term [EBM] is loosely used and can refer to anything from conducting a statistical meta-analysis of accumulated research, to promoting randomized clinical trials, to supporting uniform reporting styles for research, to a personal orientation toward critical self-evaluation."

Despite the disagreements and confusions about the basic definitions of EBM, one aspect of the EBM program has been particularly essential since the introduction of the term in 1992, namely, the view of valid evidence: "[C]omparative clinical studies, preferably from randomised trials [RCTs], are deemed to provide better evidence than mechanistic reasoning and clinical experience" (Evidence-Based Medicine Working Group 1992).

CAM and EBM are often presented as opposites, at least by the representatives of mainstream medicine, who claim to practice EBM, but often fail to define what they exactly refer to. As we have seen, neither CAM nor EBM has been defined in a satisfactory way that would give us a demarcation line between CAM and non-CAM or EBM and non-EBM.

A more pragmatic approach can be taken, however, to explore the relationship between CAM and EBM at the level of medical practice (Louhiala and Hemilä 2014). Rather than opposites, they could be seen as concepts pointing at different directions. If CAM is understood to mean therapies that lie outside mainstream medicine and EBM is understood in the light of its main principle, the requirement to base treatments on RCTs, there are, in fact, evidence-based therapies that are currently listed as CAM (e.g., high-dose zinc acetate for common cold (Hemilä 2011) or vitamin C for patients with exercise-induced asthma (Hemilä 2013)).

The opposite of CAM is thus not EBM but "mainstream medicine," and some treatments obviously belong to the CAM domain for historical reasons and because of preconceptions within mainstream medicine.

Concluding Remarks

Many treatments currently classified as CAM are not credible from the scientific point of view, and there are good reasons for them to remain outside mainstream medicine. However, the fact that a specific treatment falls into the CAM domain does not prove that the treatment is ineffective.

Some CAM treatments, such as homeopathy, are fundamentally incompatible with science. It is extremely unlikely that such treatments will ever become part of mainstream medicine, even if some occasional research findings have been positive.

Publication bias and methodological flaws are far more plausible explanations for the positive results related to homeopathy than errors in basic theories of science.

On the other hand, it is likely that some treatments currently belonging to the CAM domain will, sooner or later, be included in mainstream medicine, if their effectiveness can be demonstrated.

CAM as a concept may be useful in describing a phenomenon from a sociological or political point of view, and people in a pluralistic society should be free to choose whatever treatments they like, also CAM. From the scientific point of view, however, there is only one medicine, and the alternativity of "alternative medicine," complementarity of "complementary medicine," and integrativity of "integrative medicine" are not based on any meaningful theoretical or practical line of division.

Definition of Key Terms

Complementary and alternative medicine (CAM)	Has been defined as a "broad domain of resources that encompasses health systems, modalities, and practices and their accompanying theories and beliefs, other than those intrinsic to the dominant health system." Another possibility is an operative definition listing therapies currently classified as complementary or alternative. These definitions overlap and do not provide a demarcation line between CAM and non-CAM.
Alternative medicine	Has also been defined functionally as "interventions neither taught widely in medical schools nor generally available in US hospitals." Reflecting this, it has been suggested that CAM is a residual category, defined by its exclusion from official medicine.

Summary Points

- CAM is an acronym combining two terms, "complementary medicine" and "alternative medicine," both of which are recent.
- Alternative medicine as a term dates back to the alternative lifestyle movement that originated in the United States in the late 1960s.
- All of the commonly used terms in the debate concerning medicine and its "alternatives" are problematic, and a universal definition that would provide a demarcation line between CAM and the "dominant system" cannot be reached.
- Rather than a scientific category, CAM describes a political or sociological category.
- CAM is also a buzzword, used to promote individual treatments.
- Many treatments currently classified as CAM are not credible from the scientific point of view.

- The fact that a specific treatment currently falls into the CAM domain does not prove that the treatment is ineffective.

References

This chapter builds partly upon, and contains small extracts from, Louhiala (2010), Louhiala and Puustinen (2012), and Louhiala and Hemilä (2014). The extracts are used with permission from the publishers and the co-authors

Barrett S (1998) "Alternative" medicine: more hype than hope. In: Humber JM, Almeder RF (eds) Alternative medicine and ethics. Human Press, Totowa, pp 1–42

Bivins R (2007) Alternative medicine – a history. Oxford University Press, Oxford

Charlton BG, Miles A (1998) The rise and fall of EBM. Q J Med 91:371–374

Committee on the Use of Complementary and Alternative Medicine by the American Public, Board on Health Promotion and Disease Prevention (2005) Complementary and alternative medicine in the United States. National Academy Press, Washington, DC, p 17

Eisenberg DM, Davis RB, Ettner SL et al (1998) Trends in alternative medicine use in the United States, 1990–1997: results of a follow-up national survey. JAMA 280:1569–1575

Evidence-Based Medicine Working Group (1992) Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA 268:2420–2425

Hemilä H (2011) Zinc lozenges may shorten the duration of colds: a systematic review. Open Respir Med J 5:51–58

Hemilä H (2013) Vitamin C may alleviate exercise-induced bronchoconstriction: a meta-analysis. BMJ Open 3:e002416

Hrobjartsson A, Brorson S (2002) Interpreting results from randomized trials of complementary/alternative interventions: the role of trial quality and pre-trial beliefs. In: Callahan D (ed) The role of complementary & alternative medicine – accommodating pluralism. Georgetown University Press, Washington, DC, pp 107–121

Issit ML (2009) Hippies. Greenwood Press, Santa Barbara

Larson C (2007) Alternative medicine. Greenwood Publishing Company, Westport

Louhiala P (2010) There is no alternative medicine. Med Humanit 36:115–117

Louhiala P, Hemilä H (2014) Can CAM treatments be evidence-based? Focus Altern Complement Ther 19:84–89

Louhiala P, Puustinen R (2012) Alternative, complementary, integrative – conceptual problems in marketing healthcare ideologies and services. Focus Altern Complement Ther 17:156–159

Merriam Webster (2015) Definition of 'buzzword'. http://www.merriam-webster.com/dictionary/buzzword. Accessed 5 Mar 2015

Sackett DL, Rosenberg WMC, Gray JAM et al (1996) Evidence-based medicine: what it is and what it isn't. BMJ 312:71–72

Timmermans S, Mauck A (2005) The promises and pitfalls of evidence-based medicine. Health Aff 24:18–28

Tolstoy LN (2001) War and peace (transl. Louise and Aylmer Maude). Wordsworth Classics, London, London

Tonelli MR (1998) The philosophical limits of evidence-based medicine. Acad Med 73:1234–1240

Whitelegg M (1998) Alternative medicine. In: Chadwick R (ed) Encyclopedia of applied ethics, vol 1. Academic, San Diego

Wolpe PR (2002) Medical culture and CAM culture: science and ritual in the academic medical center. In: Callahan D (ed) The role of complementary & alternative medicine – accommodating pluralism. Georgetown University Press, Washington, DC, pp 163–171

Psychoanalysis as Science

58

Martin Hoffmann

Contents

Abstract

Psychoanalysis is one of the most prominent and most intensely discussed research programs of the twentieth century. One important debate in the philosophy of medicine centers around the question of whether or not psychoanalysis is a scientific research program. The paradigm case for the evaluation of this question is the theory of Sigmund Freud, who – in contrast to Carl G. Jung, Alfred Adler, and other proponents of psychoanalytic theory – regarded his theoretical efforts as a scientific project throughout his whole life. His project was continued by researchers in psychology and medicine, as well as

M. Hoffmann (✉)
Philosophisches Seminar, Universität Hamburg, Hamburg, Germany
e-mail: martin.hoffmann@uni-hamburg.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_41

practitioners in clinical psychotherapy and psychiatry. In order to give a more elaborate answer to the question of the extent to which this project is judged to be successful in contemporary science, it is necessary to differentiate between psychoanalytic theory, psychodynamic therapy, and the research methodology applied in the Freudian tradition.

Even if Freud himself took psychoanalysis to be a scientific, validated theory, his own research methodology faces serious problems. From the perspective of contemporary science, it constitutes the most "unscientific" aspect of his whole conception, because it is generally seen as falling victim to the *post hoc ergo propter hoc* fallacy. It is therefore deemed inappropriate for producing any substantial scientific evidence. But – contrary to Popper's prominent critique – it cannot be denied that many claims of psychoanalytic theory are empirically testable and that since the 1950s, a remarkable body of evidence that fulfills scientific research standards has been generated with the aim of confirming the central theoretical claims of psychoanalysis and the efficacy of psychoanalytic therapy. Therefore, in a processual or methodological sense, today's psychoanalysis is without any doubt a scientific research program. But at the same time, it is an open question whether the scientific endeavor to confirm the central claims of psychoanalysis will turn out to be successful. The generally accepted theorems that form the common core of today's psychoanalytic theorizing are – in sharp contrast to Freud's original theory – rather carefully formulated and are not particularly specific. For this reason, the relevance of psychoanalysis for the further development of psychology and medicine and the question of the efficacy and effectiveness of an autonomous psychodynamic therapy are matters of a deep and ongoing controversy.

Introduction

Psychoanalysis is one of the most prominent and intensely discussed research programs of the twentieth century. One important debate in the philosophy of medicine concerns the methodological status of psychoanalysis as a research program. The central question of this debate is the following: is psychoanalysis a scientific research program or does it fail scientific standards? Although there are different theories which are called "psychoanalytic" (not just Sigmund Freud's theory but also Carl G. Jung's theory of archetypes and the collective unconscious, Alfred Adler's individual psychology, Melanie Klein's object relations theory, etc.), the debate concerning the scientific status of psychoanalysis centers primarily around Freud's theory. One historical reason for this is that Freud was the only proponent of psychoanalysis who saw himself as a scientist throughout his whole life and who characterized his theory as a scientific, or at least proto-scientific, project. Furthermore, it was virtually only the Freudian tradition that gave rise to a research program aimed at validating the central claims of psychoanalysis on the basis of scientific evidence and with the help of experimental methods (Hilgard 1952a, b; Kline 1981; Fisher and Greenberg 1996; Chiesa 2010).

The conception of psychoanalysis as a science was challenged primarily by two kinds of criticisms. One line of argument was that Freud fell victim to a "scientistic self-misunderstanding" ("szientistisches Selbstmißverständnis," Habermas 1968; see especially pp. 300–332). Habermas argued that Freud's project is not a branch of the natural sciences but – rightly understood – rather turns out to be a hermeneutics of the self or of consciousness in general. Other adherents of a philosophical reinterpretation of Freud's works localized him within the methodological framework of modern phenomenology and (post-) structuralism (Ricœur 1965). This line of argument is no threat for psychoanalysis as a science if one allows for an "interpretative pluralism" and admits that it is possible to use Freud's theory as a starting point for both a distinct project in the field of the hermeneutical philosophy of consciousness and, at the same time, for a scientific project. Leaving exegetic questions aside, this seems to be an entirely plausible assumption that holds for many theoretical projects (e.g., ancient atomism, which was a theoretical source for both philosophy of nature and modern chemistry). By contrast, the other criticism is far more threatening for the project of psychoanalysis as a science. It is also the origin of the controversy about the scientific status of psychoanalysis. The proponents of this criticism accused Freud of being the founder of a pseudoscience along with astrology, homeopathy, or Marx's historical materialism. They argued that Freud's theory is not a scientific theory, because it is not empirically testable (Karl Popper), that his research methodology is deeply misconstrued (Adolf Grünbaum), and that psychodynamic therapy is at best completely ineffective and at worst dangerous for people suffering from a mental crisis (Hans-Jürgen Eysenck). Defenders of psychoanalysis react to these far-reaching criticisms with certain revisions of the theory or with refutations of the arguments.

This paper reconstructs the core issues, positions, and arguments of this controversy. It takes Freud's theory as a starting point and begins with some remarks about his reasons for classifying his theoretical conception of human mental life as a scientific theory (section "Some Central Claims of Freudian Psychoanalysis"). It then examines further developments of psychoanalytic theorizing (section "Is Freudian Psychoanalytic Theory a Scientific Theory?"), Freudian research methodology (section "Is Freud's Research Methodology a Scientific Methodology?"), and psychodynamic therapy with respect to their scientific status (section "Is Psychodynamic/Psychoanalytic Therapy Scientifically Validated?: A Reflection on Three Stages of Psychotherapy Research"). Due to the absence of a universally accepted definition of the terms "psychoanalytic" and "psychodynamic," both expressions are used interchangeably.

Some Central Claims of Freudian Psychoanalysis

Sigmund Freud was educated in the scientific tradition: he studied medicine, worked in the laboratory of Ernst Brücke on the histology of the nervous system during his studies, collected practical experience as a physician in the areas of

psychiatry and neurology, and acquired a lectureship in neuropathology in 1885. Ten years later he wrote a manuscript, later entitled *Project for a Scientific Psychology* (Entwurf einer Psychologie) by the editors, which opens with the words: "The intention of this project is to furnish us with a psychology which shall be a natural science: its aim, that is, is to represent psychical processes as quantitatively determined states of specifiable material particles and so to make them plain and void of contradictions ([Es ist die] Absicht, eine naturwissenschaftliche Psychologie zu liefern, d. h. psychische Vorgänge darzustellen als quantitativ bestimmte Zustände aufzeigbarer materieller Teile [und sie] damit anschaulich und widerspruchsfrei zu machen)" (Freud 1895, p. 387; the English translations of the quotes from Freud are taken from Strachey, 1966–1974). In this work, he tries to describe mental processes as shifts of quantums of energy within the nervous system. So in the years before 1900, he argued for a reductionist view of psychology as a field of natural science based on neurophysiological knowledge of the nervous system – which is a rather popular view in today's scientific psychology. In the following years he gave up this ambitious project, because he considered the neurophysiology of his time to be in a too rudimentary state of development in order to serve as a fruitful basis for his theoretical ideas. Nevertheless, during his entire lifetime he held the view that the psychoanalytic "hypothesis we have adopted of a psychical apparatus extended in space, expediently put together. . .has put us in a position to establish psychology on foundations similar to those of any other science, such, for instance, as physics ([u]nsere Annahme eines räumlich ausgedehnten, zweckmässig zusammengesetzten . . . psychischen Apparates . . . hat uns in den Stand gesetzt, die Psychologie auf einer ähnlichen Grundlage aufzurichten wie jede andere Naturwissenschaft, z. B. wie die Physik)" (Freud 1940, p. 126), as he wrote toward the end of his life in his work *An Outline of Psychoanalysis* (Abriß der Psychoanalyse).

According to Freud, the systematically conceptualized basic theory structure, the so-called metapsychology, is fundamental for the scientific character of psychoanalysis. The basic principles of his metapsychology are already outlined in his most famous book *The Interpretation of Dreams* (Die Traumdeutung 1900). The core of the theory consists (i) in a topography of the mental apparatus (first explicated as three mental subsystems of the Conscious, the Preconscious, and the Unconscious, subsequently superseded by the second topographic model of Id, Ego, and Superego); (ii) the dynamics of the mental apparatus, consisting of the unobservable mental forces that are causing human behavior (of special importance are the defense mechanisms such as repression, sublimation, and resistance); and (iii) the economic dimension of the mental system, explaining repression and other mental processes as shifts and exchanges of energy quantums between the different subsystems of the mental apparatus, directed from a higher level of "bound" energy to lower energy levels (see for a more detailed description of the general structure of Freud's metapsychology: Kitcher 1992, pp. 39–56). Freud defended the scientific status of the theory primarily with reference to its enormous explanatory power: psychologists, who merely theorize about conscious mental

phenomena, can only provide fragmentary, scattered, and poor explanations of the complexity and diversity of human behavior. By contrast, it is the psychoanalytic assumption of the Unconscious that allows the causes of human actions, motives, and feelings to be explained in a comprehensive and unifying way (Freud 1940, pp. 80–81).

On the basis of this metapsychology, Freud developed more specific theory elements: a theory of personality and psychosexual development, a theory of psychopathology, and a method of psychotherapy, the psychoanalytic, long-term "talking therapy" with a duration of 300, 400, or more treatment sessions. He made an explicit statement about the cornerstones of psychoanalytic theory ("die Grundpfeiler der psychoanalytischen Theorie") in a paper published in 1923: "The assumption of unconscious psychical processes, the acknowledgement of the theory of resistance and repression, the assessment of sexuality and the Oedipus complex are the chief contents of psychoanalysis and the foundations of its theory, and anyone who does not accept them all should not be considered as a psychoanalyst (Die Annahme unbewußter seelischer Vorgänge, die Anerkennung der Lehre vom Widerstand und der Verdrängung, die Einschätzung der Sexualität und des Ödipus-Komplexes sind die Hauptinhalte der Psychoanalyse und die Grundlagen ihrer Theorie, und wer sie nicht alle gutzuheißen vermag, sollte sich nicht zu den Psychoanalytikern zählen)" (Freud 1923, p. 223). These different theory elements – this was one of his central ideas expressed in the quote – are not isolated from another but are deeply interdependent: the basic principles of metapsychology, the more specific theories, and the ideas about effective psychotherapy (compare section "Is Freud's Research Methodology a Scientific Methodology?" below). According to Freud, these elements have to be seen as a holistic framework for human mental life and mental disorder (see for an introduction to psychoanalysis Brenner 1973 and for a detailed account of the whole theory and its reception Köhler 2000).

Patricia Kitcher (1992) deserves credit for having worked out a detailed reconstruction of the embedding of Freud's theory in the research context of the psychiatry, neurology, and neurophysiology of his time. Kitcher convincingly argues that in the light of his historical background, Freud can be seen as the founder of an innovative "complete interdisciplinary science of mind" and his theory as a methodologically subtle and creative reaction to the groundbreaking developments of nineteenth-century neurology, psychiatry, and psychology. But even if this historical thesis is true and if we admit that Freud's theory was a proper part of science in his time, it may nevertheless be the case that psychoanalysis shares the fate of alchemy and astrology, which were branches of science until the sixteenth century, but subsequently became decoupled from the path of scientific progress and are now considered as pseudosciences by most scientists (Newman and Grafton 2001). The next section will investigate the systematic question of whether the further developments of psychoanalytic theory during the twentieth century justify ascribing to it the status of a scientific project in the context of current scientific research.

Is Freudian Psychoanalytic Theory a Scientific Theory?

Karl Popper's Argument Against the Scientific Status of Psychoanalysis

Karl R. Popper, one of the central figures of the philosophy of science in the twentieth century, formulated a far-reaching argument against the possibility of regarding psychoanalysis as a science. His main point was that psychoanalytic theory does not satisfy the demarcation criterion for science. In his book *Logic of Scientific Discovery*, first published as *Logik der Forschung* in German in 1935, he proposed the falsifiability of empirical theories as the decisive demarcation criterion for drawing a line between science and nonscience (Popper 1935). In contrast to the members of the Vienna Circle (e.g., Rudolf Carnap, Moritz Schlick, and Otto Neurath), who developed verificationism as a semantics and methodology for scientific theories, Popper argued that empirical theories are in fact not verifiable, because the method of induction (which is, according to the members of the Vienna Circle, an indispensable inferential tool for the confirmation of empirical theories) faces serious epistemological problems. As an alternative, Popper developed his falsificationism, which, he claims, is exclusively based on deductive inference. According to this view, empirical theories have to be falsifiable, which means it must be possible that the predictions of the theory conflict with observational data. "Every 'good' scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids the better it is" (Popper 1963, p. 36). Popper notes that already in 1919, when he became acquainted with Alfred Adler, he began to think about the question of what might be wrong with Marx's theory of history, Adler's individual psychology and Freud's psychoanalysis. He found that the problem of all of these theories is that they do not "forbid" anything to happen, i.e., that *every* course of events is compatible with and can be explained by these theories. This explanatory potential makes these theories attractive and suggestive and may also explain their great popularity. But due to their lack of falsifiability, their explanatory success is merely an illusion, because the theories cannot be tested against reality. "A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice" (ibid., p. 36). Therefore, according to Popper, psychoanalysis is not a branch of science but a form of psychological metaphysics.

Popper's argument was widely discussed and, in the end, turned out to be unsuccessful, because it faces two serious problems. The first problem lies in Popper's conception of falsifiability itself, which, in contemporary philosophy of science, is almost universally considered as inadequate for demarcating the line between science and nonscience. Popper conceptualizes falsifiability as a two-place relation with one theory in one place and observational evidence in the other. But, as Imre Lakatos convincingly showed, in order to determine the scientific status of a theory, we also have to take into account that scientific theories do not exist in isolation but partake in a scientific discourse along with competing theories and are diachronically embedded in a process of theoretical changes and reformulations. Lakatos developed a more

sophisticated and adequate picture of the falsifiability of theories as a three-place relation between the observational evidence and two (or more) rival theories. Accordingly, falsifiability cannot be ascribed to single theories (as Popper claims for Freud's and Adler's theory) but has to take into account the embedding of a theory in a series of developing theories. Popper fails to take into account that these considerations are of crucial importance for evaluating the scientific status of a research program (Lakatos 1978). Furthermore, Popper's conception is a "single focus" approach to demarcation: he allows one and only one criterion for deciding the question of the theory's scientific status. By contrast, in contemporary philosophy of science, most people believe that the complex question of demarcating science and nonscience can only be answered (if at all) by a multi-criteria approach (Ruse 1982).

But Popper's argument fails for a second, even more serious reason. Popper does not present any case studies or any detailed reconstructions of Freud's theory. Other philosophers of science did so and found that psychoanalysis, e.g., Freud's theory of personality, his etiology of adult obsessional neurosis, and his theory of dreams, does in fact include falsifiable statements – which was already recognized by Freud himself (Grünbaum 1979). Furthermore, even if Popper were partially right and it would turn out that some of Freud's theories are not empirically testable in their existing formulation, it remains possible that they could be reformulated in a more precise way that makes them empirically testable.

Reactions to Popper: Establishing Psychoanalysis as a Scientific Project

During the 1940s and 1950s, several psychologists began working on the project of turning psychoanalysis into a scientific research program by looking for empirical evidence supporting it and by conducting experimental tests of psychoanalytic principles. The first person who coined the expression "psychoanalysis as science" was the Stanford psychologist Ernest R. Hilgard (1904–2001), who published a paper and a book with this title in 1952. His main idea was to collect and evaluate all of the experimental evidence available for psychoanalytic theory and psychoanalytic therapy at that time. His initial conclusions concerning the empirical validation of psychoanalysis (although they were refuted later on; see below) were quite euphoric: "[I]t has been possible to parallel many psychoanalytic phenomena in the laboratory. When this is done, the correspondence between predictions according to psychoanalytic theory and what is found is on the whole very satisfactory" (Hilgard 1952b, p. 42). Just a few years later, Ellis (1956) developed operational definitions of central terms of psychoanalytic theory (such as Id, Ego, Superego, phallic phase, libido, Oedipus complex, etc.) in order to enable a reformulation of the psychoanalytic principles in a way that makes transparent how they can be tied to an observational basis and which observable data confirm and which repudiate their existence. In the following years, the empirical methods became tremendously refined and improved, and a number of monographs were published that presented and collected empirical studies and

conducted meta-analyses in order to test and validate the basic principles of psychoanalysis in a scientific way (Fisher and Greenberg 1977, 1996; Kline 1981). This development culminated in a book series edited by J. M. Masling, systematically collecting the *empirical studies of psychoanalytical theories* (first volume Masling 1983). So now there is in fact a remarkable body of observational and experimental data generated with the aim of proving the truth of the central claims of psychoanalytic theory.

Contemporary Developments

Nevertheless, it would be too hasty to consider psychoanalysis as a generally accepted and well-established field of scientific psychology today. At present, the issue whether psychoanalysis is satisfactorily confirmed with respect to its core concepts and principles or whether it is proven wrong in the end remains unsettled and is still the subject of highly controversial debates. This can be shown, for example, with reference to the controversial assessment of one of Freud's core ideas: in his introduction to a book about the empirical investigation of the neuronal bases of unconscious mental phenomena, James Uleman concludes that indeed the "psychoanalytic unconscious is, to most laypeople and those in the arts and humanities, the only unconscious," but "it does not provide an influential framework for understanding unconscious processes in academic or scientific circles" (Uleman 2005, pp. 4–5). On the other hand, there are approaches for integrating results from psychoanalytic theorizing about unconscious mental phenomena into the context of current scientific research in the neurosciences (Mancia 2006).

These and other highly controversial assessments of the scientific merits of psychoanalytic theory in contemporary discussions in scientific psychology and medicine primarily have two sources. The first is the complex shape and inhomogeneity of the available empirical evidence. At present, certain assumptions of psychoanalytic theory are confirmed by empirical evidence, whereas others are either not sufficiently supported yet or are regarded as refuted – even by contemporary psychoanalysts themselves. The latter holds not only for negligible assumptions but also for some of Freud's most prominent claims: the existence of the Oedipus complex, traditionally seen as one of the core assumptions of his theory of the etiology of neuroses, is only confirmed by rather poor evidence (Kupfersmid 1995). The existence of the death drive, introduced as an antagonistic principle to the libido's "life drive," is currently considered to be clearly refuted in the light of modern evolutionary theory. However, defenders of Freud point out that Freud himself was very uncertain with respect to this element of his theory (introduced by him not until 1917 in rather tentative formulations) and insist that, although the idea of the death drive is wrong, "a number of lessons can be drawn" from it (Black 2011, p. 118). The empirical validation of the existence of repression and resistance, both generally regarded as centerpieces of psychoanalytic theory, is a matter of deep controversy (see the extensive discussion of an article by Erdelyi (2006) in the journal *Behavioral and Brain Sciences*). And finally the ideas of penis envy and

the castration complex as well as the negligence of female psychosexual development are interpreted as a massive gender bias of Freud's theory (Gyler 2010).

At the same time, there are other psychoanalytic claims which are confirmed by empirical evidence and even by systematic experimentation. Westen (1998) has formulated five principles that he considers to be the core assumptions of current psychodynamic theory:

1. "[M]uch of mental life – including thoughts, feelings, and motives – is unconscious."
2. "[M]ental processes, including affective and motivational processes, operate in parallel so that, toward the same person or situation, individuals can have conflicting feelings that motivate them in opposing ways and often lead to compromise solutions."
3. "[S]table personality patterns begin to form in childhood, and childhood experiences play an important role in personality development."
4. "[M]ental representations of the self, others, and relationships guide people's interactions with others and influence the ways they become psychologically symptomatic."
5. "[P]ersonality development involves not only learning to regulate sexual and aggressive feelings but also moving from an immature, socially dependent state to a mature, interdependent one." (Westen 1998, pp. 334–335)

Westen reviews the evidence in favor of these principles and rates all of them as empirically confirmed to a satisfactory degree. He concludes: "Freud advanced several fundamental propositions, once highly controversial and unique to psychoanalysis, that have stood the test of time . . . This is probably the best any thinker could hope for in a rapidly developing discipline like ours 60 years after his death" (Westen 1998, p. 362). Of course one should agree with Westen that it would be illegitimate to identify contemporary psychoanalytic theory with Freud's theory and to regard the former as refuted if central claims of the latter are shown to be wrong. But even if it is taken for granted that all of the empirical evidence that Westen refers to is of high methodological quality and therefore entirely convincing, it remains a matter of controversy whether his five principles do in fact capture the essential claims of contemporary psychodynamic theory and if they are specific to it. A closer look at the principles shows that it would be very difficult to find anyone working in contemporary psychology and psychological medicine who questions the truth of principles (4) and (5). Moreover, the other three principles do not seem to be specific to proponents of psychodynamic theory. This holds especially for principle (1), because there are several different conceptions of the Unconscious – as much in current psychology as in the history of the sciences and humanities (see for more details Uleman 2005). In sum, Westen's principles seem to be rather cautiously formulated, and in part they consist in generally accepted psychological assumptions. For this reason, critics of Westen's approach might conclude that it is not too surprising that he is able to offer an attractive number of conclusive empirical evidence for their confirmation.

This discussion leads to the second source of the ongoing controversy regarding the scientific status of psychoanalytic theory. This controversy is not merely a matter of evaluating the quality of empirical evidence alone. Rather, it cannot be solved without answering another crucial question: what is the content specific to current psychoanalytic theory? Which set of assumptions does a proponent of this theory have to accept and which of these assumptions are only accepted by the proponents of the theory? This question cannot be decided on the basis of the available empirical evidence but is related to considerations about the essential theoretical content of the claims of psychoanalytic theorizing. Therefore, this is a highly controversial question even (and especially) between the proponents of psychodynamic theory. What many defenders of psychoanalysis say in favor of their position is that it fell victim to its own success in the sense that some of its claims, historically originating from Freud's theory and empirically well confirmed today, constitute common psychological and medical knowledge, which is accepted by nearly everyone. This might be true. But still the theoretical question remains whether these claims are strong enough to denote a theory core that is specific to psychoanalytic theory (as Westen and others seem to suggest). Only when this question is answered can the controversy about the scientific credibility of psychoanalytic theory be solved.

Is Freud's Research Methodology a Scientific Methodology?

While the scientific status of the content of Freudian theory is currently a matter of controversy, it is widely accepted that the research methodology Freud has introduced as the *via regia* for the empirical validation of psychoanalysis is, from a scientific point of view, the most problematic aspect of psychoanalytic thinking.

Freud himself only used interpretations of individual cases for the empirical confirmation of his theory. In current scientific methodology, this database, especially if used as the only empirical foundation, is generally considered to be poor evidence, because the selection of individual cases is a rather arbitrary process, and the great diversity of phenomena of human behavior and mental life allows for the confirmation of almost any hypothesis by only a small number of cases. Therefore, single case studies are seen as an appropriate heuristic method in theory development and in generating innovative hypotheses, but not as a source of providing evidence for rigorous theory checking.

Freud's way of selecting and interpreting his case studies is also prone to many distortions and biases. Most of the empirical data, cited in his *The Interpretation of Dreams* (1900) with the intention to confirm the basic principles of his metapsychology, are in fact interpretations of the dreams that he himself had during his self-analysis between 1897 and 1899. The other important sources of evidence – especially for the validation of his theory of psychopathology – are detailed analyses of individual patients. Wolpe and Rachman (1960) conducted a reanalysis of his perhaps most famous case study, the first psychoanalysis of a child (published

by Freud in 1909 and entitled *Analysis of a Phobia in a five-year-old boy* (Analyse der Phobie eines fünfjährigen Knaben 1909)). Wolpe and Rachman's central criticism was that the study design violates fundamental standards of scientific objectivity: Freud saw the child only once during the treatment, and moreover, the therapy was conducted by the boy's father, whom Freud himself calls one of his "closest adherents." The emotional relation between the father and son, the partiality of the father with respect to Freud's theory, and the selection effects caused by the communication between the boy's father and Freud are all sources of systematic biases. The most important consequence is that a considerable proportion of the results must therefore be considered as a mere effect of suggestion or indoctrination during the therapy. Without any doubt, a patient in a mental crisis who expects help from the therapist (and in particular a 5-year-old boy in his relationship with his father) is predisposed to be influenced by the suggestions that lead him to accept the "truths" of psychoanalysis during the therapy.

Seven years after the publication of the *Analysis of a Phobia in a five-year-old boy*, in his *Introductory lectures on psychoanalysis* (Vorlesungen zur Einführung in die Psychoanalyse 1916/1917), Freud himself accepted that the problem of suggestion and indoctrination is the most important objection to his research method and he developed a counterargument to refute it. The decisive evidence for the truth of psychoanalytic theory consists, according to Freud, in the unique success of psychoanalytic therapy. This is now recognized, in contrast to, say, hypnosis, which Freud abandoned as a therapeutic method, because he considered it liable to suggestion. Consequently, Freud concluded that only psychoanalytic therapy yields a durable cure. His main argument to establish this conclusion is the so-called tally argument, which he presented in the last lecture of the *Introductory lectures* entitled "The analytic therapy" (Die analytische Therapie). It was reconstructed by Adolf Grünbaum (1984, pp. 135–141). This argument is based on two crucial premises:

1. Only psychoanalytic therapy provides the therapeutic option to not merely remove or shift the symptoms (as with other therapeutic procedures such as hypnosis) but to reveal the hidden (unconscious) causes of the patient's neurosis – even if these causes lie deep in the past of the patient's life.
2. Only this process of disclosure of the true causes of the mental problems to the patients can yield a durable cure from their neuroses (and not merely temporary improvements caused by shifts of certain symptoms and reactions).

From these premises Freud deduced the tally argument's main conclusion: every successful psychoanalytic therapy provides striking evidence for psychoanalytic theory, because the truth of psychoanalytic theory is the only explanation for the exclusive success of psychoanalytic therapy. This conclusion implies that a successful psychoanalytic therapy cannot be contaminated by suggestion or indoctrination. For in that case the therapy would merely remove the symptoms for a time and fail to reveal the true causes of the neurosis. But if the true causes of the neurosis remain unrevealed, no durable cure is possible.

Grünbaum criticizes this argument at length. He argues – against Popper – that his reconstruction of the argument shows the empirical testability of Freud's theory. In fact, there are several assumptions derivable from the tally argument's premises that are empirically testable, namely, (i) the only way to achieve a durable cure of a mental disorder is to reveal its true causes and (ii) psychoanalytic therapy is the only therapeutic method that can reveal a mental disorder's true causes. From (i) and (ii) follows (iii), psychoanalytic therapy alone provides a durable cure, which implies (iv), the occurrence of a spontaneous remission is empirically impossible (compare Freud 1909, p. 339), etc. Grünbaum's main point is that many of these assumptions are either not validated or are simply refuted by the available empirical data (Grünbaum 1984, pp. 141–176).

Even though Grünbaum's reconstruction and critique of the tally argument was criticized concerning certain exegetic respects (Esterton 1996), it is widely agreed that his main point is correct: Freud made the crucial mistake of an inadequate conflation of the empirical validation of causal claims of psychoanalytic theory with the empirical evaluation of the efficacy of psychoanalytic therapy (Greenwood 1996). Even the scientifically orientated psychoanalysts mostly admit that this methodological decision of Freud's is a pitfall for the scientific validation of psychoanalytic theory. The tally argument is usually interpreted as an instance of the *post hoc ergo propter hoc* fallacy, the mistake to derive a causal dependence from a temporal succession of events. This reasoning has certain established applications in medical practice – primarily the so-called *diagnosis ex juvantibus* (diagnosis on the basis of successful treatment). But even this special application is controversial and only admissible under restricted conditions: when the consequence is suddenly perceived after the preceding event and no alternative explanations for its occurrence are available (e.g., in the case of providing treacle in an acute hypoglycemia of a diabetic). None of these conditions are fulfilled in psychoanalysis. For this reason one has to conclude that Freud's research methodology fails to provide any conclusive scientific evidence for either psychoanalytic theory or psychodynamic therapy.

Is Psychodynamic/Psychoanalytic Therapy Scientifically Validated?: A Reflection on Three Stages of Psychotherapy Research

One lesson of the last section is that the areas of psychodynamic theory and psychodynamic therapy are considerably more independent from each other than Freud himself thought. This can be seen as good news for the project of the scientific validation of the methods of psychoanalytic therapy. The reason is that even if it turns out to be the case that the central claims of psychoanalytic theory have to be abandoned, psychoanalytic therapy might still be an effective method for the treatment of mental disorders. So the question about empirical evidence for the efficacy and effectiveness of psychoanalytic therapy arises.

First Stage: Clinical Studies and First Meta-Analyses

The progress of empirical research that has been carried out in order to confirm the efficacy and effectiveness of psychoanalytic therapy can be structured in three chronological stages. The first stage, beginning around the year 1950, is characterized by the first comparative experimental testing of different types of psychotherapy and by the attempt to integrate the results of these quite divergent clinical studies into several meta-analyses. In this early stage of psychotherapy research, most of the meta-analyses resulted in one of the following two results. A prominent example for the first result is the research of the psychologist Hans-Jürgen Eysenck, an influential theoretician of intelligence factor theory and defender of behavioral therapy. He conducted an oft-quoted meta-analysis of 24 effectiveness studies of psychotherapy and concluded that the recovery rate of neurotic patients after undergoing a psychoanalytic therapy is not higher than the rate of spontaneous remissions – in his own, somewhat polemic words: "[W]hen we discount the risk the patient runs of stopping treatment altogether, his chances of improvement under psychoanalysis are … slightly worse than his chances under a general practitioner or custodial treatment" (Eysenck 1952, p. 322). The second result, which is not necessarily contradicting Eysenck's verdict and can be found in many meta-analyses of that time, confirms the so-called dodo bird conjecture, named after the dodo bird in Lewis Caroll's *Alice in Wonderland* and its aphorism: "Everybody has won, and all must have prizes." The conjecture says that all types of psychotherapy (psychoanalytic therapy, behavioral therapy, and eclectic approaches) in the end show more or less equivalent outcomes – and if one type of therapy is shown to be superior in a given study, the result usually conforms with the preferences of the investigators (Luborsky et al. 1975). Sometimes this result is interpreted as a methodological artifact: most studies of that time did not reliably distinguish between different mental disorders. It could be that every type of therapy is effective only for some disorders and that the averaging evaluation of therapeutic success over all disorders merely levels out these differences. As a consequence, some psychoanalytically orientated psychotherapists recommended behavioral therapy for minor mental problems and psychodynamic therapy for the treatment of severe mental disorders (Pongratz 1973, p. 378). But there was no empirical evidence for this disorder-specific indication schema (and the current evidence seems to refute it, as shown below). From the present perspective, many of the clinical studies in that stage of research have to be criticized for their methodological deficiencies (subjective or obsolete diagnoses of the investigated mental disorders, unreliable measures of therapeutic success, failures in the statistical evaluations, selection biases in the meta-analyses), which undermine the credibility of the results.

Second Stage: Large-Scale Meta-Analyses

The second stage of psychotherapy research is characterized by the effort to overcome these methodological shortcomings with the help of more sophisticated

statistical methods and larger samples of investigated subjects. During the 1980s, Grawe et al. (1994) began to plan and undertake one of these large-scale meta-analyses, which indicated a substantial advance in psychotherapy research. First, they conducted a careful survey of the entire available research literature including all clinical studies ever carried out for the evaluation of psychotherapy – from the beginning of psychotherapy research until 1983/1984. Initially, they found more than 3500 studies. After a criteria-based selection process, 897 of these studies were found to fulfill satisfactory methodological standards. (This means that Grawe et al. included nearly twice as many clinical studies as Smith et al. (1980), a far more influential meta-analysis in the English-speaking literature that includes 475 studies.) These 897 studies served as the data basis for their systematic comparative meta-analysis of more than 40 therapeutic techniques, sorted into three broad therapy types: humanistic therapies, cognitive-behavioral therapies, and psychodynamic therapies. In the area of psychodynamic therapy, they distinguished between nine different therapeutic methods, including classic long-term psychoanalysis, psychoanalytic short-term therapy, Adler's individual therapy, and Binswanger's "Daseinsanalyse." The scientifically best-evaluated methods were the psychoanalytic short-term therapy (29 studies) and psychodynamic therapy combined with medical treatment (13 studies). For the remaining 7 psychodynamic therapies, Grawe et al. found that only 28 studies fitted their criteria. So overall, until 1983, there were merely 70 studies that assessed the efficacy of psychodynamically orientated psychotherapies. By comparison, at the same time there were 452 studies that evaluated the efficacy of the different methods of cognitive-behavioral therapy. Another indicator for the relatively small effort to prove the efficacy of psychodynamic therapy is the fact that Grawe et al. did not find a single study that fulfilled their selection criteria and evaluated classic long-term psychoanalysis, favored by Freud himself. The only systematic and controlled study to evaluate long-term psychoanalysis is the famous and oft-quoted study of the Menninger foundation, which was initiated in 1954 and lasted for more than 20 years. The study was conducted by some of the most prominent psychoanalysts of that time (Otto Kernberg, Robert Wallerstein, Merton Gill, and others) and included 42 patients, all of them suffering from severe neuroses. One reason for the long duration of the study was the average duration of psychoanalytic treatment (of the 15 patients who finished the therapy) of almost 6 years; during this time each patient received 1017 treatments on average. It is a remarkable result that even in this extremely extensive study, undertaken by renowned psychoanalysts, it was in the end not possible to show that the long-term success of psychoanalysis is superior to an alternative psychotherapy with only one third of the treatment sessions (Wallerstein 1986, p. 515).

Grawe et al. (1994) did not include the Menninger study in their meta-analysis due to its methodological shortcomings, but they also conducted a direct comparison between the efficacy of the psychodynamic therapy type on the one hand and the two types of cognitive-behavioral therapy and humanistic therapy on the other hand. They selected the comparative studies and found that, in general, cognitive-

behavioral therapy is significantly more effective than both psychodynamic therapies and humanistic therapies. A statistical effect size comparison of the 22 studies (with a total of 487 patients), which included a direct comparison, showed an averaged effect size of 0.83 for psychoanalytic psychotherapy and an averaged effect size of 1.23 for cognitive-behavioral therapy. Significance testing of this difference with the *t*-test for dependent samples showed that the difference is highly significant ($p < 0.0001$). Grawe et al. (1994, pp. 651–671) interpreted this result as strong evidence for both (i) the efficacy of psychodynamic therapy and also (ii) for the superiority of cognitive-behavioral therapy over the different methods of psychodynamic therapy.

Of course, Grawe and his colleagues' results provoked much criticism, especially from defenders of psychoanalytic therapy. Tschuschke et al. (1998) conducted a reanalysis of the 22 comparative studies from Grawe's meta-analysis. They undertook a systematized rating process by 12 independent psychotherapy researchers in order to evaluate the methodological quality of the studies. This expert rating showed the result that "only 5 or 8 of the 22 studies, respectively, could be accepted for a relatively fair comparison between the treatments under study" (Tschuschke et al. 1998, p. 430). They found all other studies to be either methodologically deficient or systematically biased. Surely, expert ratings have their own problems concerning the impartiality of and the criteria for the selection of the experts. But one systematic problem of many meta-analyses cannot be denied – regardless of how comprehensive their data base may be: the therapeutic interventions that are investigated in the multitude of the included studies (even if they are all summed up under the label of "psychodynamic therapy" or "psychoanalytic therapy") diverge considerably with respect to the dosage and realization of the treatment, the competence and practical experience of the therapist, and the duration of the therapy.

Third Stage: Comparative Psychotherapy Process-Outcome Research

In order to solve this methodological problem, which undoubtedly undermines the interpretability of the results, the third and current stage of psychotherapy research emerged, the so-called comparative psychotherapy process-outcome research. The aim of this branch of research is to empirically examine what exactly happens in the psychotherapeutic process, what the essential features of a certain method of psychotherapy are, and in which respect the properties of different methods and interventional practices diverge. Blagys and Hilsenroth (2000) conducted a study in order to isolate features that distinguish between cognitive-behavioral therapy on the one hand and psychodynamic-interpersonal therapy on the other. They did not only evaluate the theoretical literature on therapy but also generated a database in order to reveal information about the empirically perceived therapeutic processes that characterize the interventions usually labeled as psychoanalytic or psychodynamic therapy. They found seven features that reliably characterize the

empirical practice of psychodynamic therapy in contrast to the methods of cognitive-behavioral therapy:

1. A "focus on affect and the expression of patients' emotions"
2. An "exploration of patients' attempts to avoid topics or to engage in activities that hinder the progress of therapy"
3. The "identification of patterns in patients' actions, thoughts, feelings, experiences, and relationships"
4. An "emphasis on past experiences"
5. A "focus on a patients' interpersonal experiences"
6. An "emphasis on the therapeutic relationship"
7. An "exploration of patients' wishes, dreams, or fantasies" (Blagys and Hilsenroth 2000, pp. 169–182)

On the basis of these criteria, it might become possible to define the core elements of psychodynamic treatment and to make clear comparisons between different therapy methods in order to isolate the most effective techniques. "In addition, future research on the relationship between process and outcome can aid in the determination of when and with whom the use of these techniques will be most effective" (Blagys and Hilsenroth 2000, p. 185). This project seems very promising, but it is in an early stage of its development. Presently there are no definite results concerning the efficacy of psychodynamic therapy on the basis of empirically validated process-outcome criteria that would be required for the project.

To sum up, the area of psychotherapy features a research situation that is similar to the stage of the empirical validation of the principles of psychodynamic theory (compare section "Contemporary Developments"). Again, one could question whether claims like Blagys and Hilsenroth's (2000) are strong enough to define a core of methods that can serve as the basis of an autonomous therapy method. Whereas some researchers work on the further development and validation of a specific psychodynamic psychotherapy (Shedler 2010), others regard this project as "confessional" and instead favor the strategy of integrating the most successful interventions from different therapy methods into a unified "professional" psychological psychotherapy (Grawe 1998). But there is no agreement on this matter. There is a great variety of diverging definitions and approaches in today's research on the efficacy of psychotherapy in general and psychodynamic therapy in particular (Levy and Ablon 2009).

Conclusion

This chapter has addressed the question of whether psychoanalysis is a science. Even if Freud himself thought of psychoanalysis as a scientific project, his own methodological conception of the validation of his theory faces serious problems, and given today's scientific standards, it probably has to be considered as the most

"unscientific" aspect of his whole conception. His idea to construe the research methodology of psychoanalysis as deeply intertwined with its therapeutic methodology and his claim that therapeutic success is the most important validation for psychoanalytic theory are instances of the *post hoc ergo propter hoc* fallacy and therefore inappropriate for producing any substantial scientific evidence for psychoanalysis.

However, contrary to Popper's critique, it cannot be denied that many claims of the Freudian theory are empirically testable and that since the 1950s, a remarkable body of evidence that fulfills scientific research standards has been generated with the aim of proving the truth of psychoanalytic theory and of evaluating the efficacy of psychoanalytic therapy.

Nevertheless, in contemporary scientific medicine and psychology, it is highly controversial whether – and if so, to which degree – the attempt to confirm the central claims of psychoanalysis with scientific research methods will turn out to be successful. Again, Lakatos' terminology is helpful in order to adequately describe the state of the current discussions of the question about the scientific status of psychoanalysis. In his theory of research programs, Lakatos differentiates between the "hard core" of a research program, which is formed by the axioms, basic principles, and central theorems of the theory and its "protective belt," consisting of more specialized theory elements, paradigmatic heuristics and methods of experimental and observational research, ad hoc hypotheses, etc. (Lakatos 1978, pp. 47–90). Applying this terminology to psychoanalysis, its development during the twentieth century can be described as follows: in Freud's times, psychoanalysis was characterized by an ambitious "hard core" (complex and far-reaching theoretical principles formulated in Freud's extensive writings), but it lacked any substantial scientific validation. The observational and experimental research that has been carried out since the 1950s equipped psychoanalysis with a remarkable "protective belt" and turned it into an influential and well-known research paradigm in psychology, psychiatry, and clinical medicine. In this processual or methodological sense, today's psychoanalysis is a scientific research program. But at the same time, this process led to a significant thinning of the "hard core" of both the content of psychoanalytic theory and the methodology of psychodynamic therapy. The generally accepted theorems that form the common core of psychoanalytic theorizing today are rather cautiously formulated and are not particularly specific. For this reason, the progressiveness of this research program, its relevance for the further development of current psychology, and the philosophy of consciousness as well as the question of the efficacy and effectiveness of an autonomous psychoanalytic therapy remain highly controversial.

Definition of Key Terms

Unconscious — A core concept of Freud's theory, introduced as an element of Freud's first topographic model of the mental apparatus, structuring the mind into three parts: the Conscious, the

	Preconscious, and the Unconscious. Freud was convinced that every instance of human behavior, motive, or feeling must have a mental cause. He regarded the Unconscious as the source of all of the "hidden" causes that have to be assumed as the basis of a comprehensive and unified explanation of any phenomena of human mental life.
Repression	A core concept of Freud's theory, introduced in order to describe the dynamics of human mental life. Mental content that is felt to be too awkward, displeasing, or painful to cope with is repressed in the Unconscious. These mental contents cause various mental phenomena (e.g., dreams or neurotic symptoms) that represent the repressed content in a deformed way to the Conscious.
Significance level	Statistical measure to specify the probability that a certain property, effect, or group difference measured in the study sample also exists in the overall population. A significance level of 5 % ($p = 0.05$) indicates that the investigated condition measured in the sample is also present in the overall population with a probability of 95 %. In other words, a probability of 5 % indicates that the study results do not represent a condition of the population but are merely due to a sampling error.
Effect size	Statistical measure to quantify the size or magnitude of a measured effect. This statistical measure is particularly relevant in psychotherapy outcome research, because the focus here is not only to show that the investigated treatment has an effect but also to show the magnitude of the effects. Significance levels are not helpful in this respect, because they do not contain any direct information about the magnitude of the measured effects or conditions. A metric that is often used for determining effect sizes is normalized with reference to standard deviations. So if an effect size of 1 is reported in order to quantify the success of a therapy, this means that the comparison between the average health status of the patients before and after the therapy showed a gain of one standard deviation.
Randomized controlled trial (RCT)	Study type which is currently regarded as the methodological "gold standard" in (clinical) psychology and medicine. In this field, RCTs are primarily used to conduct fair checks of the effectiveness and efficacy of innovative treatments. RCTs contain at least two subsamples, a treatment group, and one or more control groups. The treatment group receives the treatment under investigation, and the control group(s) receives either an alternative treatment or a placebo. The assignment of the participants to the different

	groups is carried out randomly as a statistical means for controlling the influence of distorting effects that are unknown to the researchers.
Meta-analysis	Complex statistical procedure for integrating the results of a multitude of single studies. The aim is to strengthen the validity of the results by considering as much information as possible, avoiding the effects of one-sidedness and balancing the methodological limitations of individual studies. The main problem of meta-analyses is the diversity of the included studies, which is a challenge for the applied statistical methods and may affect the interpretability of the results.
Post hoc ergo propter hoc fallacy	The fallacy to derive conclusions about causal dependencies from the mere temporal succession of events.

Summary Points

- Although there are different theories which are called "psychoanalytic" (not just Sigmund Freud's theory but also Carl G. Jung's theory of archetypes and the collective unconscious, Alfred Adler's individual psychology, Melanie Klein's object relations theory, etc.), the debate concerning the scientific status of psychoanalysis centers primarily around psychoanalytic theorizing in the Freudian tradition.
- The controversy about the scientific status of Freudian theory originated primarily from the fundamental criticism that psychoanalysis is a pseudoscience, along with astrology, homeopathy, or Marx's historical materialism.
- In order to provide an elaborate answer to the question to which extent the project of validating psychoanalysis with scientific methods is judged to be successful in contemporary science, it is necessary to differentiate between psychoanalytic theory, psychodynamic therapy, and the research methodology applied in the Freudian tradition.
- It is widely accepted that the research methodology Freud has introduced as the *via regia* for the empirical validation of psychoanalysis is, from a scientific point of view, the most problematic aspect of psychoanalytic thinking. It is generally seen as an instance of the *post hoc ergo propter hoc* fallacy and therefore as inappropriate for producing any substantial scientific evidence.
- However, since the 1950s, a remarkable body of evidence that fulfills scientific research standards has been generated with the aim of proving the central theoretical claims of psychoanalysis and the efficacy of psychoanalytic/psychodynamic therapy.
- The scientifically validated theorems that form the common core of today's psychoanalytic theory are – in sharp contrast to Freud's original theory – rather carefully formulated. It is generally seen as an open question whether these claims are strong enough to denote a theory core that is specific to psychoanalytic theory.

- The area of psychotherapy features a research situation that is similar to the stage of the empirical validation of psychoanalytic theory. Again, one could question whether the essential claims of current psychodynamic therapy are strong enough to define a core of methods that can serve as the basis of an autonomous therapy method.
- The progressiveness of psychoanalysis as a scientific research program, its relevance for the further development of current psychology and medicine, and the question of the efficacy and effectiveness of an autonomous psychoanalytic therapy remain highly controversial.

References

Black DM (2011) Why things matter. The place of values in science, psychoanalysis and religion. Routledge, Hove/New York

Blagys MD, Hilsenroth MJ (2000) Distinctive features of short-term psychodynamic-interpersonal psychotherapy: a review of the comparative psychotherapy process literature. Clin Psychol Sci Pract 7(2):167–188

Brenner C (1973) An elementary textbook of psychoanalysis. International Universities Press, New York

Chiesa M (2010) Research and psychoanalysis: still time to bridge the great divide? Psychoanal Psychol 27(2):99–114

Ellis A (1956) An operational reformulation of some of the basic principles of psychoanalysis. In: Feigl H, Scriven M (eds) The foundations of science and the concepts of psychology and psychoanalysis. University of Minnesota Press, Minneapolis, pp 131–154

Erdelyi MH (2006) The unified theory of repression. Behav Brain Sci 29:499–551

Esterson A (1996) Grünbaum's tally argument. Hist Hum Sci 9(1):43–57

Eysenck H-J (1952) The effects of psychotherapy: an evaluation. J Consult Psychol 16(5):319–324

Fisher S, Greenberg RP (1977) The scientific credibility of Freud's theories and therapy. Basic Books, New York

Fisher S, Greenberg RP (1996) Freud scientifically reappraised: testing the theories and therapy. Wiley, New York

Freud S (1895) Entwurf einer Psychologie (Project for a scientific psychology). Quoted from: Freud S (1987) Gesammelte Werke Nachtragsband. Fischer, Frankfurt am Main, pp 375–486

Freud S (1900) Die Traumdeutung (The interpretation of dreams). Quoted from: Freud S (1941) Gesammelte Werke, vols 2 & 3 (GW II & III). Imago Publishing Co, London

Freud S (1909) Analyse der Phobie eines fünfjährigen Knaben (Analysis of a Phobia in a five-year-old boy). Quoted from: GW VII, pp 241–377

Freud S (1916/1917) Vorlesungen zur Einführung in die Psychoanalyse (Introductory lectures on psychoanalysis). Quoted from: GW XI

Freud S (1923) "Psychoanalyse" und "Libidotheorie" (Two encyclopedia articles). Quoted from: GW XIII, pp 209–233

Freud S (1940) Abriß der Psychoanalyse (An outline of psychoanalysis). Internationale Zeitschrift für Psychoanalyse 25:7–67. Quoted from: GW XVII, pp 63–138

Grawe K (1998) Psychologische Therapie. Hogrefe, Göttingen

Grawe K, Donati R, Bernauer F (1994) Psychotherapie im Wandel. Von der Konfession zur Profession. Hogrefe, Göttingen

Greenwood JD (1996) Freud's 'tally' argument, placebo control treatments, and the evaluation of psychotherapy. Philos Sci 63(4):605–621

Grünbaum A (1979) Is Freudian psychoanalytic theory pseudo-scientific by Karl Popper's criterion of demarcation? Am Philos Q 16(2):131–141
Grünbaum A (1984) The foundations of psychoanalysis. University of California Press, Berkeley
Gyler L (2010) The gendered unconscious. Can gender discourses subvert psychoanalysis? Routledge, London/New York
Habermas J (1968) Erkenntnis und Interesse. Suhrkamp, Frankfurt am Main
Hilgard ER (1952a) Psychoanalysis as science. Eng Sci 16:11–17
Hilgard ER (1952b) Experimental approaches to psychoanalysis. In: Hilgard ER, Kubie LS, Pumpian-Mindlin E (eds) Psychoanalysis as science. Greenwood Press, Westport, pp 4–45
Kitcher P (1992) Freud's dream. A complete interdisciplinary science of mind. The MIT Press, Cambridge, MA/London
Kline P (1981) Fact and fantasy in Freudian theory, 2nd edn. Methuen, London/New York
Köhler T (2000) Das Werk Sigmund Freuds. Entstehung – Inhalt – Rezeption. Pabst Science Publishers, Lengerich/Berlin/Riga
Kupfersmid J (1995) Does the Oedipus complex exist? Psychother Theory Res Pract Train 32 (4):535–547
Lakatos I (1978) Falsification and the methodology of scientific research programmes. In: Worrall J, Currie G (eds) The methodology of scientific research programmes. Philosophical papers, vol 1. Cambridge University Press, Cambridge, pp 8–101
Levy RA, Ablon JS (2009) Evidence-based psychodynamic psychotherapy. Bridging the gap between science and practice. Humana Press, New York
Luborsky L, Singer B, Luborsky L (1975) Comparative studies of psychotherapy. Arch Gen Psychiatry 32:995–1008
Mancia M (2006) Psychoanalysis and neuroscience. Springer, Milan
Masling JM (ed) (1983) Empirical studies of psychoanalytical theories, vol 1. Analytic Press, Hillsdale
Newman WR, Grafton A (2001) Secrets of nature. Astrology and alchemy in early modern Europe. MIT Press, Cambrdige, MA/London
Pongratz LJ (1973) Lehrbuch der klinischen Psychologie. Psychologische Grundlagen der Psychotherapie. Hogrefe, Göttingen
Popper KR (1935) Logik der Forschung. Mohr Siebeck, Tübingen
Popper KR (1963) Conjectures and refutations: the growth of scientific knowledge. Basic Books, New York
Ricœur P (1965) De l'interpretation: essai sur Freud. Edition du Seuil, Paris
Ruse M (1982) Creation science is not science. Sci Technol Hum Values 7(40):72–78
Shedler J (2010) The efficacy of psychodynamic psychotherapy. Am Psychol 65(2):98–109
Smith ML, Glass GV, Miller TI (1980) The benefits of psychotherapy. Johns Hopkins University Press, Baltimore
Strachey J (1966–1974) Standard edition of the complete works of Sigmund Freud. Vol. I-XXIV. Vintage, London
Tschuschke V, Bänninger-Huber E, Faller H et al (1998) Psychotherapy research – how it should (not) be done. An expert reanalysis of comparative studies by Grawe et al. (1994). Psychother Psychosom Med Psychol 48(11):430–444
Uleman JS (2005) Introduction: becoming aware of the new unconscious. In: Hassin RR, Uleman JS, Bargh J (eds) The new unconscious. Oxford University Press, Oxford/New York, pp 3–15
Wallerstein RS (1986) Forty-two lives in treatment: a study in psychoanalysis and psychotherapy. Guilford Press, New York
Westen D (1998) The scientific legacy of Sigmund Freud: toward a psychodynamically informed psychological science. Psychol Bull 124(3):333–371
Wolpe J, Rachman S (1960) Psychoanalytic "evidence": a critique on Freud's case of Little Hans. J Nerv Ment Dis 130(8):135–148

Part VI

Nosology

WHO's Definition of Health: Philosophical Analysis

59

Jerome Bickenbach

Contents

Abstract

The notorious World Health Organization definition of health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" has been roundly, and justifiably, criticized by philosophers more or less since it first appeared in 1948. Despite its obvious conceptual, and practical, limitations, it launched a highly productive debate about the nature of health in which two major strategies have dominated: a descriptive or naturalistic approach in which health is operationally defined in terms of normal functioning understood entirely in the language of the biological sciences and a normative approach which insists that health cannot be understood until the salient fact that health is a human good is explained. This debate has revealed a dilemma: any philosophically acceptable definition of health must make a place for our powerful intuitions that health is both intrinsically and instrumentally valuable. Yet, unless the notion is firmly grounded in the biological sciences and

J. Bickenbach (✉)
Human Functioning Sciences, Swiss Paraplegic Research, Nottwil, Switzerland
e-mail: jerome.bickenbach@paraplegie.ch

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_48

susceptible to operationalization, it threatens to lose its scientific legitimacy. WHO has more recently and with far less fanfare, developed another definition of health "for measurement purposes" that recognizes the force of the dilemma and attempts, with debatable success, to address it.

Introduction

In the Constitution of the World Health Organization, approved in 1948, health is famously defined as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (WHO 1948). The extreme breadth of the definition – "physical, mental and social well-being" – and its unrealistically high threshold of good health, "complete," made it tempting to dismiss the definition as an aspirational gesture emblematic of a new era of optimism in international public health. Yet, philosopher Daniel Callahan took it seriously in 1973 and roundly criticized the definition for fatal overreach, arguing that to define "health" in terms of "well-being" transforms human happiness into a medical outcome and social ills like injustice, economic scarcity, and discrimination into medical problems requiring medical solutions (Callahan 1973).

Despite these limitations, however, in retrospect the WHO definition has considerably enriched the philosophical debate over the nature of health. It set the stage for an important and ongoing dispute between normative accounts of health and far more restrictively biological or biostatistically grounded views. Although there has recently been a resurgence of the strongly normative, WHO-style definitions, ironically WHO itself has taken steps toward a more narrow view motivated by the need to develop a conceptualization suitable to "operationalize health for measurement purposes" (Salomon et al. 2003).

The current situation reflects a dilemma: any philosophically acceptable definition must make a place for our powerful intuitions that health is both intrinsically and instrumentally valuable. Yet, unless the notion is firmly grounded in the biological sciences and so susceptible to operationalization, it threatens to lose its scientific legitimacy. Specifically, without operationalization, scientists will be unable to compare, let alone measure, the difference in the health of two individuals, or the same individual before and after a health intervention, or by extension of the relative health of subpopulations of individuals. The capacity for ordinal, if not cardinal, comparisons of states of health is not merely a scientific desideratum; it is essential for any scientific or policy application of the notion, including in particular the assessment of the performance of clinical health care or public health systems. But if the cost of securing scientific legitimacy is to undercut the commonly held belief that health is a human good (indeed, a plausible human right), then the resulting conceptualization is philosophically objectionable for a different reason. The more recent WHO definition of health "for measurement purposes" was developed with recognition of this dilemma, but it arguably fails to address it adequately.

In this chapter, the philosophical evolution of WHO's contribution to the definition – or more accurately, the conceptualization – of health will be traced and its philosophical impact described. The original, 1948 definition, and its philosophical critique, is the starting point. The critique began a fruitful philosophical debate between two starkly different approaches to health conceptualization represented here by Christopher Boorse's biostatistical account and Lennart Nordenfelt's action-theoretical normative account. What arises out of this debate is a philosophical impasse in which both approaches fall short, for opposing reasons. After a review of a recent resurgence of normativism that so far seems only to have reprised the problems of WHO's original definition, this chapter turns to the current endpoint in the evolution of WHO's definition of health and its limitations.

WHO Definition of Health: A Philosophical Evolution

The 1948 WHO Definition and Its Philosophical Critique

The Preamble to the Constitution of the World Health Organization, adopted and signed immediately after World War II in 1946, and entered into force in 1948, set out principles governing the establishment of this first international organization devoted to human health:

> Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.
>
> The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition.
>
> The health of all peoples is fundamental to the attainment of peace and security and is dependent upon the fullest co-operation of individuals and States.
>
> The achievement of any State in the promotion and protection of health is of value to all. (WHO 1948)

The first clause, whether it had been so intended or not, was quickly picked up as a definition of health. (Few noticed that it conflicted with the next principle inasmuch as the "highest attainable standard of health" suggests a flexible threshold of health, but the definition itself sets that threshold at "complete.") Implicit in the Preamble as a whole was a view attributed to sociologist HE Sigerist that health must be more than the absence of a problem; it must also be something positive (Sigerist 1941; Breslow 2006). Health is not merely an enjoyable state; it is something people seek out because it is both intrinsically and instrumentally valuable. The other innovation of the WHO definition – that health had mental and social dimensions – reflected the commonplace view that people are complex, biological, psychological, and social entities. Neither these two aspects of the definition were particularly controversial; it was the identification of health with human well-being that critics balked at.

In 1973, philosopher Daniel Callahan argued that the definition caters to a "cultural tendency" to define social problems as health problems, thereby blurring the lines of responsibility between the political order and the medical profession (Callahan 1973). Callahan noted, as others before him had (e.g., Wylie 1970), that the current rhetoric "medicalized" social problems because of a "grandiose" faith in science to cure sickness in all forms, biological, psychological, and social. This unbounded optimism, he insisted, was simply without empirical support. Neither is it plausible to suggest that all social evils are either caused by or examples of bad health: it is far more likely that political injustice and economic scarcity are the causes of these problems. Finally, transforming all human evils into health problems undermines human freedom and responsibility.

The ideological assumptions bound up in the WHO definition led philosophically to an abuse of language and common sense, Callahan concluded. Surely, the normativity of health can be preserved without insisting that it is the source of all human value. Health is undoubtedly a human good, but it is not the only human good. Some minimal level of health is probably essential to achieve any possibility of human happiness; yet, at the same time, some degree of ill-health is perfectly compatible with happiness, given that no one could hope to be in a state of "complete physical, mental, and social well-being."

To explain what might have gone wrong, Callahan observed that health is intuitively both a natural norm and an ethical ideal. Viewed as a norm, health is simply a matter of the heart, lungs, kidneys, and other body parts functioning up to a threshold of normality that can be established empirically and statistically. Yet, Callahan noted that thinking about health as a norm is unsatisfying because it does not address the obvious question why anyone would care about statistically normal functioning unless dipping below that threshold was unpleasant, inconvenient, painful, or generally a bad thing. Why too should society take any interest in subnormal bodily functioning unless, in the aggregate, it has socially adverse implications? There is no escaping the intuition that health is not merely the description of a state of biological affairs, matched against some statistically determined norm; it is also an ideal people take very seriously indeed. Health is a morally significant normal bodily functioning.

The philosophical challenge, however, is to do justice to both health as norm and as ideal. To insist that health describes a state of affairs, in principle reducible to biological and psychological functioning, and assessed in light of norms generated by population statistics (the basis for what came to be called descriptive theories of health) fails to capture intuitions about what makes health valuable; yet, accounts that focus on the normative significance of biological and psychological functioning (normative theories) fail for their part to provide a sound conceptual basis for the health sciences. The WHO 1948 definition thus became the starting point for an increased philosophical interest in the conceptualization of health. Whether motivated by a rejection of the WHO definition or an affirmation of its underlying insights, the subsequent philosophical literature took the definition as its starting point.

Normative and Descriptive Accounts of Health: Boorse and Nordenfelt

The most prominent advocate of the descriptive approach to health is Christopher Boorse who in a series of seminal articles in the mid 1970s mapped out what has come to be called the "biostatistical theory of health" (Boorse 1975, 1976, 1977). Initially, his concern was to reject normativism in health, especially in the characterization of mental health and in particular a rejection of the WHO definition. He did so in terms of the conceptual difference between a disease and an illness, the first being a biological state of pathology and the second a normative disvalued experience, roughly linked to pathology. Only the first is directly relevant to the conceptualization of health.

Boorse argued that biological functions can be fully described in terms of a hierarchy of goals ascribable to different levels of organisms: cells have metabolism functions, organs have body level functions such as blood circulation, whole organisms have eating and moving around functions, and all of these functions causally contribute to the species-typical goals of survival and reproduction. But this teleology need not be normatively understood since at the end of the day this is simply how organisms behave. So understood, the health of an organism is functional normality. The notion of a biological function is central to Boorse's approach (Boorse 1976), and philosophically it has drawn the most criticism from those who, in general terms, are otherwise quite sympathetic to Boorse's descriptivism (e.g., Engelhardt 1984; Caplan 1993; Beauchamp and Childress 2001).

All descriptivists concur that functional normality can be neutrally described, despite the fact that the state of functional normality tends to be judged as desirable. This is because the evaluation of normality is based on grounds and for reasons that are only tangentially relevant to individual biology or evolutionary theory. Thus, it is quite easy to imagine a "better" normal functioning than that which evolution has provided human beings; on the other hand, in some circumstances having a disease contributes to overall well-being (if, e.g., the disease would disqualify a person from military conscription). Describing and valuing are fundamentally different operations, and there is no reason to think they must be essentially linked in the conceptualization of health.

Pressed to explain the biological significance of "normal functioning," Boorse argued that normality is primarily a statistical construct guided by scientific assumptions about or hard evidence about species-typical functioning levels. Diseases are theoretical entities health scientists defined in terms of signs and symptoms of less-than-optimal functioning at some biological level – ultimately, reflecting the evolutionary imperative of species survival. Other descriptivist accounts have put more reliance that Boorse did on the power of evolutionary theory and homeostasis to account for normality in functioning (Bechtel 1985; Kovács 1998; Ananth 2008).

For Boorse's part, he acknowledged that functional normality was neither a necessary nor sufficient condition of health (red hair is not statistically normal, and there are diseases such as tooth decay that are nearly universal). But as an

operationalization of disease (or more generally ill-health), functional normality is the most reliable indicator. Should the biological sciences devise a more sensitive indicator – perhaps one that incorporates epigenetic insights or some other more fundamental level of explanation – then scientists would turn to it. But, philosophically, the quality and reliability of the indicator of functional normality are irrelevant: at the bottom, the concept of health is in principle fully describable in normatively neutral terms. That health is universally valued and decrements in health caused by disease and injury universally disvalued are sociological facts that explain health-seeking behaviors, but they are conceptually independent of the nature of health and decrements of health.

Although on the first blush nothing could seem to be further from the WHO definition than Boorse's account (and such was his intention), in fact they are not incompatible in at least one respect: no advocate of the WHO definition would deny that health and mechanisms involve in impairing health are at the bottom biological phenomena. Arguably, the WHO definition only tells us the manner in which health is valuable to human beings – why it is individually and socially important – but leaves to biological scientists the description of states of health and ill-health. The normativist, in short, need not advocate the abandonment of the biological sciences or medical practice – he or she is merely interested in a different, but more salient, conceptual feature of health: why we value it.

Lennart Nordenfelt has been the leader in this second, normativist approach to health conceptualization, arguing that health cannot be understood philosophically unless and until it is clear why it is valuable. Health is not merely a biological norm, it is an ideal (Nordenfelt 1987, 1993). Health is about the capacity to act and to live a full life according to one's life plans. More formally, a person is healthy just in case he or she is in a bodily and mental state such that he or she has the ability to realize all his or her vital goals, in standard circumstances. A vital goal for an individual is one that is necessary for minimal happiness (understood robustly as a version of Aristotle's eudaimonia and not merely positive affect).

Nordenfelt took care to avoid some obvious traps of his theory. He took into account and sought to explain some apparent counterexamples, e.g., that people are often mistaken about what they believe will make them happy and that they can sincerely hold unrealizably vital goals or can, by pure luck, achieve minimal happiness despite utterly lacking the ability to do so. In particular, he recognized that a person may achieve minimal happiness with acceptable health, a level far below complete health. Like Boorse, in short, Nordenfelt begins with the WHO definition, but in his case he is more sympathetic to it and hoped to preserve it by crafting a philosophically sophisticated version that avoids obvious criticism.

But Nordenfelt was also keen to reject Boorse's biostatistical theory, not because he thought that health was not rooted in biology but because the mechanisms that limit individual health cannot be identified as diseases or injuries simply because they result in statistically abnormal levels of biological functioning at some level of the organism. That is not how medical theorists have identified diseases and other decrements of health, he insisted. Always in the forefront is the view that only abnormalities in functioning that also reduce the ability of the individual from

realizing his or her vital goals, and so achieving minimal happiness, are decrements in health. Diseases are identified through the lens of vital goals in the first instances and only then in terms of biological abnormality of functioning.

Nordenfelt and normativists generally characterize their views as being holistic in the sense that health is intuitively attributed to individual persons and only metaphorically and by extension to cells and organs (or by aggregation to populations). And on this they have common intuitions on their side: "To be healthy is to function well. It is to feel strong and vital. It is to lack pain and disability. It is to be able to work, to be able to handle one's daily life and enjoy one's life." (Nordenfelt 1993, 83) A concern about cells, organs, and biological functions is the (perfectly legitimate) concern about the mechanisms behind the phenomena of health and disease. Scientists need to know about the bodily machinery to inform their health sciences. But conceptually, the biomedical sciences cannot explain why it is commonly understood that hearing or vision loss, pain, infections, or diseases like diabetes, spinal cord injury, or cancer matter very much to human being or why societies invest social resources into responding to these problems in living. Conceptually, the only way to explain these hard facts is an account of health that centers on what matters to people with respect to their bodily and mental functioning, and this must, in one way or another, analytically connect with human well-being.

Recent Resurgence of Interest in Normativism

The philosophical debate between Boorse and Nordenfelt and their defenders was at its height from the late 1970s to the early 1990s, primarily in the English-speaking philosophical world. It was also during this period that the WHO discovered that it could make good political use of its 1948 definition to further the cause of international public health. In a series of important declarations and other pronouncements during this period, WHO was able to transform its definition into a successful advocacy tool by highlighting an implicit theme of the definition: that health promotion is not exclusively a matter of developing more and more sophisticated medical diagnostic and prevention tools; it is also, and often more importantly, a matter of isolating the social determinants of ill-health across the population. As one of WHO leading advocates of the human right to health Jonathan Mann put it, the WHO definition "helped to move health thinking beyond a limited, biomedical and pathology-based perspective to the more positive domain of "well-being."" In addition, by explicitly including the mental and social dimensions of well-being, WHO radically expanded the scope of health and, by extension, the roles and responsibilities of health professionals and their relationship to the larger society (Mann et al. 1994; Kickbusch 2003).

Perhaps because of the lasting significance of the 1948 WHO definition in international public health, there has been a resurgence of interest in normative conceptualizations of health in recent years. Although some philosophers found some common ground in the two approaches (Schramme 2007), others, especially in the area of mental health, argued that the Boorsian natural function approach was

unable to account for why mental illnesses are viewed as problematic (Varga 2011). Normativism seems to have won out. Unfortunately, the rejection of descriptivism has also led to normative accounts that lack the philosophical rigor of Nordenfelt's theory with the result that they have reprised some of the peculiarities of the WHO 1948 definition.

In 2011, Machteld Huber and colleagues proposed an "adaptation" of the WHO definition made necessary by the profound epidemiological shift in the worldwide burden of disease since 1948 from acute and communicable diseases to noncommunicable diseases, a shift made more dramatic by population aging and the fact that people are living longer with chronic diseases (Huber et al. 2011). These facts convinced the authors of the need to take into account the increasing importance, in public health, for individuals to adapt to environmental changes and to self-manage their chronic illnesses.

For a descriptivist, adaptation and self-management are irrelevant to the conceptualization of health and ill-health, although certainly significant to frame health intervention at clinical and population levels. If self-management, for example, helps to limit the range of potential comorbidities or functional consequences of a chronic condition such as high blood pressure, then interventions should properly focus on developing self-management skills. Chronic health conditions are by definition incurable – although their onset may be preventable – so addressing adaptation and self-management seems imminently sensible.

For a normativist, the importance of adaptation and self-management takes on a very different role in helping to explain the underlying human value that effective health interventions enhance. This focus leads Huber and colleagues to conclude that since an adequate level of capability to adapt and self-manage enhances one's well-being, it follows that health *is* the capability to adapt and self-manage. Moreover, since they are eager to affirm that "social health" is an essential component of health, they require a version of this self-management capability for the social sphere. For this purpose, they included in their account the capability "to participate in social activities including work."

The end result is a definition of health that falls victim to two substantial logical confusions (that normativist accounts tend to be prone to). The first is to conflate cause and effect: in our example, to confuse the impact of a plausible social determinant of health – for example, unemployment rates or some other force limiting the effectiveness of an individual to secure "social health" – with a component of the concept of health. Another more blatant example of this confusion at work can be found in the so-called Meikirch Model of Health in which good health is conceptualized as "individual potentials" – either biologically given or "personally acquired" – that produces a capacity that allows an individual to adequately or optimally respond to the "demands of life" in a context shaped by social and environmental determinants (Bircher and Kuruvill 2014). Personally acquired individual potentials are claimed to include "all of the physiological, mental, and social resources a person acquires during life" – that is to say, resources such as a good job, loving family relationships, educational attainment, and income level. Here again, plausible determinants of health are conflated with components of the concept of health.

The second logical error inherent in the Huber et al. definition is reductivism. For his part, Nordenfelt was careful to characterize the normative essence of health in very open and general terms, namely, as "a bodily and mental state sufficient for the ability to realize one's vital goals." Arguably, the ability to adapt and self-manage is part of that general ability, and indeed it may well be a necessary condition of the ability to realize vital goals. But it is very unlikely to be a sufficient condition for that general ability. If a person has a low level of self-esteem or personality characteristics that undermine his or her motivation to use highly developed skills to adapt and self-manage, then it is unlikely that this person would be able to realize his or her vital goals. Alternatively put, although it would be helpful to one's health to be able to adapt and self-manage, it is certainly imaginable that a person who was a terrible self-manager, by good luck, nonetheless enjoys full health. By reducing the normative essence of health to a single, albeit important, capability, the Huber et al. account is vulnerable to damning counterexamples.

Recent normativist accounts have also reprised what, to many critics, was the main problem with the WHO definition: an exaggeration of the importance of health as a human value. One recent normativist theory demonstrates this problem in stark terms. Building on Amartya Sen's influential capability theory (see, e.g., Sen 1999), Sridhar Venkatapuram has conceptualized health in terms of its potential as a "meta-capability" (Venkatapuram 2011). Incorporating but greatly expanding Nordenfelt's account of health, Venkatapuram has argued that health is both a necessary and sufficient capability to achieve all aspects of the good human life, well-being at its most expansively defined – a veritable *summum bonum*. The social impact of this normative inflation is noteworthy: Venkatapuram argues that the importance of health is such that a truly just society will be organized so as to effectively respond to every potential determinant of health so as to eliminate all forms of inequalities, physical or social, in the name of population health. This is health overreach on a grand level.

WHO's New Approach

The recent proliferation of normativist definitions of health reflects a continuation of the tradition which begun with the WHO definition in 1948, inspired by the insight that health is an aspect of human flourishing and so intrinsically a good thing for all to enjoy. What makes health a good thing and whether it is the only human good or just an especially or uniquely important one are open questions, and a normative theory will gain or lose credibility depending on how it addresses them. But though the WHO definition can be credited with the "normative turn" in conceptualizing health, recently as part of a multiyear project for health system performance assessment (WHO 2000), WHO has taken a step clearly in the direction of a descriptivist approach to health, a conceptualization of health "for measurement purposes."

Although informed by Boorsian descriptivism, WHO's more recent account of health was only possible because the development of WHO's *International Classification of Functioning, Disability and Health* (ICF) (WHO 2001). ICF is an

epidemiological standard, a classification and coding system for health and disability data. Significantly, it is grounded in the notion of human "functioning," which parallels Boorse's own notion of "function" (Boorse 1976). ICF is a classification of domains of human functioning, discrete body functions (including mental functions), bodily structures, and the full range of simple to complex human behaviors, actions, and complex social patterns of behaviors and actions (such as being a sibling, being employed, participating in community activities). The ICF, in short, is a complete classification of human functioning for the purpose of operationalizing health.

The motivation for WHO's new definition of health is measurement, without which it is not possible to compare health over time between individuals, individuals over time, and across populations and over time (Salomon et al. 2003). Without meaningful measures of health, the goals of public health are unachievable: it would be impossible to know whether public health interventions changed health or reduced health inequalities across subpopulations. Without measurement there is no proper science of health. It has been a standard practice, at least in public health, to "measure" health states of populations in terms of standardized health indicators, such as incidence of chronic illnesses, infant mortality rates, or population survivorship rates (see examples in Goldsmith 1972; Bergner 1985; McDowell 2006). Indicators are, of course, proxy measures, and it was the goal of WHO to achieve a more robust measurement of health by means of an operational conceptualization of the notion. At the same time, the authors appreciate that little would be gained if the resulting conceptualization was too distant from the common notion of health and in particular our intuitions about health as a human value. Thus, the first step in the development of a new WHO definition of health, therefore, was to identify "consensus points" about the concept of health:

1. Health is a separate concept from well-being, and is of intrinsic value to human beings as well as being instrumental for other components of wellbeing;
2. Health is comprised of states or conditions of functioning of the human body and mind, and therefore any attempts to measure health must include measures of body and mind function; and
3. Health is an attribute of an individual person, although aggregate measures of health may be used to describe populations. (Salomon et al. 2003, 303)

It follows from these simple propositions that there is a clear, conceptual distinction between health and its determinants and consequences, a confusion that is the downfall of many normativist accounts of health. The distinction between determinant and concept follows straightforwardly from the first clause of the third consensus point, as does the core descriptivist premise that the language of health is that of the biological sciences. Income levels, employment rates, and social networks – all of these phenomena are likely determinants of a person's health, but for all of that, they are not attributes of an individual person and so not part of the concept of health.

The second consensus point is the essence of the new WHO account of health as "an intrinsic, multidimensional attribute of individuals" with universal, cross-population, and cross-cultural validity. The account is universal simply because it is grounded in states or conditions of functioning of the human body and mind. The ICF

is a classification of these domains of functioning, decrements in which are impairments if the limitation is in a body function (or structure) or activity limitations and participation restrictions if the limitations is in what the person does or performs. The account, however, requires that these "states or conditions of functioning" refer to intrinsic capacities of an individual, rather than descriptions of what individuals do or perform in their actual environments. This is an important qualification, and as the philosophical plausibility of the WHO conception of health depends on it, it is worth developing the distinction between capacity and performance more fully.

As the model of functioning and disability embodied in the ICF makes clear, the nature, quality, and extent of what a person *does* (acts, executes, performs, behaves, and so on) often depend considerably on features of the environment in which the person acts. This is especially significant when the concern is to determine the state of a person's health, with limitations on what the person can do because of their intrinsic biological state. Thus, a person who has an impairment in hearing may in fact be able to hear with a hearing aid; similarly, a person with lower body muscle wastage may not be able to climb stairs in a public building because they are too steep but will be able in their own home where the stairs have been modified to accommodate this impairment. In short, to accurately assess a person's functioning in different domains – hearing, seeing, walking, climbing, grasping, carrying an object, and so on – it is important to discount the impact of the environment in which the person performs actions that depend on these functionings. Features of the physical and social environment may make it possible for the individual to perform better than he or she can intrinsically (when assistive technology or environmental modification facilitates performance); by the same token, other features may hinder performance. In either instance, to get at a person's health, the positive or negative effect of the individual's environment needs to be discounted. The result, in the ICF language, is the person's intrinsic functioning capacity.

But given the substantial number of bodily and person-level functionings that constitute the full repertoire of human functioning, it would be impractical to define health operationally in terms of all of these functionings. Though a practical rather than a conceptual issue, it is a measurement challenge that the WHO conception needs to resolve. Conceptually, the new WHO definition is completed by the three guiding principles quoted above, but as the point of the conceptualization is practical operationalization for measurement purposes, the authors are very much obliged to offer a solution to the challenge of identifying which human functionings are at the conceptual heart of the notion of health.

They approach this challenge by sketching out functioning domain selection guidelines: the domains of functioning sufficient for operationalizing the concept of health for measurement purposes should be those that have intuitive, clinical, and epidemiological significance; are classified in the ICF; are amendable to self-report, observation, or direct measurement; are cross-population comparable; and, finally, are "comprehensive enough to capture the most important aspects of health states that people value" (Ibid. 310).

This last criterion is not so much a measurement concern as a matter of face validity. When measuring health, it is important to measure what it is about health

that makes health something perceived to be both intrinsically and instrumentally valuable. This should be taken as a gesture toward the normativist challenge, but it is not a complete answer to it. In effect, the new WHO definition of health turns the issue of the normative significance of health into a technical challenge, leaving unexplained why health matters to us. Given that the definition is held out to be cross-culturally universal as well as scientifically adequate, the failure to pinpoint the source of the value of health can fairly be seen as a significant failure of the WHO definition, at the conceptual level. Even if we are confident that the domains of functioning we select serve the purposes of scientific measurement, the resulting operationalization does not, on its own, give us an explanation why, in every culture, health is conceptually understood as a human good.

Conclusion

The 1948 WHO definition of health and the current, descriptivist WHO definition "for measurement purposes" reflected a persistent dilemma in the philosophical challenge of defining health. Any philosophically acceptable definition must take into account our powerful intuitions about the intrinsic and instrumental value of health. Health may not be the same as well-being or the summum bonum, but it is a component (or determinant) of human well-being and indisputably a human good and a central one at that (see Daniels 2008). Yet, unless the notion is firmly grounded in the biological sciences and understood as an attribute of the person, the concept resists operationalization and threatens to lose its scientific legitimacy. It is not just the World Health Organization that requires a notion of health in terms of which we can compare the health of an individual before and after a clinical intervention or a population of people before or after a health promotion or other public health intervention. As an unexplained, ineffable, indefinite, or inherently subjective phenomenon, the notion of health is not of particular use to us, nor would it have any useful input into how we structure our social institutions and systems to respond to actual human need. This is the philosophical challenge of defining health.

Definitions of Key Terms

Descriptive theory of health	A philosophical theory of health based on the premise that health is an attribute of an individual fully explainable in the language of the biological sciences.
Normative theory of health	A philosophical theory of health premised on the view that it is of the essence of health that it is an intrinsic and instrumental human good.
Operationalization of health	The process by which a conceptualization of health is transformed into a set of operations, procedures,

	or explicit criteria that define elements of health that can be measured in one manner or another.
Functioning	(In the International Classification of Functioning, Disability and Health, WHO 2001) a domain of health including specific body functions and structures and all human behaviors, movements, and actions, from the simplest individual movement or action to the most complex, socially constructed, action that constitute human activity.

Summary Points

- The 1948 Constitution of WHO defined health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity."
- Though strongly criticized, the WHO definition of health set the stage for an ongoing philosophical debate about the definition of health.
- The dominant theories of health emphasize either the biological and scientific core of the notion (descriptivist or "naturalistic" accounts) or the consensus that health is an intrinsic and instrumental human good (normativist accounts).
- Despite decades of high-quality philosophical debate about the concept of health, there remains a persistent dilemma: neither a descriptivist nor a normativist account of health is adequate, but these two approaches are in fundamental conflict.
- After two decades of relative inactivity in philosophical treatments of the concept of health, recently there has been a resurgence of interest in normativist definitions.
- It is essential for the scientific status of health sciences, and in particular for assessing the effectiveness of individual and population health intervention and comparing the health of individuals and populations, to use a conceptualization of health that is operationalizable for measurement.
- Although the 1948 WHO definition remains in use, WHO itself has based its own scientific work on a very different, basically descriptivist, account of health "for measurement purposes."
- The most recent WHO definition of health, although it gestures toward the normativist approach while being firmly descriptivist, nonetheless fails to adequately account for the common perception that health is both intrinsically and instrumentally valuable.

References

Ananth M (2008) In defense of an evolutionary concept of health nature, norms, and human biology. Ashgate, Aldershot

Beauchamp T, Childress J (2001) Principles of biomedical ethics, 5th edn. Oxford University Press, New York

Bechtel W (1985) In defense of a naturalistic concept of health. In: Humber J, Almedereds R (eds) Biomedical ethics review. Humana Press, Clifton, pp 131–170
Bergner M (1985) Measurement of health status. Med Care 23:696–703
Bircher J, Kuruvill S (2014) Defining health by addressing individual, social, and environmental determinants: new opportunities for health care and public health. J Public Health Policy 35:363–386
Boorse C (1975) On the distinction between disease and illness. Philos Public Aff 5:49–68
Boorse C (1976) Wright on functions. Philos Rev 85:70–86
Boorse C (1977) Health as a theoretical concept. Philos Sci 44:542–573
Breslow L (2006) Health measurement in the third era of health. Am J Public Health 96:17–19
Callahan D (1973) The WHO definition of 'health'. Stud Hastings Cent 1:77–87
Caplan A (1993) The concepts of health, illness, and disease. In: Bynum W, Porter R (eds) Companion encyclopedia of the history of medicine, vol I. Routledge, London, pp 233–248
Daniels N (2008) Just health: meeting health needs fairly. Cambridge University Press, New York
Engelhardt HT (1984) Clinical problems and the concept of disease. In: Nordenfelt L, Ingemar B, Lindahl B (eds) Health, disease, and causal explanations in medicine. Reidel Publishing Company, Dordrecht, pp 27–41
Goldsmith S (1972) The status of health status indicators. Health Serv Rep 87:212–220
Huber M, Knottnerus J, Green L, van der Horst H, Jadad A et al (2011) How should we define health? BMJ 343:d4163. doi:10.1136/bmj.d4163
Kickbusch I (2003) The contribution of the World Health Organization to a new public health and health promotion. Am J Public Health 93:383–388
Kovács J (1998) The concept of health and disease. Med Health Care Philos 1:33–48
Mann J, Gostin L, Gruskin S, Brennan T, Lazzarini Z, Fineberg H (1994) Health and human rights. Health Hum Rights 1:6–23
McDowell I (2006) Measuring health: a guide to rating scales and questionnaires, 3rd edn. Oxford University Press, New York
Nordenfelt L (1987) On the nature of health. An action-theoretic approach. Reidel Publishing Company, Dordrecht
Nordenfelt L (1993) Quality of life, health and happiness. Averbury, Aldershot
Salomon J, Mathers C, Chatterji S, Sadana R, Üstün TB et al (2003) Quantifying individual levels of health: definitions, concepts, and measurement issues. In: Murray CJL, Evans CJL (eds) Health systems performance assessment debates, methods and empiricism. World Health Organization, Geneva, pp 301–318
Schramme T (2007) A qualified defence of a naturalist theory of health. Med Health Care Philos 10:11–17
Sen A (1999) Development as freedom. Oxford University Press, Oxford
Sigerist HE (1941) Medicine and human welfare. Yale University Press, New Haven
Varga S (2011) Defining mental disorder. Exploring the 'natural function' approach. Philos Ethics Humanit Med 6:1
Venkatapuram S (2011) Health justice. Polity Press, Cambridge
WHO (1948) Constitution. World Health Organization, Geneva
WHO (2000) The world health report 2000: health systems: improving performance. World Health Organization, Geneva
WHO (2001) International classification of functioning, disability and health. World Health Organization, Geneva
Wylie C (1970) The definition and measurement of health and disease. Public Health Rep 85:100–104

Health as Notion in Public Health

60

Thomas Schramme

Contents

Abstract

Public health is a scientific and practical endeavor. It aims at preventing disease and promoting health in a population. Public health has a specific way to use the concept of health. It is positive in the sense that it facilitates the measurement of the health status of a population over and above the absence of disease. Health in public health is a gradual, not an absolute, notion. Public health also targets health risks or health dispositions, which should not be confused with intrinsic health statuses. This chapter also discusses the aspect of referring to health within a population, which poses some issues of measurement. Finally, it is discussed what normative issues are due to the specific understanding of health in public health.

T. Schramme (✉)
Department of Philosophy, University of Liverpool, Liverpool, UK
e-mail: t.schramme@liverpool.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_49

Introduction

Modern societies are not only concerned with the health of individuals in terms of providing for curative medicine by a publicly funded system of health-care resources, professionals, and institutions. States also focus on preventing disease through measures such as provision of clean water, containment of contagious diseases, and screening programs to identify genetic dispositions. In addition, the idea of health promotion has gained more attention. The improvement of health conditions, especially for vulnerable and disadvantaged groups, has become a major political concern on a global scale (WHO 1986, 2010). Public health is the theory and practice of fulfilling this societal task of protecting the health of the population. From a philosophical perspective, it is important to ask what "health" means in public health. This issue is rarely discussed, maybe because it is thought to have a straightforward answer. But it will be seen that, on the contrary, the notion of health in public health is far from clear.

A first issue concerns whether health is understood as a negative or positive concept, that is, whether it is interpreted as the absence of disease or as something over and above such a minimal understanding. Public health indeed seems to include both these aspects, as can be seen in two definitions of public health. The first focuses on the negative aspects: "Public health is the prevention of disease and premature death through organized community effort" (Beauchamp 1995, p. 2210). The second definition states that public health is "the science and art of preventing disease, prolonging life and promoting health through organised efforts of society" (Faculty of Public Health of the Royal College of Physician of the United Kingdom, quoted in Nuffield Council on Bioethics 2007, p. 6). In this definition the idea of health promotion is explicitly mentioned. It is in relation to such advancement that the possible expansion of the concept of health over and above the absence of disease occurs.

A second issue concerns the emphasis of public health on the social determinants of health. It has been an important finding that the health status of individuals does not only depend on their internal condition, such as the well-functioning of their organs, but also, and maybe more importantly, on their living conditions. This was first mainly seen in relation to the physical environment, for instance, in regard to hygienic conditions or housing. In the last decades, the social determinants of health have been acknowledged to extend to far more factors, including working conditions, social relationships, and public safety. These more remote "causes of the causes," however, pose a theoretical problem, because being at (increased) risk of falling ill, for instance, because of a stressful and oppressive work atmosphere, is not the same as being in ill health. Conceptually, there is an important distinction between a particular health disposition and a status as being healthy. As will be seen, this distinction is related to the significant difference between a comparative and an absolute understanding of health. In focusing on healthy circumstances, public health is prone to confuse health dispositions and health statuses.

Third, public health is not only concerned with the health of individual citizens but with the health of populations. This raises conceptual issues concerning the

object to which a certain health status is ascribed. Is it the aggregate sum of individuals in a particular population that makes up the group's health status or is a population deemed to be an entity in its own right that can also have a particular health status? To ascribe a certain level of health to populations does not only raise ontological questions concerning individualism versus collectivism but also practical problems of measurement.

Finally, the very fact that public health is not just a science, but a practice aiming at a common good (cf. Parmet 2009, p. 12), makes the concept of health in public health in a distinctive way a normative notion. Health is here obviously understood as a good, which should be protected and promoted. Indeed, it is usually seen as a moral concern if populations within a society, or globally, have different health statuses, not due to their own choices, but because of the social conditions they are living in. The conceptual problem here is one of a potential confusion between a scientific and possible value-neutral account of health in contrast to an interpretation of health in terms of political interests.

Negative Versus Positive Interpretations of Health

It is important to acknowledge that the concept of health is often used in a special sense in the theory and practice of public health. The relevant understanding of health is, in certain respects, discontinuous with the received view in general medicine. In medicine, health is commonly understood in a negative way, as the absence of disease or as medical normality. This is a minimal and absolute concept of health. A person is either healthy or not, there are no grades of health. In order to be regarded as healthy, it is merely necessary not to be in any pathological condition. To be sure, there are attempts to conceptualize health in a positive way, for instance, in the well-known formulation of the World Health Organization: "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." Yet, this definition has had no impact on medical theory or practice and has actually been criticized for its lack of distinction between well-being or happiness and medical health (Callahan 1973). In terms of Fig. 1 below, medicine is interested in the left side of the spectrum.

In public health, health is usually not seen in contrast to disease, but as a condition that can be present at a certain level, even where a person has a condition that is clinically subnormal. This makes sense insofar as people can cope with disease. Their well-being or welfare might not be affected by a medical condition, so they can be healthy despite having a pathological condition. Once such a gradual understanding is introduced, health can also be found at a higher level over and above the absence of disease. In this respect, persons who are either less likely to fall ill or who are fitter than others in terms of their organismic functioning have a higher grade of health.

In summary, conditions of persons can be understood in different ways. Health can mean the absence of disease, as in the common medical perspective. Health can also mean a point on a continuum, stretching from the absence of disease up to a

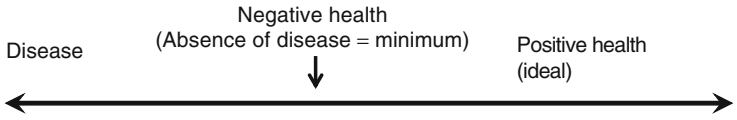


Fig. 1 Negative and positive health

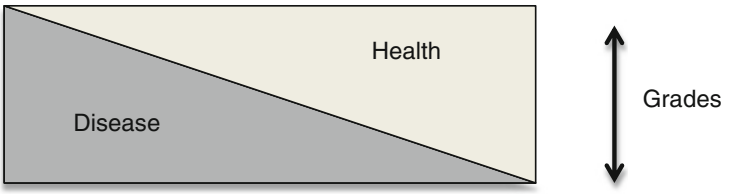


Fig. 2 Health grades

state of ideal health, as in Fig. 1. Health can finally be seen in combination with disease, in a more holistic fashion, which includes the life circumstances of persons. Here it also makes sense to talk about grades of health. This perspective can be seen in Fig. 2. Both ways of referring to grades of health, in terms of levels of quality and in terms of a holistic assessment, are found in public health. In fact, public health necessarily requires a gradual perspective, because otherwise there could be no comparisons of health statuses between individuals or populations. Also, to conceive of health in a holistic fashion naturally leads to an inclusion of social aspects of health, which is one of the distinctive theoretical and practical contributions of public health (cf. Arah 2009).

Positive health can be understood in relation to medically defined conditions, as superior or enhanced organismic functioning. The task of health promotion might be interpreted in this sense, for instance, when the lung capacity of people is increased. It might have been medically normal before, but enhanced via a public health intervention, such as incentivizing the use of bicycles. Another way to understand positive health in public health is in terms of welfare more generally, not just in relation to organismic functioning. Since in public health social determinants of health are seen in close relation to medical conditions, it is but a small step to a welfare notion of health (Venkatapuram 2011; cf. Holland 2014, p. 109 ff.). This is an interpretation of the notion of health that includes conditions that are internal and external to the person under the umbrella of health, such as being able to experience nature or to have occasions for recreation, which are not seen themselves as health conditions in medicine. Positive health here means to have superior or enhanced capabilities.

Promotion of health can therefore refer to the improvement of health over and above the absence of disease. It is focused on conditions of people. The practice of public health also includes the promotion of health in terms of advancing the awareness of the value of health. This is a focus on an abstract quality of conditions of people, the importance of health, not on the conditions itself. An understanding

of health as welfare notion might at the same time make a straightforward case for health being of significant value, but also brush over important differences between welfare and organismic conditions. When health is understood as such an overarching value and as a gradual, positive notion, we might tend to see any condition below the ideal as an impairment of welfare. It is therefore vital to always be clear about the understanding of the concept of health in public health contexts.

Intrinsic and Instrumental Health

Individuals can be healthy in the sense that they are not ill or do not have a disease. Here, their health status is intrinsic insofar as it is determined by their somatic and mental condition only. A person is either healthy or not healthy in this sense. Individuals can also be healthy in the sense that they are not likely to fall ill. A person who does not smoke and exercises regularly is healthy in this respect. Hence, it might be more fitting to say that they live a healthy life. Here, health is seen in its instrumental aspects, as a means to stay (intrinsically) healthy. This can be turned into a gradual standard: someone can be healthier than someone else or healthier than before. In addition, we can ascribe such instrumental healthiness to environments, not just conditions of people, as the circumstances causally have an impact on the ensuing physical and mental conditions of people. Clean air and nutritious food are healthy, whereas working in mines or being persecuted is not healthy.

The comparative perspective of public health therefore depends, up to a point, on the fact that people can have certain dispositions to fall ill. Yet health dispositions are not themselves health statuses, but propensities to fall ill. Even individuals who have a high propensity to develop a disease, say, because they are carriers of certain genes, are not therefore unhealthy in the absolute sense of the term. Intrinsic health and instrumental health therefore need to be kept apart (Boorse 1977, p. 553), though obviously both medicine and public health rely on assessments of instrumental health in their practice. Public health even has an understanding of instrumental health, which is not based on individual risk but on health dispositions within populations. Hence, individuals might have a worse status of instrumental health than others (in a particular respect, say, regarding respiratory capacity), in virtue of their membership in a particular targeted population, for instance, workers in a destitute area. Health risk, as in the individual case, is based on statistical findings and probabilistic theory (Parmet 2009, p. 17; Broadbent 2013, p. 129 ff.). When applied to populations, epidemiological findings also contribute to the assessment of health.

It is a mistake to confuse the intrinsic health status of individuals or populations with their internally or externally determined health dispositions or risks. This does not, of course, speak against researching on and politically discussing health risks. But it is wrong to say that risks themselves make us unhealthy in the intrinsic and absolute sense of the term (cf. John 2009). Conceptual confusion might lead to problematic decisions when developing public health policies.

Population Health: Aggregate or Distinctive Concept?

Usually, public health experts focus on particular socioeconomic groups, for instance, unemployed persons or single mothers. So when epidemiologists refer to population health, they usually mean the statistically aggregated sum of individual health traits or health statuses. The aggregation that leads to an account of population health depends on certain summary measures (e.g., Murray et al. 2002). Alternatively, but maybe more controversially, population health might be seen as a distinctive category, more than just the aggregate of individual health statuses (e.g., Arah 2009).

The particular way groups or populations are determined depends on the purpose of a study. Ultimately considerations in public health rely on hypotheses about social or socioeconomic determinants of health, or – to use another expression familiar to a public health perspective – the "causes of causes" (of health status). Hence, epidemiologists aim at findings about possible correlations between particular circumstantial aspects of citizens and their health conditions. In this respect, the population perspective is instrumental for understanding causes of individual disease that are of a social nature (e.g., Rose 1985).

Findings may be sought regarding socioeconomic aspects, such as income, educational background or gender, or behavioral aspects, such as lifestyle and diet. With these statistical correlations, it is possible to make comparisons between populations regarding their health, even on an international level. Obviously it is also possible to compare different policies in tackling possible inequalities. In more popular publications, public health scholars then end up with simple slogans, such as "inequality is bad for your health" (Daniels et al. 2001) or "uneducated people die younger," which only make sense from a population perspective. Such a collective perspective, it needs to be stressed, tends to ignore aspects on the individual level, for instance, individual responsibility for health status.

In order to distinguish grades of health, the perspective of public health needs measures of comparison. In what respect can a person (or group) be healthier than another? What may be the criteria for determining grades of health? This does not allow for a straightforward answer. In the definition of the WHO, for instance, the respective level of health is determined by a subjective state of well-being. This seems difficult to compare between persons or between different states of the same person, though there are now many efforts to turn even happiness into a quantifiable measure (Kahneman 1999). Also, it seems inadequate to call someone healthier merely because he feels better. We know that people can actually feel well and yet suffer from quite severe diseases, especially when they are symptomless. So the criteria for comparing health levels seem more likely to be measures that have to do with the organismic functions of human beings, such as lung capacity, metabolism, memory, or resilience. The more effective these mechanisms function, the more healthy is a person.

To be sure, these challenges regarding the measurement of levels of health are very difficult to surmount. This is because health is such a complex aggregation of

different aspects. For instance, we can try to measure subjectively by assessing the individual quality of life in relation to health aspects or objectively by referring to clinical data and external circumstances of individuals (Coggon 2012, p. 20; Sen 2004). In addition, we can only compare people in certain respects; we can never say whether they are healthier than others tout court (Hausman 2012). Is someone with an irritable lung but a robust psyche less healthy than a marathon runner experiencing bullying at work? Such questions cannot be answered unless we focus on certain aspects of functioning. Public health usually works with only some particular health aspects, such as mental resilience or physical fitness. It also relies on proxies of these criteria, since they cannot easily be directly measured. Accordingly public health collects data, for instance, about frequency of visits to doctors or the number of days on sick leave. Finally, there is a more general problem of collecting data in epidemiology, because it often requires certain abstractions for purposes of generating statistical data. For instance, a common statistical measure for comparing health of certain groups is life expectancy. Obviously here it is not individual health that is measured and compared but a heavily modified proxy for health conditions. This can be particularly significant when politically aiming at certain health outcomes, for instance, when an attempt is made to introduce thresholds of enough health. Any threshold, such as "enough health," relies on a certain "currency," that is, an idea of what aspect of health should be targeted and up to which level it should be accessible for citizens.

Political Dimension of Health in Public Health

As we have seen, the fact that public health allows for grades of health opens the possibility to discuss health promotion in a way that includes enhancing health over and above the absence of disease. This is exactly the area where the worries about "healthism" begin. Health, understood in a positive sense, as in the definition of the WHO, does not have an internal normative stoppage point or threshold of adequate health. More health is always better than less. For egalitarians in the debate on health-care justice, more health might also be required for some groups as a matter of justice, because they are in worse health than other populations not due to their own fault. What is more, according to the public health perspective, improvement of health is not merely, and maybe not even primarily, a matter of improving the internal resources of a person, such as stamina and nutrition, but also of the social determinants of health, such as quality of work environment, access to leisurely activities, and so on. We can accordingly think of many ways to – if only indirectly – improve health dispositions of citizens by improving their living environment as well as changing their lifestyles. So the possible scope for public health interventions is very wide to say the least. If we now add the current value that is attached to health in many societies, we can see how this emphasis on health promotion opens the door for worries about paternalistic interventions, which are even more worrisome if interventions are due to state action and coercive legal measures. One way

to avoid these problems would be to introduce a threshold of "enough" health, hence a sufficient grade of health that every citizen should be able to reach, without overreaching the target of adequate health promotion. But we have seen that it is far from trivial to determine and justify such a threshold.

One way forward is to explicitly acknowledge the political dimension of the notion of health in public health (Weinstock 2011). After all, health here is not simply a scientific notion. It is rather what might be called a functional notion, as its content is driven by public concerns about what we, as a political community, want to publicly secure for every citizen. What we regard as sufficient health within public health is therefore also influenced by theories and beliefs regarding social justice.

The health of the population is a common good. It is a public task to secure it. But it is not determined by the notion of health itself up to which level it ought to be safeguarded. This depends on issues that go far beyond conceptual issues. It is therefore important not to confuse normative aspects of the concept of health with the normative aspects of politically protecting and promoting population health. The fact that public health is both a scientific endeavor and a public policy practice makes it vulnerable to such confusion.

Conclusion

The notion of health within public health is of considerable theoretical and practical significance. There is no agreed definition to be found in the public health literature, yet it is implied that health is here understood as a gradable notion, not simply as the absence of disease. Measurement of health in certain respects and the focus on social determinants of health may lead to confusion of intrinsic and instrumental aspects of health. To have a statistically high liability or a high risk to fall ill is not the same as being in a status of ill health. Comparisons of health status – being healthier than someone else or than another population – are based on specific aspects of health and measurements that are usually proxies for health status. Again, being less healthy does not mean being in ill health. The gradual and the absolute notion of health need to be kept separate.

Public health, in its political purpose, aims at improving health within populations and often also at equalizing health statuses between socioeconomic groups. The value of health is rarely queried, but it needs to be seen in relation to other social values. However, when health becomes an all-encompassing notion via the focus on social and other determinants of health, there is a certain danger of supporting "healthism."

Definition of Key Terms

Collectivism	In the context of this chapter, collectivism is understood as a theoretical position to see society as more than just an assembly of individual

	persons. It may have a normative component, promoting the good of the public.
Grades of health	Health need not be understood as the absence of disease. People can then be deemed healthier than others and also in a certain level of health despite the presence of disease.
Healthism	The promotion of the value of health. Usually, the term is used in a pejorative way, meaning doing too much to promote health, especially using wrong means, such as social control to aim at population health.
Individualism	In the context of this chapter, individualism is understood as a theoretical position to see society as an assembly of individual persons.
Instrumental health	Disposition of a person to get a disease or to stay healthy; also statistically determined risk of health-related outcomes.
Intrinsic health	The health status of a person, either in absolute or gradual terms, but restricted to actual organismic functioning.
Population	In public health populations are statistical measures. They contain individuals combined according to a chosen characteristic, such as females living in a certain area.
Public health	The scientific and political endeavor of preventing disease and promoting health.
Social determinants of health	The aspects of the social environment, such as working conditions, housing, or security, which have an impact on people's health.

Summary Points

- Public health works with a specific notion of health.
- "Health" in public health is a gradable notion, not simply the absence of disease.
- Health dispositions and health risks are not the same as health statuses.
- The concept of health can also be applied to populations.
- The level of health within an individual or population needs to be specified in terms of particular aspects.
- Public health is concerned with environmental, especially social, determinants of health.
- Public health is a political practice as well as a scientific endeavor. In its practical role it promotes the value of health.
- The political aim of public health might lead to worries about supporting "healthism."

References

Arah OA (2009) On the relationship between individual and population health. Med Health Care Philos 12:235–244
Beauchamp DE (1995) Public health: philosophy. In: Post SG (ed) Encyclopedia of bioethics, 3rd edn. MacMillan Reference, New York, pp 2210–2215
Boorse C (1977) Health as a theoretical concept. Philos Sci 44(4):542–573
Broadbent A (2013) Philosophy of epidemiology. Macmillan Publishers, Houndmills
Callahan D (1973) The WHO definition of 'health'. Hast Cent Stud 1(3):77–87
Coggon J (2012) What makes health public? A critical evaluation of moral, legal, and political claims in public health. Cambridge University Press, Cambridge
Daniels N, Kennedy B, Kawachi I (2001) Is inequality bad for our health? Beacon, Boston
Hausman D (2012) Measuring or valuing population health: some conceptual problems. Public Health Ethics 5(3):229–239
Holland S (2014) Public health ethics, 2nd edn. Polity Press, Cambridge
John S (2009) Why 'health' is not a central category for public health policy. J Appl Philos 26 (2):129–143
Kahneman D (1999) Objective happiness. In: Kahneman D, Diener E, Schwarz N (eds) Well-being: the foundations of hedonic psychology. Russell Sage, New York, pp 3–25
Murray CJL, Salomon JA, Mathers CD, Lopez AD (eds) (2002) Summary measures of population health: concepts, ethics, measurement and applications. WHO, Geneva
Nuffield Council on Bioethics (2007) Public health: ethical issues. Cambridge Publishers, Cambridge
Parmet WE (2009) Populations, public health, and the law. Georgetown University Press, Washington, DC
Rose G (1985) Sick individuals and sick populations. Int J Epidemiol 14(1):32–38
Sen A (2004) Health achievement and equity: external and internal perspectives. In: Anand S, Peter F, Sen A (eds) Public health, ethics, and equity. Oxford University Press, Oxford, pp 263–268
Venkatapuram S (2011) Health justice: an argument from the capabilities approach. Polity Press, Cambridge
Weinstock D (2011) How should political philosophers think of health? J Med Philos 36:424–435
World Health Organisation (1986) The Ottawa charter for health promotion http://www.euro.who.int/__data/assets/pdf_file/0004/129532/Ottawa_Charter.pdf?ua=1
World Health Organisation (2010) Adelaide statement on health in all policies http://www.who.int/social_determinants/hiap_statement_who_sa_final.pdf

Identity Disorders: Philosophical Problems 61

Hugh Upton

Contents

Abstract

Any inquiry into identity disorders faces the difficulty that the ordinary understanding of personal identity is itself ambiguous and contentious. In what follows the concept of personal identity that has been of principal philosophical interest is distinguished and clarified, and ideas about the nature of the self are reviewed. The most influential approach to persons and their identity, deriving

H. Upton (✉)
College of Human and Health Sciences, Swansea University, Swansea, Wales, UK
e-mail: h.r.upton@swansea.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_50

from the work of John Locke, is then set out as a basis for reflection on disorders. Varying degrees of disruption to the unity of consciousness are then considered, together with the effect of these on the conception of the self and its continuing identity. Finally, there is a discussion of dissociative identity disorder and of the way in which its conceptualization relates to its status as a disorder.

Introduction

One of the deep problems in philosophy arises from the combination of two very familiar ideas. The first is that the entire world, including ourselves, is one of continual change. The milk curdles, a tree blows down, the curtains fade, and you become fond of the music of Mahler. The second is that things stay the same. The curdled milk is the very milk that earlier someone forgot to put in the fridge, the tree is precisely the one in which we used to climb, those same curtains have been hanging there for years, and it is still you, not someone else, who now enjoys Mahler. It is assumed, then, that physical objects continue through time as the same things, even though they undergo changes in their properties. It is also assumed that each person persists through time as the same person, despite the many changes that naturally happen to people during their lives. In other words, persons retain a personal identity. Yet although this is seemingly a simple claim, it is one that has troubled philosophers for centuries (see Martin and Barresi 2003), raising many difficulties that need addressing before the matter of disorders of identity can be broached.

An example may help in bringing out the special significance of these underlying difficulties. Supposing, for contrast, that this inquiry concerned instead the philosophical issues relating to disorders of the liver, it would hardly be necessary to take account of the possibility that people's livers exist only as some kind of fiction. This, however, has been one conception of the self. Nor can the nature of the self simply be ignored in considering personal identity, since it is not possible to consider what constitutes the continuation of the same thing without some understanding of the kind of thing being referred to in the first place. Then, as further complications, there are different conceptions of what is meant by the identity of a person, and, under any chosen conception, there are different accounts of what mental and/or physical continuity would amount to a continuing identity. Some time must thus be spent in exploring these issues, for any obscurities and ambiguities in the understanding of identity will inevitably reappear in the conception of an identity disorder.

Conceptions of Identity and Identity Disorders

Personal Identity: Numerical and Narrative

The most familiar philosophical problem with personal identity (see Noonan 2003) is this: what is it that constitutes being the same person over time? It may be helpful to clarify the question, though, since an important ambiguity appears immediately.

For illustration, let us suppose that someone is asked: "Are you the same person that you were ten years ago?" After reflection on the events of the last 10 years, the reply might be: "No, I have changed a lot." Here, it should be noted that in saying "I have changed a lot," it is actually implied that something has remained constant: the "I" that has continued through the changes. It is this sense of "I" that has mainly preoccupied philosophical inquiry under the heading of "personal identity." To distinguish it from related areas of inquiry, it is often referred to as an example of "numerical identity," indicating the concern with one and only one person who continues as that person over a period of time.

It is important to be clear how this conception of personal identity differs from certain other matters that may be of related interest. Firstly, the sense of identity outlined here is plainly not that found in the expression "identical twins." This is a matter of qualitative identity, a situation of (nearly) indistinguishable properties rather than of there being just one person. Secondly, and more directly relevant, numerical identity is not what is typically referred to in talking of what gives meaning to people's lives, or how people see themselves, or how others see them. That is, it does not directly relate to those characteristics with which people might, as is often said, "identify themselves," such as being a scientist, having a deep commitment to socialism, or an allegiance to a football team. It is true that such characteristics are often referred to as forming part of a person's identity, but this concept (often called a narrative identity) is not that of numerical identity. To see this, it can be noted that although these characteristics are usually well entrenched, they are nevertheless aspects of a person that might change. As an example, someone who has for many years been an atheist might have become religious a month ago, telling her old friends that she is now a changed person; yet, her friends will still believe that they have known her for more than a month. Further, not only may people deliberately try to change these entrenched aspects of themselves, but their conception of success will be based on the assumption that the person will still be them, before and after the change. For example, suppose that after many years of unpleasant behavior, someone wishes to become a kinder person. In that case, this person's aim is not to disappear altogether and be replaced by someone kinder, but to continue living while becoming more kind. The question of (numerical) personal identity is thus: what is it that makes this continuing self the same one, despite changes even at the level of character or personality? Importantly, whatever it is, it is taken to be something beyond people's control. By contrast, people's narrative identity, involving those deep interests with which they identify ourselves, perhaps thereby giving a sense of unity to their lives, may to some extent be deliberately constructed.

It is worth noting one further point about this concept of numerical personal identity: it seems to be "all-or-nothing" rather than a matter of degree. That is, where numerical identity is concerned, someone is either precisely the same person as 10 years ago or not that person at all. Thus, it would make no sense to say that they were "rather identical" or "somewhat identical," expressions that appear to be in conflict with the linguistic role of this concept of identity. If ever there appears to be a reference to a degree of identity, there is usually an obvious way of resolving

this into some other claim. To recall the earlier example, if twins are described as being "nearly identical," it is not being said that they are very nearly the same person but that they are very similar in appearance.

In what follows the focus will be on numerical rather than narrative identity, not least because of its central place in philosophical inquiry in this area. There is, though, another reason. While there are many problems that may arise with respect to people's narrative understanding of themselves, and while, were these problems to lead to seriously distressing and dysfunctional states for the subject, these states might be regarded as disorders, there would still be some doubt as to whether *identity* disorder would be the best way to characterize them. As an example, someone who had devoted many years of her life to being a writer might become depressed on coming to feel that she had "failed as an author"; yet although the role with which she had identified herself is one source of her problem, it might nevertheless be thought more appropriate to diagnose depression rather than an identity disorder. As another example, it might be felt that someone had developed a damagingly false narrative of himself, seeing himself as leading a life of great significance and entitlements, while regarding others as of no importance, to such an extent that it might prompt a diagnosis of narcissistic personality disorder. As with other personality disorders, this might be taken to involve "problems with one's identity or sense of self" (Butcher et al. 2015, p. 463) without it seeming appropriate to classify it as a disorder of identity. The risk, then, in taking problematic narrative identities as a basis for identity disorders is that the latter category would become too broad in its scope, going well beyond the central concern with the continuity of a single unified self.

Numerical Identity: Fact or Fiction?

Returning, then, to numerical identity, it is necessary to be aware of some of the different theoretical approaches within this conception of the self and personal identity, differences that may have a significant effect on the conception of disorders of identity (Fulford et al. 2006, pp. 761–763). The initial concern will be with just one fundamental distinction.

Firstly, there is the possibility of a realist conception of the self and its continuing personal identity. Underlying this conception is the belief that where people continue as the same person over time, this is in virtue of facts obtaining about them, perhaps simply some fact about a self or perhaps facts about their minds, bodies, or some combination of the two that constitute the self. It typically goes with the assumption that there is a truth to be discovered about personal identity. One particularly important implication of this conception is that it allows a distinction to be made between a person's identity and a person's *sense* of identity. Thus, a patient might be described as suffering a sense of alteration in identity, as reported in some schizophrenic patients (Oyebode 2008, p. 230), where it would be possible for this sense to be correctly contrasted by the clinicians with the fact that they have nevertheless been treating the same patient, before and since the onset of the

disorder. Equally, on this realist conception, it is coherent to regard some conditions (perhaps those arising from dementia or severe amnesia) as providing legitimate grounds for uncertainty among observers regarding the current identity of the patient in front of them.

Secondly, there is what might be called an "anti-realist" position, connected in the history of western philosophy particularly with David Hume, which is skeptical of claims to there being a real self at all and thus, of course, skeptical of there being one with a continuing personal identity. On such a view there is no more to the self and its identity than a subjective awareness of the successive elements of a mental life, together with a (mistaken) supposition that they reflect facts about an actual self and its identity; a supposition induced, in Hume's opinion, by a sense of the resemblance between these elements (Hume 1739/1967, p. 254). Thus, although Hume was addressing the problem of numerical identity, and was proposing an account involving a process that simply happens to us, rather than being a deliberate construction on our part, he concluded that the self and its identity was a kind of fiction. On this basis, then, an "identity disorder" could not strictly be a disorder of the identity of the self (there being no self) but would presumably be a disorder of the mechanism for creating the illusion of an identity of a self. The result would thus be a disruption to the *sense* of having an identity but, since the sense of it is all there can be on this view, the contrast with a *real* identity (a possibility normally implied by this way of speaking) would in fact be lacking.

Self, Identity, and Psychiatry

Thus, one major philosophical issue is to what extent a realist account of the self and its identity can be defended against Hume's skepticism. This is too large a question to try to answer here. Instead, it can be noted that either the realist or the Humean (anti-realist) conception could constitute a framework for understanding references to the self and its disorders. It might of course be argued that the realist conception provides a better foundation for understanding abnormal states, if only because a disordered self sounds more serious than a disordered fictional self. Yet this need not be so: Hume was not denying the great psychological significance of a sense of personal identity, only questioning its basis. Anti-realism, then, does not in itself cast doubt on the seriousness of a disruption to one's sense of self and self-identity.

However, a second issue arises here. Even if the adequacy of anti-realism on this point were accepted, it might still be thought an advantage of a realist account that it usefully extends the conceptual framework beyond the subjective, in the way outlined above. Certainly, it provides the clearest framework for what might be regarded as identity disorders par excellence, those involving the loss not just of a sense of identity but of an actual identity; cases where those other than the sufferer might think that identity had been lost even while personhood continued. Valuable though this might be philosophically, however, some caution is needed, since this conceptual opportunity is not necessarily one that psychiatry will wish to take

up. That is, even if realism regarding the self were to be accepted, it would not follow that the disorders will be conceptualized as disorders of the real self, rather than disorders in the sense of self. It is important, then, to have some awareness of both the identification and description of identity disorders to be found in psychiatry and abnormal psychology.

The current standard text for psychiatric classification is *DSM-5* or, to give its full title, the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association 2013). Here, the relevant general category is "dissociative disorders." These are described as involving "a disruption of and/or discontinuity in the normal integration of consciousness, memory, identity, emotion, perception, body representation, motor control, and behaviour" (p. 291). Of the three main examples of dissociative disorders, only one, "dissociative identity disorder" (DID), is explicitly called an identity disorder, and it will be discussed in detail later. The second, "dissociative amnesia," will also be discussed, on the grounds that although not labeled as such, it seems directly related to personal identity. On the other hand, the third, "depersonalization/derealization disorder," is arguably more directly related to a sense of self and only indirectly to identity, in that depersonalization involves, in various ways, a feeling of detachment from all or part of the self, sometimes involving an "out-of-body" experience.

By way of contrast, the slightly different taxonomy found in a well-known introduction to psychiatry, *Sims' Symptoms in the Mind* (Oyebode 2008), may be noted. Here, the relevant general classification is that of "disorders of the self," this being sub-divided on the basis of five aspects of a person's self-awareness: awareness of existing, of activity, of unity, of identity, and of the boundaries of the self (p. 222). Of these, disorders of identity are said to involve various kinds of discontinuity in the awareness of a continuing identity, though interestingly the category does not include DID, this being classified as a disorder in the awareness of unity rather than of identity. Of the three kinds of condition that it does include, the first (in extreme form) is a sense of having been "completely changed from being one person to another" (p. 230), a condition where even this description of the awareness involved seems deeply puzzling. To recall the opening discussion of numerical identity, if someone has a sense of having *changed into* someone else, this suggests also some sense of continuity between the two states; otherwise, there seems no reason to call it a change rather than simply a coming into existence. If that is correct, might this also be a sense of continuing identity, seemingly in conflict with the description of the condition? The other two kinds of identity disorder in *Sims* are a feeling of possession (included under dissociative identity disorder in *DSM-5*) and that of near death experiences (included, to some extent, under depersonalization in *DSM-5*).

In what follows, rather than attempting to resolve the realist/anti-realist debate, the possibility of greater objectivity mentioned earlier will be kept open by focusing on a realist account of numerical identity as the context for discussing identity disorders. Also, following what seems to be philosophically indicated, the scope will extend to issues of singularity, unity, and continuity of the self that go somewhat beyond the classification of disorders of identity found in *Sims* so as to

include DID. Next, then, it is necessary to give some consideration to the nature of realist accounts of identity.

Are Persons Material Objects?

The issue of personal identity actually raises two questions at once: what exactly is a person and (whatever it is) what makes each of them the same one over time? One possibility is that persons are properly understood as no more than physical, human animals. Is it best, then, to focus on the body as the key to the self and its identity? Certainly, the body seems highly significant, since it is usually the means by which people immediately identify those known to them. Yet the fact that the appearance of the body is typically the way in which people identify each other does not necessarily mean that bodies are persons or are what make persons the same over time. Consider, for example, the use of fingerprints in forensic science. These aspects of the body may uniquely identify each person, but it does not follow that what makes people the same persons over time are their fingerprints. Further, even if bodies were taken to be persons, it must be remembered that the physical matter that constitutes the cells of the body changes completely over a period of a few years. So, if bodily continuity is what constitutes personal identity, the same problem would immediately arise as with objects such as tables, cars, and computers: how can they retain the same identity despite changes in the physical material that constitutes them? For example, if a car is gradually re-built with (some or all) new parts, will it be the same car?

One point about objects that seems to be unavoidable, and which might be unwelcome if transferred to persons, is that it seems unlikely that there is always a truth to be discovered that will settle the question of their identity. In the absence of such truths, it seems entirely appropriate that a decision may have to be made as to the identity of an object, perhaps for a specific purpose such as ensuring the legality of its description for sale. Here, although those who are making the decision will base their judgment on factual matters (on how much has changed or which parts), an assessment will also have to be made of the significance of these changes, something that is not a matter of fact but requires a value judgment. The idea that there may be only this kind of pragmatic or conventional answer may not seem philosophically troubling where objects are concerned, but it may well seem strange as an implication of a conception of personal identity.

Personal Identity and Mind Transfer

There may, of course, be reasons to reject such a reductive physicalist account of persons. It might be said that persons are "souls" (see Quinton 1975), perhaps something that animates a body and maybe survives its death. However, this will probably not help the inquiry much, since, particularly in a secular context, there is too much uncertainty about the supposed nature of souls. Unless it is simply another

way of referring to the mind, talk of them will likely add to the obscurity rather than solve the problem in any informative way. This leaves the possibility that, although dependent upon bodies, what persons actually are resides in some way in their mental life. As mentioned at the outset, this too raises the familiar problem for identity, which is that changes occur precisely where continuity is sought. The mind, or at least its mental life, is something that changes all the time, as old experiences are forgotten and new sensations, ideas, dreams, hopes, and memories appear. Nevertheless, might a person be constituted by this succession? Or, if not the succession itself, might a person be the consciousness that has this succession of mental activity?

At this point, use can be made of an idea from Bernard Williams (1973), who provided an imaginary "test case" to aid reflection on the problems of what persons are and how they continue as the same ones. In brief, everyone is asked to consider this: if your mind was transferred to another body, and that body's mind transferred to yours, where would you be when you woke up after the operation? (As an alternative, this procedure may be imagined as a brain transplant, provided that it is remembered that it is the transfer of the mind that is really at issue in the example, not that of the brain.) Though Williams himself is skeptical about this answer, it is at least plausible that the person and their continuing identity would go with the mind into the new body, in virtue of the transfer of memories and other elements of mental life. Perhaps the friends of the subject would have to accept this too, after their initial puzzlement at the new appearance of the person who gives them a familiar greeting on their arrival at the bedside. In any event, this proposal will be taken as a cue for further inquiries.

John Locke on Persons and Identity

Perhaps the most influential philosophical account of personhood and personal identity derives from material included in the second edition of *An Essay Concerning Human Understanding* by John Locke (1694/1975). It is one that is consistent with the response to the mind transfer case just mentioned, in that it maintains that persons are to be understood in mental terms and that thus personal identity is the identity of a particular mental life.

For Locke, what defines a person is not that it is a certain kind of physical object but that it is a "thinking, intelligent being, that has reason and reflection" (1694/1975, p. 335). In particular, it must be self-aware, conscious of its own thoughts and perceptions, and capable of reflecting on its own experiences. It is an account with some important implications. For example, if the capacity for consciousness is essential to being a person, then the persistent vegetative state (PVS) victim who has no possibility of regaining consciousness would presumably have ceased to be one (see McMahan 2002, pp. 446–447). The requirement for self-awareness and reflection is also significant, since it would have to be accepted that babies would probably not meet this condition and thus would not count as persons. If so, it would

follow that strictly, for Locke, adult persons were never babies, since their personal identity could not be shared with nonpersons. However, at least it may be accepted that there were babies who developed into adult human beings.

Turning, then, to the idea of identity, it will be useful to note that Locke uses the identity of animals by way of contrast with persons and personal identity. The identity of a particular cat, for example, will be a matter of the continuity of a specific cat's body. Given the physical changes that are bound to occur to it, this is not necessarily an entirely straightforward matter but, in any event, the account that will be needed here is one of the continuing identity of a physical object despite the occurrence of changes in that object. Then, since Locke includes human beings (or "man" as he says) under animals, what makes for the continuing identity of a particular human being (as opposed to a person) would in principle be just the same as what makes for the continuing identity of the cat: the continuation of a particular physical body despite the inevitable physical changes.

Locke saw the idea of personal identity in quite different terms. In setting it out, he drew upon one of the essential features of his conception of personhood, that of self-awareness. For Locke, what makes someone the same person over time is not the continuity of their body, or indeed of any substance, but a continuity that relates to a particular kind of self-awareness, that is, the present awareness of earlier experiences. Thus, Locke says, someone is the same person as far back (and no further) as the time of the occurrence of those experiences of which he or she now has memories (Locke 1694/1975, p. 335).

Whether or not it is ultimately accepted, there certainly seems to be something of value in this kind of account. There are also, though, complications that must be considered. Firstly, as Locke was well aware, there are always gaps in people's memories. These relate most obviously to times when they were asleep, but there are also those gaps relating to ordinary forgetfulness about periods in their waking life. However, the gaps are perhaps not too great a problem for Locke's account of identity; as he says, the same consciousness will extend back to remembered earlier experiences, despite any intervening gaps. Secondly, there may be uncertainties as to whether something is a genuine memory or not. If so, these will presumably result in uncertainty about how far back in time someone's identity actually extends. Thirdly, it seems difficult to accept that the natural loss of memories of early childhood will, in effect, reduce the lifespan of a person. However, a modified version of Locke's theory may help here, to the effect that you do not need to be able to recall every earlier experience now, so long as you can recall a time when you could recall them (Noonan 2003, pp. 55–56). In the context of this issue, it is also worth noting that Locke took personal identity to be a "forensic" concept, one that is concerned with the attribution of responsibility. Given that, it is possible to appreciate his reluctance to attribute continuing personal identity (and thus responsibility) to periods and actions that a person simply could not recall. Nevertheless, it does constitute a controversial aspect of Locke's theory. It can be explored further by returning to one of the clinical conditions mentioned earlier, that of memory loss.

Memory Loss and Identity

What should be said about the situation of someone in adult life who suffers serious loss of memory? One example of this might be the gradually worsening memory that is associated with dementia. Here, the discussion is complicated by the fact that the condition of its sufferers may raise questions about Locke's definition of personhood, since the capacity for "reason and reflection" will be a controversial requirement in this context. However, dementia has also been seen as raising questions specifically about the continuing personal identity of the sufferer. One practical manifestation of this (DeGrazia 2005) has been a question concerning the validity of an advance directive: if the earlier experience of writing it can no longer be recalled by the sufferer, is it correct to regard it as having been written by the same person who is now suffering from dementia?

By contrast, in cases of dissociative amnesia, there is no issue regarding loss of personhood but simply an inability to recollect experiences from before the onset of the condition, an inability that may persist for years, during which new experiences are retained in the normal way (Butcher et al. 2015, pp. 298–300). Here, on Locke's account, it would seem that the onset must mark the start of a new person, since the sufferer's current experiential memories go back as far as that time but not earlier. Yet this is bound to be a contentious claim. For one thing it may not accord with the judgment of the sufferer. For another we have to consider the reaction of those people who regard themselves (at least initially) as friends of the victim and call to see her. The face and body will be familiar to them but she, by contrast, will not recognize them. Are they meeting an old friend who has lost her memory, or have they lost their friend and are thus meeting a new person for the first time?

This dilemma brings us back to the question of whether there is a right answer here, a truth to be discovered about her identity. It might be argued instead that there is a decision to be made by the visitors, one that could reasonably go either way. Or, if not exactly a decision, perhaps a pragmatic acceptance of what turns out to seem the more appropriate response over a period of time. Thus, although they might initially be disposed to accept that their friend still exists (despite having forgotten them and everything that has happened earlier) they might simply find this belief impossible to sustain. More generally, as Derek Parfit (1987) has argued, whatever account of personal identity that is adopted, the assumption that there is a determinate answer to all puzzling cases may have to be abandoned.

Successive Selves and Multiple Selves

Having claimed that memory was the basis of personal identity, and having noted the facts of ordinary forgetfulness that result in gaps in memory, Locke discovered an intriguing possibility: he saw that in theory, there could be a succession of different persons in the same body. The form in which he envisaged this was one involving the alternation of "two distinct incommunicable consciousnesses acting the same Body, the one constantly by Day, the other by Night" (Locke 1694/1975,

p. 344). In such a case, where it can be assumed that there would also be two separate sequences of memories, Locke suggests that there would be two persons as distinct (as he puts it) as Socrates and Plato, regardless of the fact that they have a single body in common. In a similar vein, Jennifer Radden (2004) has suggested that even the cycles of bipolar mood disorders such as manic depression might constitute different selves. Perhaps, though, Locke's insight leads further still. If his conception is extended somewhat, so that it may include several more selves, together with the possibility of them existing concurrently rather than only successively, the outcome is the situation familiar from descriptions of dissociative identity disorder. Before reflecting on this condition, however, it is worth recalling some of the usual presuppositions about the unity of the mind and of the self and looking at the extent to which these might be questioned even without the radical possibility of the multiple personalities associated with DID.

The Unity of Consciousness

Just one thought existing in isolation seems to be inconceivable. It seems that thoughts, together with the other elements of a person's mental life, have to be understood as existing with others and as being related to them. However, for a plausible account of a unified self, there is a need for more than just this minimal condition of the relatedness of mental items. After all, the beliefs of different people may very easily be related; for example, one person's belief may be the negation of that of someone else. So an understanding of the self appears to presuppose some further requirement, one that brings thoughts into a closer relation and thus to form the sort of group that is regarded as being in (or perhaps constitutive of) a particular mind. There is a need, in other words, for some conception of the unity of each particular mind and of the distinction between one mind and another. Such a conception is part of a broader and very familiar idea of persons and their identity: that for each human being, there is just one mind and one person.

What then unifies a mind and separates it from other minds? It might naturally be said that these are achieved through its dependence on a particular brain. But even if it were to be accepted that this dependency of a mind on the physical brain is relevant to a general understanding of the mind, there are nevertheless other notions of unity that are important here. In particular, there are some that seem to need describing essentially in mental terms and which thus, arguably, relate more directly to the understanding of the unity of a person's mental life. Above all there is the idea of the unity of consciousness: that for each person there is a single consciousness that has a direct awareness of all that person's thoughts but no direct awareness of those of other minds. Admittedly, neither the ideas of the unity nor of the directness of the awareness are as clear as might be wished; yet the belief that, for each person, there is one continuing awareness of all that person's thoughts as they occur, and that this awareness cannot have the same relation to the thoughts of other people, seems entrenched in human experience.

Just how basic this is to the conception of a person is perhaps best revealed by the difficulties in attempting to abandon it. By way of illustration, schizophrenic patients may report a sense of "thought insertion," the feeling that some of the thoughts they are experiencing are, even at that time, not really their own but have been placed into their minds by others to whom the thoughts still really belong (Oyebode 2008, p. 167). It is a condition where, once again, an account of the symptoms is deeply puzzling, even considered as a delusional state. While a delusion of hearing voices, for example, seems at least to have a comprehensible description, there is by contrast a particular incoherence in supposing that a thought occurring in someone's mind is not now solely that person's own thought, whatever its source and however alien it might seem to the sufferer. This problem with the intelligibility of the symptoms makes it extremely difficult for non-sufferers to have any imaginative grasp of the feeling of an intrusion into the self that matches this description. To recall the earlier distinction, even if disruptions to real selves are left aside, it is a challenge to understand this kind of disruption even as one simply to the sense of self.

It is also worth mentioning two other features that seem important to the idea of a single mind and thus to the idea of a single person. Firstly, there is consistency of belief, to the effect that a person cannot knowingly hold inconsistent beliefs. It cannot be the case that someone genuinely believes some proposition P and also not -P, for example, that here and now it is both raining and not raining. This has implications from the third person point of view as well, in that inconsistent beliefs cannot properly be attributed to another in circumstances where the subject would be aware of the inconsistency. Secondly, there is the idea of a unified will, such that it is assumed that each person has a single "decision center" in the mind which can consider various options before deciding what to do. Thus, even if someone says "I was caught in two minds," what is normally meant is simply that one single decider was finding it difficult to choose between two options.

However, some bodily conditions raise problems for these ideas of unity. Two such conditions will be considered next as a way of exploring the degree of disunity that may be possible in what, at least arguably, remains one mind.

The Split Brain

The first condition derives from attempts to treat epilepsy by a surgical severing of the cerebral commissures, the nerve fibers linking the two cerebral hemispheres of the brain. As a result of this surgery, while there are still links lower down in the brain, the usual direct flow of information between the two hemispheres is lost. One intriguing result (Nagel 1979) was the absence of any immediately obvious effects on the behavior of the patients; only with carefully devised experiments were any effects eventually discovered. To take a single example of many, the right nostril, exclusively, would be exposed to a strong smelling substance, with the effect that the smell would be registered in the right hemisphere, as in normal

cases. However, the patient would deny smelling anything, since the information reaching the right hemisphere could not reach the left, and the left is the one responsible for speech. In contrast to the denial, the patient would show the usual facial signs of detecting a strong smell. Also, from a selection of objects, while still denying smelling anything, the patient would point to the object related to the smell, this being done with the left hand, which was controlled by the well-informed right hemisphere.

On the assumption of the unacceptability in one person of beliefs known to be contradictory, these contrasting responses are problematic. After all, an implication of this assumption is that for any strong and obvious smell, either a person smells something, or they do not. The split brain case presents a challenge to this, in that there seems to be one person who at the same time both smells something and does not smell it. That is, while admittedly there are not two inconsistent statements of belief, there is nevertheless behavior normally clearly indicative of detecting the smell and behavior normally clearly indicative of not detecting the smell. It is thus hard to know how to describe the subject. If it is inappropriate to say that there are two minds here, then the case may at least show that a single functioning mind, and therefore a single self, can be less unified than is usually thought. Perhaps too it is suggestive of an inherent vagueness in our conception of minds. For example, Jonathan Glover (1989, p. 46) argues that these patients do have a divided consciousness but suggests that in counting minds we are dealing with something that has "fuzzy edges."

Alien Hand Syndrome

The second condition may also be found following brain surgery and typically involves one hand seemingly obstructing what someone has decided to do by means of the other hand. This condition is perhaps even more challenging, in that it seems to involve a conflict of decision-making, or of wills, rather than just the question of whether or not a person believes something. Thus, to take one example, a person may light a cigarette and attempt to smoke it, yet find that the other hand has extinguished it, where this extinguishing occurs without the usual feelings of indecision and change of mind that would normally explain such an action. Nevertheless, although the person is puzzled by the act, and feels thwarted by it, the movements of the "alien" hand can hardly be regarded as random. In fact, they appear typical of purposeful behavior, albeit behavior that is rejected as contrary to the person's will and, as it seems to the sufferer, is not owned by them (Gallagher and Vaever 2004). Much as with the split brain example, if the alien hand does not appear to warrant talk of two minds, or two persons, it does seem to challenge the usual conception of their unity. That is, if a person can feel both surprised and thwarted by a seemingly purposeful action, where this action involves his or her own body and has its source within it, this threatens at least the straightforward idea of a person as invariably having a single decision center and a unified will.

Dissociative Identity (Multiple Personality) Disorder

Two kinds of condition have thus been considered, each of which might be thought to cast some doubt on the unity of the mind, whether this is understood as relating to consistency among cognitive states or to the unanimity of the will. With those issues in mind, consideration can now be given to what is perhaps the best known disorder concerning identity, that of dissociative identity disorder (DID), a condition still sometimes called by its earlier name of multiple personality disorder. In *DSM-5* it is described as being "characterized by (a) the presence of two or more distinct personality states or an experience of possession and (b) recurrent episodes of amnesia" (American Psychiatric Association 2013, p. 291). Not surprisingly, the diagnosis is a controversial one in practice (Oyebode 2008, pp. 228–229), and how the condition is even to be conceptualized depends on some fundamental philosophical assumptions about the nature of human beings and persons.

One basis for a conception of DID would be the familiar belief that there can only be a single person per human being. Thus, in these cases, the assumption would be that the patient could only possibly be one person, though a person whose mind had suffered major disruptions. This would be the natural view if, for example, it was thought that persons actually *are* bodies and that our personal identity is constituted by the continuation of the same body. (Though here the precise structure of the body may be critical, since in the case of dicephalic conjoined twins, where much of the body though not the brain is shared, it seems clear that there are two persons.) Certainly, a "one-person" assumption links well with the problems of the split brain and alien hand syndrome just considered, where doubts may be raised about the degree of unity in what might nevertheless be regarded as still one mind and one person. On this approach DID might be taken as simply further evidence of just how great the disunity may be in a single person's mind.

Yet, as was mentioned earlier, there is the option of a more radical conceptualization, one that involves abandoning the belief in an invariable one-one relation between human being and person. Drawing upon Locke's account of personal identity, it seems possible that two or more minds, and thus two or more genuinely distinct persons, may coexist in a single human being. On this basis it could be said that cases of DID present, within one human being, an alternation between different persons, each of whom has his or her own distinct personal identity. Note that it could still be accepted that they all depend on the body for their existence; the claim would be that nevertheless each of them could be a distinct consciousness with distinct memories and thus constitute a distinct person. It is this possibility, to an extent foreseen by Locke, which makes his theory of personal identity particularly relevant to the understanding of DID. But is it a conception that can be accepted? Or, if not, is there a coherent way of understanding DID as a disruption to the mind of what is never more than a single person?

The well-known case of Miss Beauchamp may usefully be taken as an example of DID, one that began in 1898 and was documented at length by the American physician Morton Prince (1978). A few basic elements can be taken from his very

long and detailed account, enough to enable consideration of the philosophical issues that arise from this kind of case. It involves three main personalities (labeled B1, B3, and B4) plus one (B2) of initially rather indeterminate status. The original patient, Christine Beauchamp (or B1) was a quiet, conscientious nurse who consulted Prince when suffering mental health problems after a traumatic incident at her hospital. B2 was not initially thought of as a different person but was just a name for B1's character when under hypnosis. B2 did however have considerable significance, since by the end of the case, when some kind of unity was achieved – initially under hypnosis – B2 came to be thought of as the real Miss Beauchamp, just in need of being "woken up" or brought out of hypnosis (Prince 1978, p. 519). B3, or "Sally," originally appeared when B1 was hypnotized but later appeared spontaneously, that is, without hypnosis. She was lively, carefree, and rather unkind, with a tendency to play tricks on B1. B4 (or the "Idiot," as Sally called her) appeared spontaneously one evening during a visit by Prince to see Christine Beauchamp. The first indication was a change in Miss Beauchamp's demeanor from extreme agitation to calm, which he later realized marked the emergence of a new personality. B4 had suffered amnesia with respect to the previous 6 years, since the traumatic incident, and was confused, rather silent, stubborn, and sometimes aggressive in nature.

Here, as with such cases in general, just one of the alternate persons (if that is what they were) would present themselves and be "in charge" at any one time. The question of the mutual awareness between them, however, was more problematic. The original Miss Beauchamp (B1) had no direct knowledge of B3 or B4, and thus there were simply gaps in her memory for those periods when either of the other two was conscious. Similarly, B4 knew nothing of B1 or B3 and also had memory gaps from those times when the other two were in charge. Most strange of all, though, was the situation of Sally (B3). She was not only aware of the existence of both B1 and B4 but was seemingly aware of B1's thoughts, even though denying ownership of them. Her situation thus involves two deeply puzzling issues (Radden 1996): an asymmetry of awareness between the alternating persons, and the idea, mentioned earlier, of a direct awareness of thoughts that are not your own.

For and Against the Idea of Distinct Persons

To return to the radical conception of DID, is the mental life of the various characters in the Miss Beauchamp case sufficiently distinct for them to count as different persons? Certainly, there do seem to be distinct histories and distinct memories. The histories may not have been very long relative to most lives, but then we do not normally set a minimum length for a life to qualify as that of a person. A more obscure issue, perhaps, is whether it matters that there is an earlier shared history, before the distinct strands appear. Here, there is no normal situation to which we can appeal for guidance, though the ideas explored by Derek Parfit by means of imaginary cases of fission may be relevant here (Parfit 1987). By contrast, J.L. Mackie (1985) drew attention to something more familiar, which is that

different interests and responsibilities are typically associated with different persons. Once allowance is made for the fact that a shared body means that some interests are inevitably shared, Sally and B1, for example, do indeed seem to have different interests. Likewise, it seems natural to hold Sally responsible for her unkindness toward B1, much as would be done if they were quite evidently different people. There are also different capacities: B1 knew French, for example, while Sally did not. Further, B1 had no direct access to Sally's thoughts, a barrier that is one of the crucial features of the usual conception of different minds and different persons. However, matters are complicated here, as has been mentioned earlier, since the direct access that Sally seems to have to the thoughts of B1 is normally (as one aspect of the unity of consciousness) something that counts strongly against the idea of two distinct persons. Yet the fact of the asymmetry, that it is a one-way access between Sally and B1, may perhaps weaken the force of this assumption.

Prince's Conception of the Case

Prince's own presuppositions are intimated by his rejection of the label "multiple personality" for such cases in favor of "disintegrated personality" (Prince 1978, p. 3). In his view there was one and only one real person, the real Miss Beauchamp, to be recovered, and for this reason he thought it appropriate to try to achieve a single unified consciousness. He was aware that this might be contentious, though, and suggested three considerations in defense of his approach to the problem (pp. 231–234). Firstly, he claimed that one of the multiples would be the one best adapted to any environment and that this would be the real Miss Beauchamp. Secondly, he claimed that any other self would be a "sick self," suffering from such conditions as amnesia and poor motivation. Thirdly, he claimed that the real person would be one that was not "artificial," not the product of "special influences."

In response it should perhaps first be acknowledged that the concept of a person is itself contestable (Braude 1991) and has to be applied with caution in a philosophical context. However, even with this proviso, it might be suggested that the first of Prince's points seems clearly questionable as a test for genuine personhood. The ability to adapt and cope successfully with various environments is clearly important but that some (putative) persons are less able than others in this respect is hardly grounds for denying their status as persons. After all, many people may be badly adapted to life yet are unquestionably persons nonetheless. A similar response seems appropriate to the second of Prince's considerations, regarding the sick self: plainly, many people have cognitive and motivational difficulties yet are as much persons as those who are better off in these respects.

The third consideration is rather different. If the splitting into multiples results from a *special* event, does this in itself render the newcomers *artificial* and thus disqualify them from being genuine persons? More generally, can the conception of a "natural" person (see Lizza 1993) be relied upon or has this been undermined by

the reflections on the idea of personal identity discussed earlier? These issues are too complex to follow up in detail here, but the features regarded by Prince as relevant may be briefly considered. Of these, the concept of an event (such as a traumatic experience) as "special" is a difficult category to apply as it stands. It might be understood to mean "abnormal," but even then it would not be clear why an abnormal event would necessarily produce an "artificial" outcome, as opposed to one that was simply unusual. For example, if being born is a normal stage on the way to personhood, a caesarean birth could be regarded as abnormal, yet the resulting baby is not regarded as in any way artificial and nor is it doubted that it is (or will soon become) a person. Perhaps IVF provides an even clearer counter-example. It is a procedure that might reasonably be regarded as special, abnormal, even in some sense unnatural, and artificial, yet (when successful) persons are undoubtedly the eventual outcome. Thus, even if it were supposed that these descriptions applied to traumatic shock, or to any other supposed cause of multiples, it is not obvious that this would provide grounds for the denial of personhood.

Multiples as Real Persons

So far the usual assumption has been made: that if the existence of DID as a phenomenon is accepted, then it must be a disorder. To conclude, though, a different possibility may be considered. Suppose DID is conceived radically as involving genuine multiple persons in one human being, would this situation necessarily be in itself an illness? Arguably, there is no reason to see it in this way, since the appropriate question with respect to health would be whether any of the individual multiples were unwell. Though all of them might share a bodily illness, and any of them might have a mental illness, the mere fact of multiplicity does not seem obviously pathological. To say the least, the presence of multiples might sometimes be awkward, but then so are relations between people in different bodies. Further, if they were held to be genuine persons, the plan to unify them would be conceptually puzzling and (if it were possible to carry it out) open to moral objections. As for suppressing any one of them to enable another to flourish, this too raises obvious ethical issues (Saks and Behnke 2000, pp. 63–66). In fact, if they had problems as individuals or as a group, some form of counseling for each of them might be the most appropriate response.

In general, though, perhaps not surprisingly, there has been a reluctance to accept the possibility of genuine multiple persons in one body. The uncertainty of the whole issue is sometimes the reason for this, rather than any sense that it can be shown to be impossible. Kathleen Wilkes (1988, p. 128), for example, writes that perhaps our concept of a person has "fractured" in the face of DID; meaning, presumably, that we can no longer be sure how to apply it in this context. Yet, as Carol Rovane (1998) argues in her defense of the possibility, the reluctance may be no more than an understandably entrenched way of thinking about persons in general, while Saks and Behnke (2000) regards it as too soon to judge the issue. And it is worth noting that even if there are insufficient grounds for regarding them

as different persons, it is not straightforward to think of them as fragments of a single person either: partly because they each have a reasonable degree of coherence in themselves and partly because, more generally, it is not clear whether we can make sense of the idea of a fragment of a person or indeed of a mind.

Definitions of Key Terms

Narrative identity	Someone's own conception of a meaningful personal history that gives a sense of unity to his or her self.
Numerical identity	This is what holds in virtue of some X being one and the same thing or person. It can hold over time despite changes to X's properties.
Qualitative identity	Being alike in virtue of having the same properties.

Summary Points

- The nature and importance of the distinction between numerical and narrative identity.
- The possibility of skepticism regarding the reality of the self.
- Is personal identity bodily identity?
- John Locke: the self as a thinking being and memory as the key to identity.
- That personal identity might be altered by amnesia.
- The significance of the unity of consciousness in the understanding of the self.
- The problem of split brains and alien hands for the unity of the self.
- Considering dissociative identity disorder from a Lockean standpoint suggests the possibility of genuine multiple persons.

References

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, fifth edition, DSM-5. American Psychiatric Publishing, Washington, DC

Braude SE (1991) First person plural: multiple personality and the philosophy of mind. Routledge, London

Butcher JN, Hooley JM, Mineka S (2015) Abnormal psychology. Pearson Education Limited, Harlow

DeGrazia D (2005) Human identity and bioethics. Cambridge University Press, Cambridge

Fulford KWM, Thornton T, Graham G (2006) Oxford textbook of philosophy and psychiatry. Oxford University Press, Oxford

Gallagher S, Vaever M (2004) Body: disorders of embodiment. In: Radden J (ed) The philosophy of psychiatry: a companion. Oxford University Press, Oxford, pp 118–132

Glover J (1989) I: the philosophy and psychology of personal identity. Penguin, Harmondsworth

Hume D (1739/1967) A treatise of human nature (ed: Selbey-Bigge LA). Clarendon, Oxford

Lizza JP (1993) Multiple personality and personal identity revisited. Br J Philos Sci 44:263–274

Locke J (1694/1975) An essay concerning human understanding (ed: Nidditch PH). Clarendon, Oxford
Mackie JL (1985) Multiple personality. In: Mackie JL (ed) Persons and values. Clarendon, Oxford
Martin R, Barresi J (2003) Personal identity and what matters in survival: an historical overview. In: Martin R, Barresi J (eds) Personal identity. Blackwell, Oxford, pp 1–74
McMahan J (2002) The ethics of killing: problems at the margins of life. Oxford University Press, Oxford
Nagel T (1979) Brain bisection and the unity of consciousness. In: Nagel T (ed) Mortal questions. Cambridge University Press, Cambridge, pp 147–164
Noonan HW (2003) Personal identity, 2nd edn. Routledge, London
Oyebode F (2008) Sims' symptoms in the mind. Saunders Elsevier, Philadelphia
Parfit D (1987) Reasons and persons. Clarendon, Oxford
Prince M (1978) The dissociation of a personality: the hunt for the real Miss Beauchamp. Oxford University Press, Oxford
Quinton A (1975) The soul. In: Perry J (ed) Personal identity. University of California Press, Berkeley/Los Angeles, pp 53–72
Radden J (1996) Divided minds and successive selves: ethical issues in disorders of identity and personality. MIT Press, Cambridge, MA
Radden J (2004) Identity: personal identity, characterization identity, and mental disorder. In: Radden J (ed) The philosophy of psychiatry: a companion. Oxford University Press, Oxford, pp 133–146
Rovane C (1998) The bounds of agency: an essay in revisionary metaphysics. Princeton University Press, Princeton
Saks ER, Behnke SH (2000) Jekyll on trial: multiple personality disorder & criminal law. New York University Press, New York
Wilkes KV (1988) Real people: personal identity without thought experiments. Clarendon, Oxford
Williams B (1973) The self and the future. In: Williams B (ed) Problems of the self. Cambridge University Press, Cambridge, pp 46–63

Personality Disorder: Philosophical Problems

62

Peter Zachar

Contents

Abstract

The concept of personality disorder was introduced in the twentieth century, emerging from a small collection of prior concepts such as constitution, temperament, self, character, and personality. Among the key events in the development of the concept are the introduction and subsequent rejection of degeneration theory, the work of Kurt Schneider, the DSM-III, and the recent proposals to dimensionalize personality disorder in DSM-5 and ICD-11. As the patchwork of ideas that belong to the domain of personality disorder are residues of its conceptual history, that history is herein used to guide an exploration of ongoing philosophical problems. Constitution and temperament raise the issue of the biological basis of personality and personality

P. Zachar (✉)
Department of Psychology, Auburn University Montgomery, Montgomery, AL, USA
e-mail: pzachar@aum.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_77

disorder. Recent work in behavioral genetics supports the hypothesis that personality has a genetic component but not that it is genetically determined. Surprisingly, the genetic component in personality may become more important in open societies where people can self-select into environments. Under the concept of self, the notions of causal explanation and self-continuity are important. Recent work on the psychometrics of latent variable modeling has given new life to traditional empiricist suspicions about reifying personality traits as causal entities. Longitudinal studies indicate that there is both continuity and variability in personality and personality disorder across the life span. Character is a concept drawn from moral theory and draws attention to the close association between some personality disorders and moral vices. In clinical settings, separating responsibility and blame is an important skill for working with patients diagnosed with a personality disorder. The concept of personality once referred to self-presentation but was gradually interiorized, leading to the problem of distinguishing surface versus deep features of personality. This chapter concludes with a survey of six different models regarding the nature of "disorder" in personality disorder.

Introduction: The Historical Development of Personality Disorder

The concept of personality disorder is a child of the twentieth century. Despite having similar names such as mania and melancholia, very few of the categories used in nineteenth-century psychiatry align with current concepts. Throughout the nineteenth century, the psychiatric landscape was expanded by the introduction of new diagnostic concepts. The most important new concepts for personality disorder were *manie sans délire*, monomania, moral insanity, and *folie lucide*. Encompassed under these new diagnoses were compulsions, impulsive acts, overvalued ideas, and rigid affective states – all in the absence of active psychosis. Although personality disorders could be included in this collection, only from our current historical vantage point is "personality" evident.

In the late nineteenth and early twentieth centuries, the psychological concept of personality was emerging out a collection of historically diverse concepts. According to Berrios (1996), these concepts include:

Constitution
Temperament
Self
Character
Personality

Theoreticians in different countries mixed and matched intellectual traditions in such a way that no simple story can be told of how these various strands led to our current conceptions. Indeed, the patchwork of ideas that constitute the concepts of personality and personality disorder are residues of this history.

Four Milestones: Degeneration, Schneider, DSM-III, and DSM-5

The kernel around which the concept of personality disorder developed was degeneration theory. Introduced in 1857 by Benedict Morel, degeneration came to be thought of as a process of de-evolution or a regression to a more primitive stage of development. Once initiated, a trajectory of degeneration was supposedly transmitted to offspring, with each new age group becoming increasingly degenerate. In literature, both Mr. Hyde and Count Dracula were late-nineteenth-century depictions of degeneration. The makeup for Mr. Hyde in the 1931 film starring Frederic March presented Hyde as a Neanderthal. This portrayal was closer to Robert Louis Stevenson's atavistic concept than is seeing Hyde as a manifestation of multiple personality disorder.

By the time Kurt Schneider published *Psychopathic Personalities* in 1923, degeneration was on the wane, somewhat. Both Schneider and Sigmund Freud had rejected it – although it remained influential in the eugenics movement throughout the 1930s, especially in *Nazi* Germany. It was not until the aftermath of World War II when many aspects of *Nazi* ideology were newly considered unacceptable that degeneration theory was abandoned.

Julius Koch's 1891 concept of psychopathic inferiority was formulated, in part, under the auspices of degeneration theory. The same is true of Emil Kraepelin's 1904 notion of the morbid personality. Schneider (1923/1950), however, explicitly claimed that the psychopathic personality was not a degenerate state. Nor was it even "psychopathic" in the current sense of the term. At that point in history, psychopathic was a synonym for "psychological pathology." Schneider viewed psychopathic personalities as statistical abnormalities – either an excess or deficit relative to the mean. When people suffer or they make others suffer because of these abnormalities, the personality can be considered disordered (i.e., psychopathic).

The notion of personality disorder as different in kind from both psychosis and neurosis (mood and anxiety disorders) was clearly articulated in Schneider's book. Soon thereafter the general term psychopathic personality/personality disorder became a center of gravity drawing into its orbit many phenomena that needed a home base in what was a rapidly changing psychiatric landscape. These phenomena included the various formes frustes (milder, incomplete forms) of other mental disorders such as schizoid and cyclothymic disturbances, substance abuse problems, and maladaptive stress reactions.

In the individualist culture of the USA, many psychiatric phenomena came to be considered personality like. Also in the USA, in the middle part of the twentieth century, the concept of personality disorder was augmented by psychoanalytic, neo-Freudian, ego-psychological, and object relation perspectives culminating in the introduction of borderline and narcissistic disorders (Kernberg 1975; Kohut 1971). Hervey Cleckley's (1941) work on psychopathy, which was descriptive rather than psychodynamic, was a parallel line of development.

The third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III), published in 1980, rivals Schneider's book in its historical importance.

The DSM-III gathered together a variety of personality disorder concepts used in American psychiatry and placed them into a single domain. These disorders were grouped into three clusters. The first cluster included paranoid, schizoid, and schizotypal personalities. The second included histrionic, narcissistic, antisocial, and borderline personalities. The third included avoidant, dependent, compulsive, and passive-aggressive personalities.

The DSM-III also highlighted the distinction between personality disorder and other psychiatric disorders by placing them on separate "axes." For instance, in a DSM-III diagnostic formulation, schizophrenia and depression were coded on axis I, while personality disorder was coded on axis II. After the introduction of axis II, personality disorder became a more distinct topic of specialization in psychiatry.

The stability that the domain of personality disorder has possessed since the publication of the DSM-III is increasingly fragile. The committee that developed the DSM-5 of 2013 hoped to introduce a new hybrid model of personality disorder composed of categories and a hierarchy of dimensions. This hybrid model requires that all cases of personality disorder have deficits in two or more dimensions of self and interpersonal functioning. These self and interpersonal dysfunctions are an amalgamation of borderline, narcissistic, and psychopathic features. In addition all cases of personality disorder have to possess one or more pathological personality traits. There are five broad trait dimensions (or *domains*) that can collectively be decomposed in 25 narrower facets. The domains are negative affectivity, detachment, antagonism, disinhibition, and psychoticism. Examples of facets are emotional lability and hostility for negative affectivity and risk taking for disinhibition.

The hybrid model reduces the number of personality disorder categories to six: borderline, narcissistic, antisocial, schizotypal, avoidant, and obsessive-compulsive. The retained categories are all identified by self and interpersonal deficits and profiles of pathological personality traits. Cases of personality disorder that do not fit one of the retained categories are also identified by self and interpersonal deficits and pathological personality traits, potentially making the categories peripheral.

This new model was controversial (Zachar et al. 2016). It was rejected for inclusion in the main text of the manual but was placed in section III of the book and titled *Alternative DSM-5 Model of Personality Disorders*. Its advocates treat it as being in competition with the old DSM categorical model, which was reprinted in section II of the DSM-5.

European and British psychiatrists have been less enthusiastic about the topic of personality disorder than Americans – although more recently the concept of dangerous and severe personality disorder has been important in the UK. The *International Classification of Diseases* (ICD) has mostly utilized the DSM personality disorder categories. The current plans for the ICD-11, however, are to replace all categories of personality disorder with a parsimonious model of five dimensions that will be rated for levels of severity (Tyrer 2014). The five proposed domains are negative emotional (affective), dissocial, disinhibited, anankastic, and detached. Presumably, the ICD approach will compete with the models in sections II and III of the DSM-5.

As might be expected from a conceptually complicated domain such as personality disorder, an exploration of the relevant philosophical problems could fill a handbook of its own. For this reason, selecting which problems to address is likely to provoke disagreement and second guessing. A useful map of the more perennial issues can be found in Berrios' list of the nineteenth-century concepts that preceded the current notion of personality disorder. In what follows, important philosophical and conceptual problems pertaining to constitution and temperament, self as substrate, self-continuity, character, and personality will be explored. In addition, this chapter will address the perennial problem of the nature of "disorder" in personality disorder.

Constitution and Temperament

Constitutional factors refer to innate features that are considered to be biological predispositions for psychiatric disorder. Temperament refers to affective and behavioral dispositions with a genetic component. Temperament in particular encompasses one of the oldest perspectives on psychological types – best exemplified by the humoral theory of Hippocrates and Galen. Their four temperaments were sanguine (extroverted and happy), choleric (energetic and irritable), melancholic (moody and reserved), and phlegmatic (thoughtful and calm).

Temperaments emerge early in life and are relatively stable over time, albeit expressed differently as people become psychologically complex (Kagan et al. 1994). Many different conceptualizations of temperament have been offered in the past 50 years. Among the temperaments considered important for psychopathology are negative affectivity, positive affectivity, and self-control.

According to Lee Anna Clark (2005), temperament is a predisposition for both personality and psychopathology. For instance, the temperament of negative affectivity is a common feature in the personality trait of neuroticism, borderline personality disorder, and major depressive disorder. As a common feature, it partly explains the tendency of neuroticism, borderline personality, and depression to co-occur. Each is also distinct. Neuroticism is a more complex phenomenon than negative affectivity because it also includes self-concepts, motivations, and coping styles. The borderline personality involves specific forms of self-concept, coping styles, etc. Both neuroticism and borderline personality can moderate the course of a depressive disorder.

An important philosophical issue with respect to temperament is its implications for the biological basis of personality and personality disorder. The behavioral genetics of the 1990s initiated a rebirth of biological models of personality as scientific evidence about the high heritabilities of personality traits accumulated (Bouchard and McGue 1990). For instance, even when reared apart, identical twins develop similar political attitudes and interests in religion. Later, the heritabilities of traits like neuroticism were used to argue that such traits are innate, universal features of human design selected for during evolution (Livesley et al. 1998; McCrae and Costa 1997).

Part of the difficulty in conceptualizing the genetic findings is a tendency to conflate high heritability with "highly inherited." They are not the same. Heritability is a technical statistical concept and not a simple proxy for "inherited." To illustrate, for mammals, being an oxygen breather has a heritability of zero, although it is a highly inherited, innate trait. Heritability is a numerical estimate of the percentage of the phenotypic variance that is due to genetic effects. Because all mammals are oxygen breathers, the variance of being an oxygen breather is zero. If the total phenotypic variance is zero, the percentage of that variance due to genetics must also be zero.

By the beginning of the twenty-first century, the optimism of the 1990s, at least among behavioral geneticists, was tempered by additional research (Johnson et al. 2009). As Turkheimer (1998) has pointed out, nearly every psychological trait studied is heritable. This is a strange finding because the ubiquity of heritability also includes time spent watching television and divorce. Time watching television cannot be a universal trait directly selected for during evolution as television was not available prior to the twentieth century. Nor can divorce be said to be inherited rather than acquired.

Additional research also showed that candidate genes for personality traits account for only a small percentage of the variance, and most of these findings do not reliably replicate across studies (McGue and Gottesman 2015). According to current views, there is a genetic basis for personality, but there is no such thing as a *gene for* a personality trait in the same way there is a *gene for* Huntington's disease. The genetic research supports the classical view that there is a genetic component to personality, but not the more modern view that personality is genetically *determined*.

The behavioral genetics of personality is also pertinent to the philosophical problem of freedom versus determinism. In an interesting and unexpected twist, allowing people freedom of choice may increase the extent to which personality traits are heritable in a population. In fact, as populations age, the heritabilities of traits can increase.

How can this be? In an open society, people are able to select the environments to which they are most often exposed. In doing so, they are likely to select environments that allow them to exercise their basic traits and talents (Scarr and McCartney 1983). For example, in the eighth century, people with intellectual interests had limited options. They could work in farming and agriculture, raise children, become a cleric, or practice the art of war. The paucity of available environments constrained the traits that could be developed. In contrast, modern people with intellectual interests have more options to self-select into environments where intellectual traits and abilities can be developed. With self-selection, the traits that are developed are aligned with individual dispositions, resulting in higher heritabilities.

If one's dispositions are maladaptive and the diversity of available environments is large, self-selection might not have such a healthy outcome for everyone. Consider the salesperson who spent a career needing to get along with customers and coworkers despite a disposition toward intense suspiciousness and mistrust. He gets along only with great difficulty and modest success. After retirement, he is

freer to self-select into environments such as those that reinforce conspiracy theories, resentment, and black-and-white thinking. Over time, his suspiciousness can increasingly manifest as something more akin to a paranoid personality style.

The Self

The self with its long and rich philosophical history is a general topic rather than a specific concept. Concepts that fall under the general topic include self-consciousness, self-concept, self-esteem, identity, and personhood. For personality disorders, two key philosophical problems pertaining to the self are the self as substrate and self-continuity.

In the discussion that follows, both (categorical) types and (dimensional) traits will be mentioned. For this reason, some background information on types and traits will be helpful before proceeding.

The classification of personality disorder in psychiatry has tended to utilize types such as borderline, narcissistic, and psychopathic. One of the problems with types is that patients with complicated symptom presentations tend to meet diagnostic criteria for more than one type. For instance, they may be borderline and narcissistic. Generally speaking, mental health professionals believe it is more accurate to say that such patients have a single complicated personality disorder, not two comorbid disorders.

A second problem with types is that there is likely no single, privileged classification of types. Many different personality types exist, only a few of which are formally named. For instance, there were 11 types in the DSM-III. These 11 types do not comprehensively represent the domain of personality disorder as exemplified by the finding that most diagnoses of personality disorder using DSM categories are classified as personality disorder not otherwise specified.

Rather than expand the menu of types, many psychologists and psychiatrists believe that it would be better to develop a comprehensive model of the personality disorder domain with respect to pathological personality traits. Both the alternative DSM-5 model and the proposed ICD-11 model are trait oriented in this way. Ideally, a trait model would be empirically based, rather than founded only on clinical tradition. In addition, personality traits are almost always a matter of degree. For example, neuroticism and suspiciousness are traits on which everyone has a value from low to high. Traits therefore span across the normal and the abnormal. One of the attractions to such a "dimensional model" approach is that it could unify both normal and abnormal personality in a single domain.

Self as Substrate: Basic Issues

The self as substrate refers to the notion that the self is the possessor of psychological features such as perceiving, thinking, and emoting. According to René Descartes, just as length does not exist on its own but is a property of some material object, thoughts and emotions do not exist in on their own; rather, they are properties of the self.

An important contrast to this Cartesian view is that of John Locke who pointed out that "substrate" is inferred to explain why properties co-occur in a regular pattern, but as a concept the substrate is obscure, being little more than a name for "we know not what" supports these observable patterns. David Hume likewise denied that we have any experience of a self, viewing it instead as a speculative inference. Locke and Hume established an empiricist tradition which is wary of making inferences to hidden entities behind the appearances or more specifically wary about attributing metaphysical importance to inferred entities such as the self.

In the seventeenth century, the playwright Moliére mocked the pseudo-explanations used by the physicians of the day by having one of his characters explain the sleep-inducing properties of opium by claiming that opium has a *virtus dormitiva*. Roughly translated, *virtus dormitiva* means the "capacity to induce sleep." Moliére's mocking of the reification of technical phrases into causal entities is a favorite example of empiricists. An analogy in psychiatry would be to claim that a patient has unstable affect because of her borderline personality disorder.

The empiricist perspective has had a profound influence on philosophical thinking about personality traits – particularly the notion that hidden personality traits in the head are psychological causes of behavior. Consider the following:

Feeling tense and jittery
Having a quick temper
Being thin-skinned and readily insulted

In certain personality structures, these features regularly co-occur. One way to explain why they co-occur is to infer an underlying causal disposition – called neuroticism. In addition to explaining why these attributes co-occur, the explanatory construct of neuroticism allows us to make predictions about the presence of additional attributes such as feeling overwhelmed by daily stress.

However, one can ask – is the personality trait of neuroticism a psychological entity that causes people to both feel tense and be overwhelmed by daily stress or is it a general name referring to the regular co-occurrence of feeling tense and overwhelmed? According to the latter position, the name neuroticism is a placeholder for the various causes of these behaviors, but not a cause itself. If we knew what all those causal processes were, there would be no need to ever infer an additional causal process called "neuroticism."

Self as Substrate: Scientific Relevance

Some readers may question the relevance of armchair empiricist metaphysics, but similar considerations have recently gained renewed importance in the scientific study of personality disorder. Let me explain further.

As noted above, advocates for the dimensional approach have been seeking comprehensive, empirically based models of the domain of personality disorder. One of their strategies for doing so is the statistical technique of factor analysis.

Let us consider the 92 symptoms of the eleven DSM-III personality disorders as demarcating the domain of personality disorder. A factor analysis would examine the pattern of correlations between these symptoms and derive a small number of dimensions that statistically explain those correlations. Such dimensions are called latent variables because they are causes hidden in the correlation matrix. The five domain level traits of the alternative DSM-5 model are conceptualized as latent variables in this way.

The problem is that the interpretation of factors as causally potent latent variables may originate in metaphysical assumptions about the self. The important assumption in this case is that personality traits cause behavior. There are two reasons why these assumptions might not be justified in all cases.

First, the "factors" in these statistical models represent individual differences in a population, not causal processes inside the heads of individuals. As Borsboom et al. (2009) have shown, the mathematical requirements for taking individual difference factors derived from between-persons data and treating them as within-persons causal variables are rarely met.

Second, according to van der Maas et al. (2006), artificially constructed data sets in which the correlations between variables are a function of underlying common causes generate the appropriate factors using factor analytic models. However, these factors are also generated by data sets in which the variables are in direct causal relationship with themselves in the absence of underlying common causes (Cramer et al. 2010). For example, the neuroticism factor might be a concept that emerges from a reciprocal causal relationship between

Feeling tense and jittery
Feeling overwhelmed by daily stress

According to this perspective, the pattern named neuroticism refers to an interlocking network of causal relationships that maintains itself over time by means of feedback loops. Rather than *feeling overwhelmed by daily stress* being a surface indicator of a latent entity called neuroticism, it is a part of a pattern of activation in a causal network that we name neuroticism.

Self as Substrate: Situations Versus Traits

The psychologist Walter Mischel (1968) has argued that there is too much variability in behavior across situations to support inferences to causally important traits. If you want to explain why someone is suspicious at one time but not another, according to Mischel, the nature of the situation is a better place to look.

Social psychologists have also discovered that people are too quick to explain an individual's behavior with respect to internal states such as personality traits while ignoring situational influences (Nisbett and Wilson 1977). In some social psychological experiments, people resort to trait explanations even when they know that the situation is the primary causal factor (Jones and Harris 1967).

An example of erroneous trait explanation in psychiatry would be to see most behaviors of a person diagnosed with a personality disorder as expressions of that disorder. For instance, "She is in a bad mood today – there goes her borderline personality disorder again!" Besides the problem of propagating an overly simplistic casual theory, reducing a whole person to a diagnosis is dismissive. It is always prudent to augment explanations with respect to personal qualities by looking for situational influences.

In response to Mischel, advocates for studying personality traits claim that for any particular trait, people will differ on how important that trait is in their personality structure (Bem and Allen 1974). An individual can be suspicious without suspiciousness being a defining trait. Furthermore, for people for whom suspiciousness is central to their personality structure, there will still be variability across situations. To detect the consistency, it is important to aggregate situations and examine trends (Epstein 1979).

If treating personality traits as latent causal entities is justified, doing so requires experience and training. For example, Funder (1997) notes that shy people are often inaccurately judged to be aloof and cold. Shyness and aloofness share many of the same behaviors. Flawed inferences about traits are heightened once we develop conceptual expectations about what another person is like (e.g., she is aloof) and thereafter interpret behaviors in accordance with those concepts (i.e., "Eating lunch alone again? She is awfully aloof."). Rather than using behaviors to confirm trait inferences, it is important to actively differentiate manifestly similar traits by asking the person to report on their likes, dislikes, thoughts, emotions, and perceptions.

Self-Continuity

The problem of continuity versus change is one of the oldest in philosophy. What does it mean for something to change but still be the same thing versus becoming a different kind of thing? With respect to personality, change occurs between age 5 and age 30, between 30 and 55, and between 55 and 80. Some people change more than others.

Is the adult who as an adolescent expressed his shyness by spending most of his time reading alone in his bedroom expressing the same trait when at age 40 he rises early in the morning to have 2 hours of quiet reading? What if that same adult is sociable, interacts with people all day, and teaches courses on interpersonal skills?

Within our large behavioral repertoires, there will always be resemblances between what we were like in the past and what we are like now. Any observer makes a choice about which of these resemblances to call the same "trait." Such choices can be justified, but it is important to not minimize variation and change in order to preferentially see continuity over time.

As a general rule, clinical psychologists are reluctant to make attributions about pathological personality traits until someone reaches late adolescence or early adulthood. One reason for this reluctance is that young people are immature. The normal immaturities of children's and adolescents' personalities would likely be suggestive of personality disorder were they to occur in adults.

A second reason for this reluctance is that personality traits are less stable in our early years (Roberts et al. 2006). It makes sense to expect that if you get to know a

5-year-old girl very well but then you did not see her again until she is 30, you should expect to encounter someone you do really know. If, after getting to know her as an adult, you did not see her again until she was 55, you could reasonably anticipate meeting someone with whom you are familiar.

To an even greater extent, personality disorders are assumed to be fixed, with "inflexibility" being one of the features that make them maladaptive. For instance, the DSM-5 diagnostic criteria for personality disorder include:

An enduring pattern of inner experience and behavior.
The enduring pattern is inflexible and pervasive across a broad range of personal and social situations.
The pattern is stable and of long duration, and its onset can be traced back at least to adolescence or early adulthood.

More recent research, however, has shown that even for personality disorder, continuity versus change is not a simple matter. When personality disorder symptoms are evident in adolescence, if there are also other psychiatric difficulties such as mood, anxiety, and conduct problems, the personality disorder symptoms are more likely to be maintained into adulthood. Otherwise they decrease. This decrease may be correlated with maturation of normal personality traits.

What about the stability of adult personality configurations once diagnostic criteria for a personality disorder have been met? At the beginning of the twenty-first century, data from several longitudinal studies of personality disorder started to become available. According to Morey and Meyer (2012), the early indications were that even in severe cases, pathological symptoms decline over time, and personality disorders are less enduring than psychiatrists had assumed. However, as the time interval in these studies has increased, the picture has become more complicated. Even after a person no longer meets diagnostic criteria for a personality disorder, psychiatric distress and impairment are still evident; and remissions are also common.

An important argument for implementing a dimensional approach to personality disorder, claim Morey and Meyer, is that pathological personality traits such as affective instability and their associated functional impairments are more stable over time than are types such as borderline and narcissistic. The problem with types is that they include features that are stable (traits) and those that are more transient (states). In borderline personality disorder, for instance, affective instability seems to be more enduring, whereas frantic efforts to avoid abandonment are more situation bound.

Character

Prior to the nineteenth century, "character" was a term in moral theories that emphasized virtues and vices. So close was the association between moral theory and character traits that the term moral was often used to denote "psychological."

Considered as traits, virtues are stable dispositions. Examples include benevolence, fairness, and honesty. In the ideal exercise of virtue, a person's cognition,

emotion, and action are coordinated. The virtuous person knows what is good, has good sentiments, and performs good acts.

In this tradition, Peter Goldie (2004) argues that character traits are concerned with a person's moral worth. He also believes that character traits are more important than personality traits because character traits can color all of the personality. For instance, someone can be outgoing, witty, and diligent, but these positive personality traits can be subservient to an all-consuming self-centeredness.

John Sadler (2013) has pointed out that many psychiatric disorders are vice laden, raising the problem of moral taint for diagnostic constructs. More specific to personality disorder, Louis Charland (2004) argues that some personality disorders are moral, not medical conditions. If so, then to have a certain type of personality disorder might be equivalent to being a certain type of bad person.

This problem is most evident for psychopathy and its DSM sibling called antisocial personality disorder in which all seven diagnostic criteria are morally tainted: *lawbreaker*, *deceitful*, *impulsive*, *continually fighting and assaulting others*, *recklessly disregarding others' safety*, *irresponsible*, and *lacking in remorse*. Although not our primary concern in this chapter, the concept of psychopathy may have as much relevance for moral philosophy as it does for personality disorder (Kiehl 2014; Schramme 2014).

Key features of borderline personality disorder, like impulsivity, are also associated with behaviors typically considered immoral such as infidelity. The diagnosis of narcissistic personality disorder is likewise vice laden as many of its clinical features parallel the seven deadly sins, e.g., "grandiosity" is *pride*, "enviousness of others" is *envy*, and "reacting to perceived insults with rage" is *anger*. Paranoid personality disorder and histrionic personality disorder also intersect with the moral realm. People who are paranoid display unjustified resentment and blaming of others. People who are histrionic exhibit a shallow self-centeredness.

This issue of moral taint is equally problematic for the alternative DSM-5 model because the self and interpersonal deficits that are required for *every* diagnosis of personality disorder are largely made up of borderline, narcissistic, and psychopathic features. These deficits include being unconcerned about the effect of one's behavior on others and cooperating predominately for personal gain.

Among the many issues falling under the problem of moral taint for the domain of personality disorder are:

Are personality disorders objective and value free, or must they be value laden?
If value laden, are the values moral values or nonmoral values?
Are failures of moral capacities pathological processes themselves or consequences of deficits in nonmoral capacities?
Under what conditions, if any, does a personality disorder either attenuate or increase responsibility for wrongful acts?

What implications do these philosophical issues have for clinical work? One worry is that emphasizing the moral dimension can mean that blame and

stigmatization are increased if the people diagnosed with certain personality disorders are seen as being either born bad or irreparably bad.

Often, mental health professionals who work with drug and alcohol populations or in prison settings learn to expect deception and persistently insist that clients take responsibility for their actions. Their experience tells them that it would be naïve to hope for virtue rather than expect vice.

This confrontational strategy would seem to conflict with the empathic stance of the general psychotherapist who tries to understand the patient's perspective. However, the consequences of being empathic might not be positively uniform in a therapeutic sense. For instance, one way to achieve empathy is to view behaviors as reason responsive, meaning either (a) from the patient's perspective the behaviors are enacted for reasons or (b) we can understand the behavior with respect to reasons. But if behaviors are seen as being enacted for reasons, then the person is also seen as having a degree of control and responsibility – and therefore as being potentially blameworthy.

Hanna Pickard (2013) observes that the experience of being blamed often has a "sting" that can be anti-therapeutic. The sting, she argues, is related to an emotional form of blame in which one feels entitled to blame the other and believes that the other deserves the blame. Part of clinical training involves learning to manage emotional reactions that might interfere with the professional role. With such training, it is possible to hold patients responsible for their behaviors without engaging in emotional blame.

An important feature of the professional relationship is that it is limited to therapeutic contexts – and the limited nature of that relationship contributes to a clinician's ability to adopt attitudes that would be harder to maintain across all sectors of life. In this vein, clinicians should also be cautious about applying the concept of "personality disorder" outside of clinical settings as it might enhance blaming. Instead of the term "disorder," it is often possible to talk about immaturity instead.

Personality

Of all the concepts in our list, none has undergone a more fundamental transformation than the concept of personality. Our current notion of an individual's personality as those psychological features which (a) make her or him distinct from others and (b) the same over time is of recent origin. In both antiquity and the Medieval period, personality referred to self-presentation – or how we appear to others. In that sense, personality was a name for surface features. Psychological questions about internal features under the auspices of self, character, and soul became increasingly important in the modern era.

For example, Descartes wrote his *Meditations on First Philosophy* from the perspective of the experiencing "I." Half a century later, Locke wrote about "personality" as the awareness of a self-same I extending back in time in his *An Essay Concerning Human Understanding*. Still, the various conceptual strands were

jumbled up. For Locke personality was similar to "personness," referring to a person's awareness of moral agency and responsibility across time.

According to Lombardo and Foschi (2003), it is against the backdrop a more spiritualist and speculative French tradition that the concept of personality began to evolve toward our present notions. A key aspect of this evolution was to understand personality as referring to the unity of the conscious self.

In late nineteenth-century France, with the arrival of positivism and its own suspicion of metaphysics, personality was naturalized. It became a descriptive concept for use in psychiatry and psychology. Among the contributors to this tradition were Eugène Azam, Théodule Ribot, and Pierre Janet – each emphasizing in some way alterations in self-consciousness as forms of *psychopathology*.

It is through William James' (1890) study of French thinking that the concept of personality gained a foothold in the USA. James' concept of personality is not well developed, but it seems to refer to a subjective awareness of self that can vary over time (echoing Locke).

In the first decades of the twentieth century, personality came to be preferred as a secular alternative to character because it lacked strong moral connotations. It was also a psychological alternative to the more biological concept of temperament.

The introduction of personality as a general topic under which the concepts discussed in this chapter were integrated was spearheaded by the psychologist Gordon Allport (1937) and others (Lewin 1935; Murray 1938; Stagner 1937). One of the most important features of Allport's work is that he advocated for the measurement of personality traits, but inspired by his time in Germany studying with William Stern, he also advocated for the importance of the qualitative study of individuals as historically unique (Nicholson 2003). Allport's worry was that restricting our scientific understanding of individuals to the measurement of traits was too shallow.

This tension in Allport's approach continues to exist today in psychiatry with respect to types versus traits. Types of personality disorder such as borderline were initially based on case studies and narratives – which represented a Germanic, qualitative approach to the study of personality. The rich intellectual traditions in the Germanic lands that influenced the concept of personality are too extensive to survey here. Among the various traditions that would need recounting are Immanuel Kant and his heirs, the organicist perspective, the influence of Wilhelm Dilthey, and the study of phenomenology.

Advocates for types state that constructs describing the integration of psychological processes in an individual patient are richer than dimensional profiles which tend to be lists of traits. In their view, coherent types offer bridges to deeper aspects of a personality, whereas dimensions are primarily research tools for identifying relationship among variables in a general population (Shedler et al. 2010). A trait profile might offer an overview of the personality to help initiate a diagnostic formulation, but it is largely a screening instrument.

In contrast, advocates for dimensional trait models believe that types are heuristics constructed from unsystematic clinical observation that lacks the validity of empirically derived dimensions (Livesley 2012). They allow that profiles need to be

augmented with clinical conceptualizations, but in their view, the empirical grounding of traits should lead to better formulations in the long run.

Another manifestation of the contrast between surface versus deep features in psychiatry is the problem of nonconscious influences on behavior. The difficult conceptual issue, especially for scientific psychologists, is the notion of the dynamic unconscious in Sigmund Freud's sense. The Freudian unconscious encompasses impulses, emotions, and affect-tinged representations that are repressed, but that can influence behavior if the repression weakens. These influences are deeper because they were formed early in our development. In addition, because we lack awareness of them, we cannot moderate their influence as repression begins to fail.

The main philosophical problem of the metaphysics of self, i.e., the legitimacy of inferences to unobservable and unexperienced causal entities in the head, remains important. In the history of psychiatry, the unconscious is the ultimate latent variable. This problem is made thornier by the added complication of attributing intentionality and purpose in the absence of awareness. Here are some examples of such inferences:

Psychopaths seek to control others in order to avoid feelings of shame.
Paranoia is a defense against homosexual feelings.
Compulsive behaviors are strategies for undoing an imagined transgression.
Narcissistic grandiosity is rooted in fear of dependency on others

Irrespective of the validity of inferences about unconscious processes, algorithmically applying such attributions to all cases with a particular personality disorder diagnosis is best discouraged. As we saw earlier with the problem of being shy versus aloof, it is hard to appropriately name abstract psychological processes. The danger is that a fallacious version of confirmatory hypotheses testing can be used to transform inferences from behaviors to psychological processes into a conviction that whenever those behaviors occur, the inferred psychological processes are responsible (e.g., "Wow this guy is paranoid. Obviously pathologically unaware of his attractions to other men.").

What the contrast of deep versus shallow ultimately denotes is that the concept of personality helps us understand an individual by looking for patterns that are not immediately apparent. The concept of personality disorder seeks to inform us about patterns that have specific relevance for psychiatric settings and the professional problems of clinicians.

The "Disorder" in Personality Disorder

The problems discussed up to this point in this chapter largely pertain to the personality part of personality disorder. That leaves untouched the question of the nature of personality *disorder*. From its very inception, the concept of personality disorder was different from other psychiatric disorders. Unlike psychosis and many neurotic states, personality disorders are not usually considered afflictions. They are often understood to be ego-syntonic expressions of what someone is like.

An important problem with respect to the validity of personality *disorder* is that these diagnoses are applied to personality styles that people find disagreeable or unlikeable (Saulsman and Page 2004). Indeed, within clinical traditions that emphasize the importance of countertransference, unusually strong feelings toward a patient – including dislike – may be used as a diagnostic indicator of personality disorder.

To say that personality disorder is an appellation for clusters of disliked behaviors seems to make the concept thoroughly subjective. For this reason, it is helpful to have conceptual models that justify including personality in the domain of psychiatric disorder. Zachar and Krueger (2013) describe six different models.

The *vulnerability model* claims that personality disorders are disorders in the same way that essential hypertension is a disorder. Hypertension is a risk factor for heart disease and stroke. Personality disorders are risk factors for the development of other disorders such as depression, panic disorder, and substance abuse. One, however, can cogently argue that vulnerabilities are not disorders.

The *pathoplasticity model* holds that personality disorders are included in the psychiatric domain because they affect the course and outcome of other psychiatric disorders. People with personality disorder develop other psychiatric disorders earlier in life, experience more psychiatric disorders over their lifetimes, and have worse outcomes. Quite often, the diagnosis of a personality disorder is an indicator that a case may have a complicated symptom pattern.

The *spectrum model* claims that personality disorders and other psychiatric disorders share common genetic predispositions. Personality disorders are milder manifestations of those predispositions. The concept of a spectrum refers to the different ways and degrees of severity by which the predispositions can be expressed (Lenzenweger 2006). For instance, schizotypal, schizoid, and paranoid personality disorders have all been considered to be part of a schizophrenic spectrum.

According to the *decline-in-functioning model*, personality disorder symptoms are siblings to the psychological scars that appear in the wake of traumatic brain injury, severe emotional trauma, and severe psychiatric disorder. These "morbid changes" are associated with unambiguous declines in functioning. In cases of psychological scar, the aberrant causal history is known. For personality disorders, the causal history is less certain, but as they share the same pathological symptoms as seen in the trauma-induced cases, they are also considered disorders.

The *impairment-distress model* states that personality disorders are pathological by being directly associated with clinically significant distress or impairment in social, occupational, or other important areas of function. The earliest proponent of this model was Kurt Schneider who viewed personality disorders as statistically abnormal personalities that led to suffering on the part of their bearer. More recent proponents of this model emphasize both distress and impairment (Widiger and Sanderson 1995).

The *capacity failure model* asserts that personality disorders represent dysfunctions in normal, adaptive psychological capacities. These dysfunctions are the underlying pathological processes of any personality disorder. The difficulty with all capacity failure models is that they rely on speculative inferences about normal, healthy functioning.

Both Christopher Boorse (1975) and Jerome Wakefield (1992) advocate for some form of capacity failure model, with an important difference between them being that Wakefield holds that the term "disorder" should be applied only to those dysfunctions that are harmful to their bearer. Livesley's and Jang's (2000) application of Wakefield's harmful dysfunction model to personality disorder was an important inspiration for the self and interpersonal deficits that are part of the alternative DSM-5 model. According to them, these deficits represent failures to find adaptive solutions to universal life tasks.

There are multiple reasons why personality disorders are considered to be clinically relevant in psychiatry. Just as there is no single model of diseases that covers all the things we call disease (tuberculosis, cancer, systemic lupus, essential hypertension, etc.), no single model of disorder currently applies to everything that might be considered a personality disorder. From the standpoint of the medical model, a capacity failure approach may be the most preferable option, but a capacity failure model will need to be justified by auxiliary psychological and social concepts rather than being exclusively a biological or genetic model.

Conclusions

Many important philosophical problems in metaphysics, epistemology, and ethics are relevant to our understanding of personality disorder. Important problems in the philosophies of science, psychology, psychiatry, and medicine are relevant as well. In turn, the phenomena of personality disorder can enrich these philosophical domains. This chapter represents only a small sample of the wealth of material in the domain of personality disorder that is waiting to be explored in future interdisciplinary work.

Definition of Key Terms

Constitutional factors	Innate features that are considered to be biological predispositions for psychiatric disorder.
Dimensional model	A view of psychological and psychiatric traits that views them as being continuous with normality. In a dimensional model, anxiety would be an emotional state on which every person has a value from high to low. An anxiety disorder would occur when the amount of anxiety interferes with normal functioning. The contrast to dimensional model is a categorical model. In a categorical model, an anxiety *disorder* would be discontinuous from a state of normality. For example, posttraumatic stress can be seen as a qualitative change in the structure of one's psychological makeup that is different from excessive anxiety.

Neuroticism	A personality trait exemplified by frequent experiences of negative emotions and related thoughts and perceptions. Negative emotions include anxiety, anger, fear, and sadness. Feeling overwhelmed is another important feature of neuroticism. In contrast, the Freudian concept of neurosis refers to anxiety resulting from unresolved psychological conflict.
Personality disorder	Inflexible personality functioning associated with impairments in social and occupational functioning as a result of disturbances in identity, self-direction, empathy, or intimacy. The impairments are present by early adulthood and are usually chronic.
Psychopathology	A synonym for abnormal psychology, mental disorder, and psychiatric disorder. In the early twentieth century, psychiatric patients were also called psychopaths. Around mid-century, the term psychopath was narrowed to refer to a particular kind of personality disturbance featuring a lack of conscience, a failure to worry, and impulsivity.
Psychosis	A decline in functioning associated with an inability to adapt to the demands of everyday life, often accompanied with a distorted experience of reality. Excessive positive or negative emotions, cognitive disintegration, or misleading sensory experience and beliefs are most commonly associated with psychosis.
Temperament	Early emerging affective and behavioral dispositions with a genetic component.

Summary Points

- The concept of personality disorder was introduced in the twentieth century.
- Research in behavioral genetics supports the view that personality has an important biological component but not that it is biologically determined.
- In open societies in which people are free to self-select into environments, the variance in personality that is attributable to genetics increases.
- There are both philosophical and scientific reasons for viewing personality traits as coherent behavioral patterns (descriptively) rather than as causes of behavior.
- Personality and personality disorders are continuous over time, but with extensive variability.
- Many personality disorder diagnoses are described using moral terms for "vices." The challenge for clinicians is to hold patients appropriately responsible without engaging in emotional blame.

- The twentieth century's transformation of "personality" from something external to something internal introduced a problem about surface features versus deep features that continues to manifest in different ways.
- Personality disorders are psychiatrically relevant for many reasons, but currently there is no single reason that covers all phenomena included in the domain.

References

Allport GW (1937) Personality: a psychological interpretation. Henry Holt, Oxford

Bem DJ, Allen A (1974) On predicting some of the people some of the time: the search for cross-situational consistencies in behavior. Psychol Rev 81(6):506–520. doi:10.1037/h0037130

Berrios GE (1996) The history of mental symptoms. Cambridge University Press, Cambridge, UK

Boorse C (1975) On the distinction between disease and illness. Philos Public Affairs 5:49–68

Borsboom D, Kievit RA, Cervone D, Hood SB (2009) The two disciplines of scientific psychology, or: the disunity of psychology as a working hypothesis. In: Valsiner J, Molenaar PCM, Lyra MCDP, Chaudhary N, Valsiner J, Molenaar PCM, Lyra MCDP, Chaudhary N (eds) Dynamic process methodology in the social and developmental sciences. Springer Science + Business Media, New York, pp 67–97

Bouchard TJ, McGue M (1990) Genetic and rearing environmental influences on adult personality: an analysis of adopted twins reared apart. J Pers 58(1):263–292. doi:10.1111/j.1467-6494.1990.tb00916.x

Charland LC (2004) Moral treatment and the personality disorders. In: Radden J (ed) The philosophy of psychiatry: a companion. Oxford University Press, New York, pp 64–77

Clark LA (2005) Temperament as a unifying basis for personality and psychopathology. J Abnorm Psychol 114(4):505–521

Cleckley HM (1941) The mask of sanity; an attempt to reinterpret the so-called psychopathic personality. Mosby, St. Louis

Cramer AOJ, Waldrop LJ, van der Mass HLJ, Borsboom D (2010) Comorbidity: a network perspective. Behav Brain Sci 33(2–3):137–150

Epstein S (1979) The stability of behavior: I. On predicting most of the people much of the time. J Pers Soc Psychol 37(7):1097–1126. doi:10.1037/0022-3514.37.7.1097

Funder DC (1997) The personality puzzle. W. W. Norton, New York

Goldie P (2004) On personality. Routledge, London

Han L. J. van der Maas, Conor V. Dolan, Raoul P. P. P. Grasman, Jelte M. Wicherts, Hilde M. Huizenga, and Maartje E. J. Raijmakers (2006) A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. Psychol Rev 113:842–861

James W (1890) The principles of psychology. Holt, New York

Johnson W, Turkheimer E, Gottesman II, Bouchard TJ Jr (2009) Beyond heritability: twin studies in behavioral research. Curr Dir Psychol Sci 18(4):217–220. doi:10.1111/j.1467-8721.2009.01639.x

Jones EE, Harris VA (1967) The attribution of attitudes. J Exp Soc Psychol 3(1):1–24

Kagan J, Snidman N, Arcus D, Reznick JS (1994) Galen's prophecy: temperament in human nature. Basic Books, New York

Kernberg OF (1975) Borderline conditions and pathological narcissism. Jason Aronson, New York

Kiehl KA (2014) The psychopath whisperer: the science of those without conscience. Crown Publishers/Random House, New York

Kohut H (1971) The analysis of the self: a systematic psychoanalytic approach to the treatment of narcissistic personality disorders. International Universities Press, New York

Lenzenweger MF (2006) Schizotaxia, schizotypy, and schizophrenia: Paul E. Meehl's blueprint for the experimental psychopathology and genetics of schizophrenia. J Abnorm Psychol 115:195–200

Lewin K (1935) A dynamic theory of personality. McGraw-Hill, New York
Livesley WJ (2012) Tradition versus empiricism in the current DSM-5 proposal for revising the classification of personality disorders. Crim Behav Ment Health 22:81–91
Livesley WJ, Jang KL (2000) Toward an empirically based classification of personality disorder. J Pers Disord 14(2):137–151. doi:10.1521/pedi.2000.14.2.137
Livesley WJ, Jang KL, Vernon PA (1998) Phenotypic and genetic structure of traits delineating personality disorder. Arch Gen Psychiatry 55:941–948
Lombardo GP, Foschi R (2003) The concept of personality in 19-th century French and 20-th century American psychology. Hist Psychol 6(2):123–142
McCrae RR, Costa PT Jr (1997) Personality trait structure as a human universal. Am Psychol 52:509–516
McGue M, Gottesman II (2015) Behavioral genetics. In: Cautin RL, Lilienfeld SO (eds) The encyclopedia of clinical psychology, vol I. Wiley, Hoboken, pp 341–351
Mischel W (1968) Personality and assessment. Wiley, New York
Morey LC, Meyer JK (2012) Course of personality disorder. In: Widiger TA, Widiger TA (eds) The Oxford handbook of personality disorders. Oxford University Press, New York, pp 275–295
Murray HA (1938) Explorations in personality. Oxford University Press, New York
Nicholson IAM (2003) Inventing personality: Gordon Allport and the science of selfhood. American Psychological Association, Washington, DC
Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. Psychol Rev 84(3):231–259. doi:10.1037/0033-295X.84.3.231
Pickard H (2013) Responsibility without blame: philosophical reflections on clinical practice. In: Fulford K, Davies M, Graham G, Sadler JZ, Stanghellini G, Gipps R, Thornton T (eds) The Oxford handbook of philosophy and psychiatry. Oxford University Press, Oxford, pp 1134–1152
Roberts BW, Walton KE, Viechtbauer W (2006) Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. Psychol Bull 132(1):1–25. doi:10.1037/0033-2909.132.1.1
Sadler JZ (2013) Vice and mental disorder. In: Fulford K, Davies M, Graham G, Sadler JZ, Stanghellini G, Gipps R, Thornton T (eds) The Oxford handbook of philosophy and psychiatry. Oxford University Press, Oxford, pp 451–479
Saulsman LM, Page AC (2004) The five-factor model and personality disorder empirical literature: a meta-analytic review. Clin Psychol Rev 23(8):1055–1085
Scarr S, McCartney K (1983) How people make their own environments: a theory of genotype→environment effects. Child Dev 54(2):424–435
Schneider K (1923/1950) Psychopathic personalities (trans: Hamilton MW). Cassell, London
Schramme T (2014) Being amoral: psychopathy and moral incapacity. MIT Press, Cambridge, MA
Shedler J, Beck A, Fonagy P, Gabbard GO, Gunderson J, Kernberg O, ... Westen D (2010) Personality disorders in DSM-5. Am J Psychiatry 167:1026–1028
Stagner R (1937) Psychology of personality. McGraw-Hill, New York
Turkheimer E (1998) Heritability and biological explanation. Psychol Rev 105(4):782–791. doi:10.1037/0033-295x.105.4.782-791
Tyrer P (2014) Time to choose—DSM-5, ICD-11 or both? Arch Psychiatry Psychother 16(3):5–8. doi:10.12740/APP/28380
Wakefield JC (1992) Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder. Psychol Rev 99(2):232–247
Widiger TA, Sanderson CJ (1995) Toward a dimensional model of personality disorders. In: Livesley WJ (ed) The DSM-IV personality disorders. Guilford Press, New York, pp 433–458
Zachar P, Krueger RF (2013) Personality disorder and validity: a history of controversy. In: Fulford K, Davies M, Graham G, Sadler JZ, Stanghellini G, Gipps R, Thornton T (eds) The Oxford handbook of philosophy and psychiatry. Oxford University Press, Oxford, pp 889–910
Zachar P, Krueger RF, Kendler KS (2016) Personality disorder in the DSM-5: an oral history. Psychol Med, 46:1–10

Philosophical Implications of Changes in the Classification of Mental Disorders in DSM-5

63

Andreas Heinz, Eva Friedel, Hans-Peter Krüger, and Carolin Wackerhagen

Contents

Abstract

The new edition of the *Diagnostic and Statistical Manual for Mental Disorders* (DSM-5) by the American Psychiatric Association (Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Association, Washington, DC, 2013) has sparked considerable debate. Allen Frances (Saving normal: an insider's revolt against out-of-control psychiatric diagnosis, DSM-5, Big Pharma, and the medicalization of ordinary life, 1st edn. William Morrow, New York, 2013) and others (Heinz A, Friedel E, Der Nervenarzt 85:571–577, 2014) have argued that this revision may increase the risk to inadequately

A. Heinz (✉) • E. Friedel • C. Wackerhagen
Klinik für Psychiatrie und Psychotherapie Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Berlin, Germany
e-mail: andreas.heinz@charite.de; eva.friedel@charite.de; Carolin.Wackerhagen@charite.de

H.-P. Krüger
Institut für Politische Philosophie und Philosophische Anthropologie, Universität Potsdam, Potsdam, Germany
e-mail: krueghp@uni-potsdam.de

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_76

pathologize socially unwanted behavior and to defocus psychiatric treatment. An undesirable result can be that more severely ill patients will not be adequately provided with services, while an abundance of problems of everyday life in modern societies receives a medical label. This may cause the ambivalent consequence that psychotherapeutic aid can be provided, but that social problems are individualized and isolated from their context instead of being open to social rather than medical or psychotherapeutic interventions. These concerns will be discussed with respect to three topics: firstly, it will be described how the general definition of mental disorders underwent a slight change with nevertheless considerable consequences; secondly, it will be exemplified how a loss of psychopathological traditions and a new definition of core symptoms in schizophrenia together with a lack of consideration of neurological disorders have widened the schizophrenia category to a degree that it may do more harm than good to patients; and thirdly, it will be discussed by way of example how the merging of the previously distinct categories of harmful substance use and substance dependence combines a diagnostically unreliable (harmful substance use) and a diagnostically reliable (substance dependence) clinical category resulting in a socially potentially abusive and poorly defined new category of "substance use disorders." It is argued that the underlying changes would have deserved a more profound discussion of their philosophical as well as social implications.

Introduction

The American Psychiatric Association revised its Diagnostic and statistical manual of mental disorders in 2013 (American Psyciatric Association, 2013). Unlike previous revisions, publishing DSM-5 sparked considerable debate. For example, Alan Frances (2013) critizised that the diagnostic manual fails to focus on the severely mentally ill and instead classifies a multitude of personal and social problems as disorders. Indeed, DSM-5 slightly altered its general definition of mental disorders and revised the required diagnostic criteria for several disorders, including schizophrenia and drug use as well as dependence, which is now relabeled as substance used disorders. In this essay, we discuss the philosophical and anthropological implications of these changes and some of their practical consequences.

Mental Disorders: How Slight Changes Can Have Profound Consequences

Throughout the history of psychiatry, there have been many attempts to define mental disorders. Karl Jaspers, whose work on "*General Psychopathology*" (1946) still has great influence on the field, already observed some 70 years ago that

clinicians are interested in the definitions of specific disorders rather than general concepts of health and disease. In the discussion about mental disorders, philosophers used to have a rather profound impact. For example, Christopher Boorse (1976) suggested that mental disorders are defined by a substantial impairment of mental functions relevant for individual survival. As a consequence, the inability to roll your tongue is not a medical problem, because being able or unable to roll your tongue (in spite of being highly heritable) is generally irrelevant for the survival of human beings. On the other hand, being unable to swallow, for example, due to a stroke or some other neurological disease, impairs a function relevant for individual survival and hence fulfills a criterion for the presence of a disease, namely, the presence of a medically relevant dysfunction.

However, it has been suggested (Sartorius 2011; Heinz 2014) that the presence of a medical dysfunction is not sufficient to diagnose a mental disorder, but that it only represents one aspect (the disease criterion) of a mental malady and has to be accompanied by either personal suffering (the illness criterion) or social impairment (the sickness criterion) in order to be clinically relevant. There are indeed patients who hear voices, thus show a perceptual dysfunction, in this case a hallucination, that can generally be crucial for the survival of human beings and hence fulfills the disease criterion, but these subjects do neither suffer from their hallucinations nor do these perceptual dysfunctions impair their personal functioning in daily life. For example, one of our patients stated that his "voices should be left alone," because he would speculate at the stock exchange and these voices would give him valuable advice, which to date has never been to his disadvantage.

This example highlights the necessity to go beyond the "disease" criterion of mental maladies, which is rightfully defined as an impairment of a mental function generally relevant for survival (see Boorse 1976; Schramme 2000. NB: we disagree with Boorse 1976 and do not suggest that impaired reproduction is a valid criterion for a mental disorder; see Heinz 2014). It suggests to also consider the personal consequences of such dysfunctions for well-being, mainly discussed under the term "illness," as well as the implications for social inclusion and participation, which are generally discussed as the "sickness" aspect of a mental malady (Sartorius 2011). Indeed, medical philosophers such as Charles Culver and Bernard Gert (1982) emphasized that a mental malady is present if a dysfunction is harmful to the individual and either causes suffering or some other state that is undesirable to human beings. Also, Jerome Wakefield (2007) suggests that beyond the medical criterion of any given impairment, individual harm has to be present in order to diagnose a mental disorder. These considerations used to be reflected in the definition of mental disorders as described in DSM-IV, where the American Psychiatric Association stated that a mental disorder "is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that *is associated with* present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom. In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example,

the death of a loved one. Whatever its original cause, it must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual. Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above" (American Psychiatric Association 2000).

Personal suffering, i.e., the illness experience, directly refers to the subjective side of any mental malady. Therefore, a person who suffers from acoustic hallucinations would rightfully be diagnosed with a mental malady because both the disease criterion (impairment of perception that can generally be relevant for survival) as well as the illness criterion (personal suffering) of a mental disorder are present. It has been suggested that a mental malady can also be diagnosed if a person does not personally suffer from his or her dysfunction, but is severely impaired in activities of daily living relevant for social inclusion and participation (the sickness criterion). For example, a patient suffering from Alzheimer's dementia with a clinically relevant dysfunction of memory (the disease criterion) should rightfully be diagnosed with a mental malady if the person is impaired in her ability to take care of herself (the sickness criterion), for example, because hygiene or nutrition is no longer possible, even if she is not aware of her impaired state and does not subjectively suffer from it, and thus, the illness criterion is not fulfilled. Likewise, an alcohol-dependent patient with a delirium during withdrawal will show symptoms fulfilling the disease criterion of a malady (e.g., disorientation and clouding of consciousness) and be absolutely unable to take care of himself with respect to basic activities of daily living (the sickness criterion), while due to a misperception of the current state, he may not subjectively suffer from his experience, may feel ill or in need for treatment, and hence would not fulfill the illness criterion. It is therefore suggested that the presence of medically relevant dysfunction has to be accompanied either by individual suffering or a profound impairment of social participation if a mental malady is rightfully to be diagnosed (see Fig. 1).

This necessity to combine medical, individual, and social aspects in order to diagnose a mental disorder is no longer upheld in the new version of DSM-5. Here, it is now stated that a mental disorder "is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental process underlying mental functioning. Mental disorders are *usually* associated with significant distress or disability in social, occupational, or other important activities. An expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above" (American Psychiatric Association 2013). The new definition thus only suggests that medically relevant dysfunctions are "usually" accompanied by the illness or sickness aspect of a mental malady. Hence, a mental disorder can now already be diagnosed if there are only symptoms relevant within the medical domain

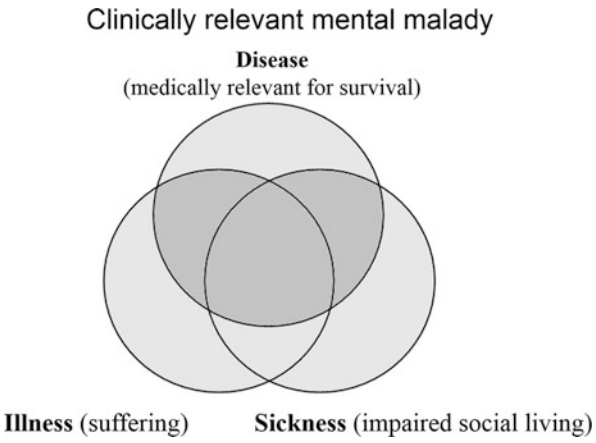


Fig. 1 It is suggested that a clinically relevant mental malady can be diagnosed if (1) medically relevant symptoms of a disease are present, defined as an impairment of functions generally relevant for human survival (disease criterion), plus either (2) such dysfunctions cause individual suffering (illness criterion) or (3) impair activities of daily living (hygiene, nutrition, etc.) and thus social participation (sickness criterion)

(the disease criterion of any disorder). This rather small change in wording has profound implications, because if taken seriously, it abolishes the necessary links between medical, social, and individual aspects of mental disorders and suggests that physicians and psychologists can diagnose a mental disorder solely on the basis of an impairment in one of the multifold functions that can be described in the cognitive or social realm. Accordingly, there is a concern that these changes in DSM-5 can lead to the pathologization of dysfunctions that do not cause subjective suffering or social impairment and thereby jeopardize a philosophically reflected, valid concept of mental disorders with rather profound social consequences.

Dropping Schneiderian First Rank Symptoms in Schizophrenia: The Development of a Dangerously Vague Category

In the 1940s, Kurt Schneider developed his concept of first-rank symptoms of schizophrenia. They include symptoms such as voices arguing or commenting on one's action, withdrawal, insertion or broadcasting of thoughts, made feelings, impulses, and volitional acts as well as delusional perception (Schneider 1942, see also Cutting 2015; Mellor 1970). Some of those symptoms describing alterations such as inserted thoughts have been called "bizarre delusions" in the American tradition, although alterations in the ownership or authorship of thoughts had never been considered a "delusion" in the German psychopathological tradition, because the term "delusion" was limited to a rigidly held, false belief concerning phenomena in the outside world (Spitzer 1988). Here, Schneider (1942) emphasized that psychiatrists tend to label a certain set of beliefs a delusion too easily,

based on their feelings or observations of supposedly "bizarre" behavior of a person rather than investigating the truth content of such beliefs. Schneider therefore suggested to only rely on "delusional perceptions," i.e., the report of certain objective facts that can be assessed by both the examiner and the patient, which are imbued by a specific meaning for the individual patient that is incorrect, yet defended against all evidence by the delusional person. Reports of inserted thoughts, on the other hand, were considered to represent a deep disturbance of inner experience, which in the Kantian tradition (Kant 1986) involves a dysfunction of the "mineness" of thoughts (Mishara et al. 2014). Kant had suggested that perceptions and ideas are unified in consciousness by the act of a perception, the "I think," and that this act by itself is not subjected to a sensual perception (Kant 1986). Unifying procession of thoughts by an active act, which by itself is not represented in consciousness, is then apparently absent in "thoughts inserted by somebody else," which holds a paradoxical position: on the one hand, a person could not communicate that there are "alien" thoughts in her mind if they were to be completely inaccessible ("not her own"); on the other hand, these thoughts are phenomenologically distinct from common experience to a degree that they appear to be "alien" and "inserted" or "controlled" by an outside power ("not authored by the subject") (Sousa and Swiney 2011; Synofzik et al. 2013).

It has been suggested (Sass and Parnas 2003; Heinz et al. 2012) that thought insertion represents an impairment of fundamental aspects of a core self, i.e., a dysfunction of pre-reflective self-awareness. Self-awareness could only be impaired during split seconds or with respect to certain thoughts, because otherwise there would be no possibility for a person to contrast these "alien" thoughts to other ideations that also appear in a person's own stream of consciousness but are obviously not attributed to an outside author (Gallagher 2004; Heinz 2014).

There are two reasons why an impairment of thought authorship, as in the case of passivity phenomena (when thoughts appear to be inserted or controlled by an outside agent), is fundamental for our understanding of psychosis: on the one hand, they interfere with personal agency, thus severely altering a person's position in her "Mitwelt" (Plessner 1975); on the other hand, such first-rank symptoms focus on the articulation of psychotic experiences by the patients themselves instead of relying on the impressions of doctors and other experts with respect to incomprehensible or disorganized psychotic behavior.

With respect to the first aspect, impaired authorship of one's own thoughts may not directly interfere with the ability of a subject to survive in isolation; however, it severely interferes with human interactions and may render it impossible for other persons to know when a certain act carried out by a psychotic person is indeed based on her own considerations or when it results from inserted thoughts or commands of an outside agent that the psychotic person experiences to be controlled by Heinz (2014). With respect to the second aspect, a focus on symptoms reported by the patients themselves is particularly important in light of stigmatization and aggression to which patients with psychotic experiences are repeatedly exposed. For example, when Schneider developed his criteria, schizophrenia patients were subject to compulsory sterilization or even mass murder due to

legislation and procedures enacted by the Nazi government and its psychiatric allies (Klee 2006). Schneider, though not actively resisting these Nazi policies, at least articulated a very cautious procedure toward the diagnosis of schizophrenia, which completely relies on symptoms reported by the patients instead of emphasizing the "impression" that the patient's behavior has on the examiner. As a consequence, disorganized speech or behavior is not a first-rank symptom, because diagnosis of such symptoms is much less reliable due to variation in the judgment of psychiatrists and psychologists.

Schneider (1942) indeed emphasized that expert-based ratings of behavior, i.e., of the "expression" of speech coherence or emotions by any person with psychotic experiences, always make an "impression" on the expert who rates them. If every "expression" is an "impression," rating of such symptoms can largely depend on prejudices and stereotypes that are shared by experts at a certain historical time (remember that Schneider wrote his comments during the Nazi rule). Again, one has to emphasize that even today, treatment of patients with psychotic experiences is often not up to humanitarian standards on a worldwide level. It may be wise to hear Schneider's warning to be cautious when diagnosing schizophrenia and to base this diagnosis mainly on reports of the patients, as impressions by experts are more prone to distortion than the report of inserted thoughts or commenting voices by the patients themselves. Schneider (1942) thus grounded the ability to reliably diagnose schizophrenia on the cooperation of the patients, which – decades before shared decision making or empowerment of patients has been discussed – at least gave some authority to the patients themselves, who have to trust the examiner in order to reveal their experiences.

DSM-5 now abolishes this specific focus on first-rank symptoms for the diagnosis of schizophrenia. While in DSM-IV, one of at least two criterion A symptoms ((1) delusions; (2) hallucinations; (3) disorganized speech; (4) grossly disorganized or catatonic behavior; and (5) negative symptoms, i.e., affective flattening, alogia, or avolition) would have been sufficient if delusions were "bizarre" or hallucinations were commentary or conversing voices; DSM-5 now only requires at least two criterion A symptoms *without any further specification* of the quality of delusions or hallucinations. This is done because a series of studies suggested that first-rank symptoms are not specific for schizophrenia but also appear in other mental conditions, for example, in affective disorders (Carpenter et al. 1973; Andreasen and Carpenter 1993; Shinn et al. 2013; Tandon et al. 2013; Oliva et al. 2014). However, this criticism appears to be controversial, because in the ICD-10, which is based on Schneider's nosological system, the presence of first-rank symptoms suffices to label affective disorders that represent in association with first-rank symptoms as "schizoaffective" (World Health Organization 1992). DSM also provides a schizoaffective category, but requires delusions or hallucinations to show in the course of a predominantly affective episode, during which affective symptoms are mainly absent, while in ICD-10, for the diagnosis of a schizoaffective disorder, affective and first-rank symptoms can be present *concurrently*. It is only by using DSM-IV criteria for affective disorders, which never incorporated the systematic approach of Schneider, that empirical studies can find first-rank symptoms in

"purely" affective disorders. This can be deemed as an example of circular reasoning, in which the assumption that first-rank symptoms are not considered to be specific for schizophrenia is used to prove that they are not specific for schizophrenia (and schizoaffective disorders), while neglecting the fact that their presence (at least for Schneider) would be sufficient to diagnose a schizophrenia and schizophrenia spectrum disorder (ICD-10 F20). Disregarding the specific quality of delusions and hallucinations in the diagnostic process defocuses attention from patient's reports and emphasizes the power of the physician or psychologist to diagnose schizophrenia, even if patients do not report characteristic symptoms.

Medically, this may lead to misdiagnosis: there is a multitude of neurological disorders, including Lupus erythematodes or dopamine-associated psychosis in Parkinson's disease, which manifest with the new key symptoms of schizophrenia in DSM-5 (any kind of hallucinations or delusions). In our own experience, Schneiderian first-rank symptoms do not occur in all schizophrenia patients; however, they are rarely observed in neurological disorders that result in delusional or hallucinatory experiences and thus help to distinguish between schizophrenia and brain organic syndromes (Marneros 1988; Heinz et al. 1995). This line of argument is supported by a recent Cochrane review (Soares-Weiser et al. 2015): when reviewing studies that rely on expert ratings (e.g., using DSM criteria) to diagnose schizophrenia, first-rank symptoms correctly identified most (i.e., 75–95 %) of so-classified schizophrenia patients. Sensitivity of first-rank symptoms, on the other hand, was 60 %; thus, a rather large number of expert-classified patients will not receive a schizophrenia diagnosis. On the negative side, this may cause a delay in proper treatment and suggests that ICD-10 (World Health Organization 1992) is correct in not limiting schizophrenia diagnosis to the presence of first-rank symptoms. On the positive side, regarding first-rank symptoms to be specifically important for the diagnosis of schizophrenia can help to reconsider diagnosis in case of their absence and thus to detect subjects with primary neurological disorders (Marneros 1988; Heinz et al. 1995). From a neurological perspective, the decision to neglect specific psychotic experiences such as first-rank symptoms in the diagnostic process has been made without adequately recognizing the multitude of neurological disorders that can cause psychotic symptoms. Indeed, the presence of such neurological disorders is an exclusion criterion for the diagnosis of schizophrenia in DSM-5, as well as in ICD-10; however, due to budgetary limitations, adequate diagnosis of schizophrenia patients with respect to the multitude of potential neurological disorders that can cause any kind of hallucinations or delusions is not provided in the majority of countries worldwide to date.

Neglecting the diagnostic value of first-rank symptoms did not start with DSM-5; however, DSM-5 finally eliminates any specific focus on them (Tandon et al. 2013). Doing so on the basis of circular reasoning and disregard of neurological disorders appears to be questionable and can undermine the clinical utility of the schizophrenia category. The positive aspect of this development is the demand to reconsider the schizophrenia category altogether. Schneider, in the 1940s, stated that when first-rank symptoms are given, "in all modesty we *speak* of schizophrenia" (Schneider 1942). This statement clearly articulates two important aspects of

Schneider's approach: (1) the function of any diagnostic system for clinical communication rather than the reification of these entities and (2) the cautious approach that any physician should take toward labeling patients with a mental disorder. Both arguments have largely disappeared in the last decades. One consequence of the current neglect of first-rank symptoms in schizophrenia diagnosis may be the potential abolishment of the schizophrenia category altogether, which could be replaced by a label acceptable to patients and relatives as well as clinicians and by diagnostic categories that reflect the necessity to listen to patient's experiences and complaints, as Schneider did some 70 years ago.

Abolishing the Distinction Between Harmful Substance Use and Substance Dependence: How Merging of an Ill-Defined and a Well-Defined Category Can Result in a Fuzzy Concept Prone to Social Abuse

In DSM-5, substance use disorders are a new category that includes the previous categories of harmful substance use and substance dependence (American Psychiatric Association 2013; Hasin et al. 2013; Heinz and Friedel 2014, for a comparison see Table 1). The two concepts were merged because epidemiological studies suggested a gradual rather than a categorical distinction between harmful substance use and dependence (Rumpf and Kiefer 2011; Schacht et al. 2013). Moreover, merging these categories increases the number of substance use patients and therefore the necessity for adequate funding of addiction research. However, the "substance abuse" category was rather ill-defined, because it mainly focused on social problems associated with drug use such as interpersonal problems, impaired social functioning, and conflicts with law (Heinz and Friedel 2014) which depend upon socially and culturally diverse value judgments. For example, social problems increase if the drug is not legal in a given country – then considerable time is required to acquire the drug of abuse, and social consequences including conflicts in the family, job loss, and legal persecution can increase dramatically. Substance dependence, on the other hand, is quite well defined, as long as one keeps in mind Edwards' recommendation stating that tolerance, development, and withdrawal are at the core of the dependence concept (Edwards and Gross 1976). In this view, craving and reduced control are necessary, but additional diagnostic criteria, which should not be used in isolation, as any passion can lead to strong urges and reduced interest in other activities as well as excessive time spent, e.g., in groundbreaking research or other passionate activities (Plessner 2003). Merging the concepts of dependence and harmful use now results in a situation where legally banning alcohol consumption suffices to diagnose a substance use disorder in case a subject wants to consume the illegal drug. These considerations show that it may be quite dangerous to rely too much on diagnostic criteria that are largely depending upon legislation, particularly with respect to addictions.

In DSM-5, pathological gambling has now been classified along with substance-related addictions as an addictive behavior. While it has been agreed on that this

Table 1 A contrasting juxtaposition of DSM-IV and DSM-5 criteria for the diagnoses of harmful substance use and substance dependence and substance use disorder (Modified according to Heinz and Friedel (2014), Hasin et al. (2013))

	DSM IV Abuse		DSM IV Dependence		DSM-5 Substance use disorder	
Hazardous use	X	≥ 1 criterion	-		X	≥2 criteria
Interpersonal problems related to use	X		-		X	
Impaired social functioning related to use	X		-		X	
Conflicts with law	X		-		-	
Wihtdrawel symptoms	-		X	≥3 criteria	X	
Tolerance	-		X		X	
Increasing amounts consumed	-		X		X	
Unsuccessful attempt to quit	-		X		X	
Large amount of time spent for supply/consumption	-		X		X	
Physical/psychological problems related to use	-		X		X	
Narrowing activities on consume	-		X		X	
Craving	-		-		X	

may be useful (Heinz and Friedel 2014) there is an abundance of behaviors that have been suggested to fulfill criteria for behavioral addictions, starting from sex addiction to shopping addiction or even workaholism. While behavior of single cases may well fulfill criteria of behavioral addictions, and it may hence be helpful to classify such behavioral patterns as addictive, it must be cautioned that social pressures and demands can result in political abuse of such addiction categories. Imagine, for example, the power of a state to label an oppositional blogger as an "internet addicted" or the impact of some religious groups on standards of sexual behavior, which could result in an abundance of "sex addiction" diagnoses for behaviors that have just recently (i.e., since the 1960s) been accepted as individual rights of a person in an open society. Moreover, pressures on subjects to not only fulfill work requirements but also to take meticulous care of their own health may result in labeling passions that can interfere with a perfect work performance as

addictions. Continuous pressure to perform in an age of omnipresent electronic devices for communication may thus be individualized as "workaholism" instead of being addressed on a social level. These considerations suggest that a wider discussion is required before more substance-related addictions beyond pathological gambling can be defined.

Conclusion

The current discussion about DSM-5 may miss some of the central points that are indeed worth considering. It is suggested in this chapter that not every psychological or social disorder is rightfully named a disease, but that, instead, the disease term should be limited to impairments relevant for survival or at least for the ability of a person to live with others in the "Mitwelt" (Plessner 1975). Moreover, a clinically relevant mental malady should not be diagnosed if only medical criteria for the presence of a disease are fulfilled, while there is no evidence for individual harm by such symptoms. A point in case are subjects reporting auditory hallucinations, which neither impair their social functioning (i.e., there is no presence of sickness) nor cause any individual suffering (i.e., the illness aspect of a mental malady is not present). It has been suggested (Wakefield 2007) that a mental malady with clinical relevance is not simply given if functions that are generally relevant for individual survival are impaired, as it is the case if there is a perceptual dysfunction resulting in a hallucination, but that in addition, such symptoms of a disease need to either cause suffering in the given individual or impair the person's social participation. This means that evidence for individual harm is required in addition to the presence of disease symptoms to diagnose a clinically relevant mental malady. In contrast to these considerations, in DSM-5 a mental disorder is already to be diagnosed if one out of a multitude of cognitive, affective, and behavioral dysfunctions is present, which do not even have to fulfill the criterion of impairing a function necessary for survival. This development can indeed result in a proliferation of mental disorders with the potential to pathologize socially unwanted behavior and to defocus support in the mental healthcare system from those who are severely ill to subjects suffering from everyday problems.

Furthermore, there is a concern that alterations in key criteria for psychosis may result in mislabeling a multitude of neurological disorders as schizophrenia, because the internal logic of previous schizophrenia classifications is misunderstood and lost in current discussions about key psychotic symptoms. Finally, merging the categories of substance abuse (harmful use) and dependence renders the diagnosis of substance use disorders prone to political interference and abuse. The concept of mental health and its impairment requires discussion on a broader level, including the participation of philosophically informed experts (Boorse 1976; Culver and Gert 1982; Schramme 2000), but also the participation of patients and their relatives in an open dialogue in international settings, as it had occurred and still happens with respect to the United Nations Convention on the Rights of

Persons with Disabilities and its implementation (United Nations 2006; UNBRK 2008; Müller et al. 2012).

Definition of Key Terms

Disease criterion	A medical dysfunction relevant for individual survival
DSM-5 substance use disorder	In DSM-5, all criteria of harmful substance use and substance dependence are joined under the term substance use disorders which are graded in mild (2–3 symptoms), moderate (3–5 symptoms), and severe (6 or more symptoms); social problems are treated equally with symptoms of reduced control, tolerance, and withdrawal.
Harmful substance use	In DSM-IV, harmful substance use is diagnosed if an individual has been using a psychoactive substance for at least 1 year, which was associated with problems in social functioning, hazardous situations, and conflicts with law or was continued although it led to interpersonal conflicts.
Illness criterion	Subjective experience of suffering due to a medical dysfunction
Mental disorder	While in DSM-IV a mental disorder is conceptualized as a behavioral or psychological syndrome *associated with* increased risk of death or pain, personal distress, or social impairment, DSM-5 defines a mental disorder as a clinically significant disturbance of the psychological, biological, or developmental process underlying mental functioning, which is *usually associated* with personal distress or social disabilities. In a philosophically reflected definition, mental disorders can be conceptualized as a medical dysfunction relevant for survival (disease criterion) which is either accompanied with personal suffering (illness criterion) or social impairment (sickness criterion).
Schneiderian first-rank symptoms	These include voices arguing or commenting on one's action, withdrawal, insertion or broadcasting of thought, made feelings,

	made impulses or volitional acts, and delusional perception.
Sickness criterion	Impairment of social (and/or occupational) functioning caused by a medical dysfunction
Substance dependence	In DSM-IV, substance dependence is narrowly tailored to reduced control of substance intake (in ICD-10 also craving) associated with tolerance development and withdrawal symptoms.

Summary Points

- From a philosophical perspective, a clinical relevant mental disorder can be defined as a medical dysfunction generally relevant for individual survival (the disease criterion), accompanied by personal suffering (the illness criterion) and/or social impairment (the sickness criterion).
- Reforms in DSM-5 have diminished the importance of personal suffering and social impairment as diagnostic criteria for mental disorders, which may cause a societally unfavorable distribution of therapeutic care.
- Schneiderian first-rank symptoms focus at the patient's own experience of disrupted personal agency as well as on "delusional perceptions," i.e., interpretations of objective facts that are held in spite of clear evidence to the contrary, which can be assessed by both the examiner and the patient. Relying on first-rank symptoms can be a protective mechanism against misdiagnosis of schizophrenia and stigmatization and can also be crucial for the distinction of schizophrenia from other neurological disorders that manifest with any kinds of hallucinations or delusions.
- Changes in DSM-5 have eliminated the particular importance of first-rank symptoms for the diagnosis of schizophrenia, which can undermine its clinical utility. On the positive side, these changes may open up a discussion to create a new label acceptable to patients, relatives, and clinicians.
- In DSM-5, categories of substance abuse (harmful use) and substance dependence were merged to substance use disorders, which, on the one hand, account for the gradual transition between substance use and dependence, but, on the other, by treating social problems as an equal criterion, make the category prone to political interference and abuse.

References

American Psychiatric Association (2000) Diagnostic and statistical manual of mental disorders, 4th edn (text rev.). American Psychiatric Association, Washington, DC

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Association, Washington, DC

Andreasen NC, Carpenter WT (1993) Diagnosis and classification of schizophrenia. Schizophr Bull 19:199–214

Boorse C (1976) What a theory of mental health should be. J Theory Soc Behav 6:61–84

Carpenter W, Strauss J, Muleh S (1973) Are there pathognomonic symptoms in schizophrenia? An empiric investigation of Schneider's first-rank symptoms. Arch Gen Psychiatry 28:847–852

Culver CM, Gert B (1982) Philosophy in medicine: conceptual and ethical issues in medicine and psychiatry. Oxford University Press, New York

Cutting J (2015) First rank symptoms of schizophrenia: their nature and origin. Hist Psychiatry 26:131–146

Edwards G, Gross MM (1976) Alcohol dependence: provisional description of a clinical syndrome. Br Med J 1:1058–1061

Frances AJ (2013) Saving normal: an insider's revolt against out-of-control psychiatric diagnosis, DSM-5, big pharma, and the medicalization of ordinary life, 1st edn. William Morrow, New York

Gallagher S (2004) Neurocognitive models of schizophrenia: a neurophenomenological critique. Psychopathology 37:8–19

Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, ... Grant BF (2013) DSM-5 criteria for substance use disorders: recommendations and rationale. Am J Psychiatry 170:834–851

Heinz A (2014) Der Begriff der psychischen Krankheit (Auflage: Originalausgabe). Suhrkamp Verlag, Berlin

Heinz A, Friedel E (2014) DSM-5: wichtige Änderungen im Bereich der Suchterkrankungen. Nervenarzt 85:571–577

Heinz A, Przuntek H, Winterer G, Pietzcker A (1995) Clinical aspects and follow-up of dopamine-induced psychoses in continuous dopaminergic therapy and their implications for the dopamine hypothesis of schizophrenic symptoms. Nervenarzt 66:662–669

Heinz A, Bermpohl F, Frank M (2012) Construction and interpretation of self-related function and dysfunction in intercultural psychiatry. Migr Ment Health 27(Suppl 2):S32–S43

Jaspers K (1946) Allgemeine Psychopathologie. Ein Leitfaden für Studierende, Ärzte und Psychologen, 4th edn. Springer, Berlin

Kant I (1986) Kritik der reinen Vernunft. Reclam, Philipp, jun. GmbH, Verlag, Stuttgart

Klee E (2006) Dokumente zur » Euthanasie « im NS-Staat, 6th edn. Fischer Taschenbuch Verlag, Frankfurt am Main

Marneros A (1988) Schizophrenic first-rank symptoms in organic mental disorders. Br J Psychiatry 152:625–628

Mellor CS (1970) First rank symptoms of schizophrenia. Br J Psychiatry 117:15–23

Mishara AL, Lysaker PH, Schwartz MA (2014) Self-disturbances in schizophrenia: history, phenomenology, and relevant findings from research on metacognition. Schizophr Bull 40:5–12

Müller DS, Walter H, Kunze H, Konrad N, Heinz A (2012) Zwangsbehandlungen unter Rechtsunsicherheit. Nervenarzt 83:1150–1155

Oliva F, Dalmotto M, Pirfo E, Furlan PM, Picci RL (2014) A comparison of thought and perception disorders in borderline personality disorder and schizophrenia: psychotic experiences as a reaction to impaired social functioning. BMC Psychiatry 14. doi:10.1186/s12888-014-0239-2

Plessner H (1975) Die Stufen des Organischen und der Mensch. Einleitung in die philosophische Anthropologie (3. Unveränd.). de Gruyter, Berlin

Plessner H (2003) Der kategorische Konjunktiv. Ein Versuch über die Leidenschaft. In: Gesammelte Schriften in zehn Bänden: VIII: Conditio humana. (Vol. 8, pp. 338–352). Frankfurt/M: suhrkamp taschenbuch wissenschaft.

Rumpf H-J, Kiefer F (2011) DSM-5: removal of the distinction between dependence and abuse and the opening for behavioural addictions. Sucht 57:45–48

Sartorius N (2011) Meta effects of classifying mental disorders. In: Regier DA, Narrow WE, Kuhl EA, Kupfer D (eds) The conceptual evolution of DSM 5. American Psychiatric Publishing, Arlington, pp 59–80
Sass LA, Parnas J (2003) Schizophrenia, consciousness, and the self. Schizophr Bull 29:427–444
Schacht JP, Anton RF, Myrick H (2013) Functional neuroimaging studies of alcohol cue reactivity: a quantitative meta-analysis and systematic review. Addict Biol 18:121–133
Schneider K (1942) Psychischer Befund und psychiatrische Diagnose. Thieme, Leipzig
Schramme T (2000) Patienten und Personen: Zum Begriff der psychischen Krankheit. FISCHER Taschenbuch, Frankfurt am Main
Shinn AK, Heckers S, Öngür D (2013) The special treatment of first rank auditory hallucinations and bizarre delusions in the diagnosis of schizophrenia. Schizophr Res 146:17–21
Soares-Weiser K, Maayan N, Bergman H, Davenport C, Kirkham AJ, Grabowski S, Adams CE (2015) First rank symptoms for schizophrenia (Cochrane Diagnostic Test Accuracy Review). Schizophr Bull sbv061
Sousa P, Swiney L (2011) Thought insertion: abnormal sense of thought agency or thought endorsement? Phenomenol Cogn Sci 12:637–654
Spitzer M (1988) Ichstörungen: in search of a theory. In: Spitzer M, Uehlein F, Oepen G (eds) Psychopathology and philosophy. Springer, Berlin/Heidelberg, pp 167–183
Synofzik M, Vosgerau G, Voss M (2013) The experience of agency: an interplay between prediction and postdiction. Front Psychol 4. doi:10.3389/fpsyg.2013.00127
Tandon R, Gaebel W, Barch DM, Bustillo J, Gur RE, Heckers S, ... Carpenter W (2013) Definition and description of schizophrenia in the DSM-5. Schizophr Res 150:3–10
UNBRK (2008) aus un.org: Bundesgesetzblatt, Teil II, 35, Bonn. Gesetz zu dem Übereinkommen der Vereinten Nationen vom 13. Dezember 2006 über die Rechte von Menschen mit Behinderungen. Bonn. Retrieved from http://www.un.org/Depts/german/uebereinkommen/ar61106-dbgbl.pdf
United Nations (2006) Convention on the rights of persons with disabilities. http://www.un.org/disabilities/convention/conventionfull.shtml. Retrieved 6 May 2015
Wakefield JC (2007) The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. World Psychiatry 6:149–156
World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

Part VII

Health as a Social and Political Issue

Medicalization of Social Problems

64

Ashley Frawley

Contents

Abstract

Medicalization is a key concept in sociology, referring to the process by which an increasing array of personal and social phenomena come to be described and understood in medical terms. Concerned primarily with the ways by which social problems are described and defined, constructionist approaches to social problems have utilized medicalization to examine the ways that medical language has been used to describe an increasing array of social problems. Drivers of the proliferation of medical definitions have been identified as the expansion of expertise, the interests of pharmaceutical and biotech companies, and consumerism. Contextual factors include secularization, the growing power of medical and scientific knowledge, the decline of tradition, and the shift of political focus from production to consumption. Though benefits are generally recognized,

A. Frawley (✉)
Department of Public Health, Policy and Social Sciences, Swansea University, Swansea, UK
e-mail: a.frawley@swansea.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_74

medicalization studies usually foreground the process's negative results. Studies utilizing both medicalization and constructionism are subject to general criticisms effecting either approach, including overstating the problem and theoretical inconsistencies.

Introduction

Medicalization is one of the few sociological coinages that have successfully permeated popular vocabularies. Literally meaning "to make medical," it refers to the process by which an increasing array of issues come to be described and understood in medical terms, often through the language of syndromes, diseases, and disorders (Gabe 2013). Its origins can be traced to the mid-twentieth century when many critics began to challenge what they perceived as the rising and potentially pernicious power of the medical profession and of psychiatry in particular. The term gained momentum in the 1970s, particularly through its links with understandings of social control, as social scientists began to describe apparent shifts in the means by which Anglo-American societies were defining and disciplining deviance. Since then, medicalization has been used to understand how a wide variety of phenomena have come to be considered medical issues including pregnancy, childbirth, alcoholism, obesity, sleep, educational underachievement, madness, drug addiction, and death.

The focus of medicalization has encompassed a wide variety of phenomena, which can be conceptualized as falling into two broad categories: (1) the medical redefinition of, and control over, hitherto unproblematic or at least common aspects of everyday life including pregnancy, birth, and death and (2) the medicalization of deviance or the process through which nonnormative or morally condemned attributes, beliefs, or behaviors come under medical jurisdiction (Mcgann and Conrad 2011). This chapter attempts to focus more acutely upon the latter category. In particular, it focuses on that form of deviance, or the transgression of social norms, that sociologists and other social scientists mostly espousing a social constructionist outlook study under the rubric of "social problems." From this perspective, the medicalization of social problems refers to the ways by which medical categories have come to be applied and increasingly accepted as one of the dominant means through which modern societies define and approach their problems. In other words, it is the process of defining "troubling conditions" as medical problems (Best 2008, p. 99).

After exploring the origins and definitions of medicalization and associated terminologies, this chapter moves to a consideration of the use of medicalization to understand social problems. Potential underlying drivers of the proliferation of medical definitions are considered as well as the consequences of this process for individuals and society. The final section considers general criticisms that have been launched against both medicalization and constructionist approaches to social problems that are relevant to studies combining the two.

Emergence of Medicalization Thesis

The twentieth century saw a rise in the stature of the medical profession, encouraging its gradual expansion into, and authority over, growing domains of individual and social life. Until the mid-twentieth century, this authority largely went unchallenged (Cockerham and Ritchley 1997). However, early analyses and critiques of medical power began to emerge as early as the 1950s. Talcott Parsons' notion of the "sick role" may be read as one of the first examinations of medicine as a form of social control for the way in which it defined, legitimated, and contained illness as a deviation from "wellness" as a social norm (Conrad 2007, p. 51). Though not necessarily employing the terminology of medicalization, theorists writing in the 1960s and early 1970s including Thomas Szasz (1961, 1970), Michel Foucault (1965), and Eliot Freidson (1970) began to forward critical analyses of the broadening scope of the medical profession, and psychiatry in particular, in its power to define and control deviance once considered in moral, religious, or criminal terms. For Thomas Szasz, the movement of "madness" into the realm of medicine effectively reified a social and moral category, unduly removing it from the cultural sphere and placing it in the realm of the biological. This transformation did not necessarily represent an enlightened progression toward more humane treatment of those once seen as witches or possessed, but rather served to obscure the social origins of human difference and distress.

At the same time, the growing feminist movement began to question the role of male doctors and a male-dominated health system in controlling women's bodies and reproductive health. Women began to expose and challenge the tendency for their experiences to be treated as forms of illness and deviations from an implied male norm (Nettleton 2013). Medicine was accused of attempting to discipline the "unruly" female body, and interventions into pregnancy and birth were criticized as unnecessary or even harmful interferences. Beginning in the late 1960s, organizations like the Boston Women's Collective and the Jane Collective variously encouraged women to take back control of their bodies, their sexuality, and their reproductive health.

Also exploring the growing power of medicine and its role in social control, Irving Zola drew attention to the ways by which medicine was "nudging aside, if not incorporating, the more traditional institutions of religion and law" in the regulation of everyday life (1972, p. 487). In one of the earliest explicit articulations of the medicalization thesis, he worried that not just childbirth, but nearly everything from "sex to food, from aspirins to clothes, from driving your car to riding the surf," had become associated with health and health risks, famously concluding that "I at least have finally been convinced that living is injurious to health" (Zola 1972, p. 498). While building his description of medicalization on earlier critiques of psychiatry, Zola thought this myopic focus on one form of medical knowledge was misplaced. Psychiatry had simply pushed to its logical conclusions the task long taken up by the medical profession as a whole: the control and containment of deviance and the disciplining of daily life.

Writing almost simultaneously with Zola, Ivan Illich argued the success of the medical profession had been overstated and that it often did more harm than good. Using instead the term "iatrogenesis," literally referring to harm "brought forth by the healer," he pointed to medicine's tendency to undermine people's capacity for self-care, paralyzing healthy responses to life processes and encouraging a dependence upon professionals. He also described the ways by which the side effects of medical interventions could sometimes outweigh the harm wrought by the initial condition. Indeed, Illich himself refused to have a large facial tumor removed since doing so could remove his ability to speak. His critique of modern medicine formed part of his broader questioning of the fruits of industrial progress and development. People would revolt, Illich argued, if medicine did not exist to tell people that their problems lay primarily in their bodies rather than the outside world (Sheaff 2005).

Thus, the medicalization thesis emerged from critiques of medical dominance and described the ways by which medicine's power was mobilized to control and discipline behaviors, bodies, and beliefs society deemed dangerous, unruly, or deviant. Although not always or initially using the terminology of medicalization, these early analyses all essentially drew attention to the growing use of medical language to describe human experiences and problems in the twentieth century.

Medicalization and Definitions

While many aspects of medicalization remain open to debate, it is clear that the key to the process lies in language or the way social phenomena are *defined* at a given time (Conrad 2013). Medicalization studies focus on the application of a medical framework, sometimes referred to as a medical model, to a wide range of phenomena previously considered the domain of religion, culture, criminality, or other nonmedical structures (Best 2008). While sometimes restricted to the explicit application of a diagnosis to new areas, in its broadest sense, medicalization refers to the process whereby individual and social life comes to be talked about and understood through the use of medical language *of any kind*. What is important is a definitional and rhetorical shift in the way that people talk about a given phenomenon from one sociohistorical moment to the next.

This means that although the profession of medicine is often pointed to as the source of medicalization, sometimes expressed through a critique of "medical imperialism," the medical profession itself need not play a key role. It is the power of medical language, rather than the medical profession itself, that is most significant and to which more recent critiques have pointed. There are numerous cases in which doctors have been uninvolved or even opposed to claims aiming to extend medical definitions to new domains. A famous example is the movement to define alcoholism as a disease which was led not by physicians but rather by individuals and groups associated with the Alcoholics Anonymous program and only loosely followed a strict biomedical model of disease (Best 2008). Indeed, there is some evidence that far from achieving unparalleled dominance, the medical profession is experiencing a process of deprofessionalization as other institutions

challenge its authority. Deprofessionalization refers to a "decline in professional status and power resulting from a deterioration in those characteristics which distinguish professions from other occupations, especially the loss of autonomy over work and control over clients" (Cockerham and Ritchley 1997, p. 31). State regulation, managerial surveillance, suspicion toward doctors, the rise of litigation, the growth of alternative and complementary medicine, and the professionalization of other occupational groups including nurses have all been pointed to as sources of a decline in the medical profession's dominance (Furedi 2008). Thus, the expansion of medicalization is increasingly overseen by a wide variety of experts and professionals, advocacy groups, and even laypeople.

The growing use of medical language to understand new domains and new problems may be considered as part of a broader expansion of scientific knowledge in general into nearly all aspects of life. As Best (2008) describes, medical knowledge is merely a subcategory of the larger domain of scientific expertise. In the same way that medicine saw an expansion of authority due to prominent successes in the nineteenth and twentieth centuries, so too did staggering advancements made possible by science contribute to its cultural authority (Best 2008). This rise in the authority of science also oversaw or at least coincided with a decline in the ability for older structures like tradition and religion to lend meaning to everyday life and legitimacy to pronouncements. Although these structures have not entirely disappeared, it is difficult to argue that the voice of the priest and language of sin continue to hold the same power as they might have done even a century ago. At least in the West, audiences are more likely to be convinced by claims communicated in the language of diseases, syndromes, and disorders – "words that seem more grounded in medical, scientific classifications" (Best 2008, p. 97–98).

The increasing significance of science and new technologies has led to the coinage of biomedicalization as a terminology that more acutely grasps ways by which medicalization has developed and intensified in the twenty-first century. As Clarke and Shim (2011) describe, biomedicalization refers not only to medical control over phenomena but their actual transformation by technoscientific means. That is, the emphasis is shifting more toward scientific and technological interventions that actually change bodies and identities (Clarke and Shim 2011). These high-tech interventions often promise not only to treat or cure but also to enhance and optimize. While medicalization has continued, biomedicalization draws attention to the increased centrality of science and technology in the process. For instance, the past few decades have seen a revolution in genetics, which has led to claims that it will not be long before genetic markers will be found to underlie nearly all personal and social problems (Best 2008). As with medicalization, the biomedicalization of life encourages the view that it is biology that ultimately underlies almost every trouble and that it is toward biomedical advancements that one must turn if solutions are to be found.

In a similar way, psychologization has been suggested as a terminology that more acutely grasps the inclination within medicalization toward the discovery of new psychiatric categories and the expansion of old ones, so that more and more of individual and social life comes to be encompassed by psychiatric labels. It also

captures the ways by which social problems are often conceptualized as stemming from problems at the emotional and psychological level. As Furedi (2008) describes, emotional and therapeutic terms like "stress, rage, trauma, low self-esteem, or addiction" increasingly offer up quasi-medical labels for interpreting virtually any human experience or issue. However, in spite of the potential specificity offered by terminologies like psychologization and biomedicalization – and indeed the many other "izations" that have since emerged (see Conrad 2013) – medicalization continues to be the most popular term utilized by social scientists and has developed to encapsulate many of these and other trends.

Finally, it is important to note that medicalization is not a straightforward process toward a foregone conclusion. Rather, the degree to which a phenomenon is considered an entirely medical category may not be total, and medicalization of a phenomenon may exist at many stages of cultural recognition and affirmation. It can also operate in reverse, a process referred to as "demedicalization." Some of the most oft-cited examples include the waning concern of medical professionals with masturbation and the declassification of homosexuality as a mental disorder (Mcgann and Conrad 2011). By the same token, categories can be "re-medicalized." Indeed, while homosexuality remains the most commonly cited example of a successful demedicalization campaign, it may be experiencing a process of re-medicalization through, for example, prominent associations with HIV/AIDS, discourses of genetically predetermined sexuality and searches for a "gay gene," and the introduction of "gender identity disorder" to the DSM (Conrad 2007). Although the overarching trend appears to be toward increasing medicalization, it is important to recognize that it is not a linear process toward a static end point (Gabe 2013).

Medicalization of Social Problems

The study of the medicalization of social problems has both contributed to and grown from broader medicalization debates discussed thus far. The first medicalization theorists were influenced by a variety of theoretical traditions; Zola and Freidson took inspiration from labeling theory and the then fledgling social constructionism, drawing attention to medicine's ability to define what illness is and to label as illness that which was not previously labeled (Busfield 2006). From the perspective of labeling theory, deviance does not exist within the people or actions themselves, but rather in the act of assigning a label. Similarly, social constructionism (also called constructivism) argues that ideas are not simple one-to-one reflections of objective reality, but are socially and historically conditioned. From infancy, people learn to assign names and categories to the world in order to understand it and from which they are able to attribute meaning to experiences (Best 2008). These names and categories in turn influence the ways by which people understand themselves and the world around them. Consider, for example, that although nearly every human society assigns gender classifications, the meanings that these classifications carry, and the implications for the identities and

behaviors of the people to which they are attached, represent enormous cultural variation (Best 2008). Thus, what many may take for granted as a straightforward biological classification represents a complex process of meaning construction.

Although not all examinations of medicalization adopt a constructionist framework (Nettleton 2013), the thesis was adopted and extensively developed within the constructionist literature specifically pertaining to social problems beginning in the early 1980s. While social problems can be understood in a number of ways, the constructionist orientation adopts a subjectivist rather than objectivist orientation to their study. From an objectivist perspective, social problems are simply harmful conditions that affect society in some way (Goode and Ben-Yehuda 2009). By contrast, a subjectivist orientation views social problems in terms of people's subjective judgments about whether or not something is troubling (Best 2008). For instance, it may seem relatively uncomplicated that a condition such as racism or sexism should represent a problem for society. Rationales for defining these as problematic may rest on threats to well-being or violations of fairness or justice (Best 2008). However, such definitions seem less straightforward when one considers other imaginable foci for critiques of discrimination or unfairness that go unnoticed or unlabeled by society. As Best (2008) describes, while there is considerable evidence of discrimination based on height, which might also be considered unfair and detrimental to well-being, "heightism" as a social problem largely does not exist. If people do not think it is a problem, it will not become a social problem with the attendant societal recognition that both affirms it and seeks its amelioration.

As with medicalization, it is the definitional process that is key to understanding how social problems come into being. From this perspective, social problems are processes through which individuals and groups develop definitions and ways of understanding conditions they find problematic and work to bring these to the attention of the broader public, policymakers, and others they feel ought to recognize and respond to the issue. The medicalization of social problems therefore refers to the specific analysis of the convergence of these two processes, to the ways by which medical language has come to be bound up in definitional activities relating to social problems.

From at least the early twentieth century to the present day, many aspects of life once considered the product of bad people (volitional deviance) have come to be considered as illness, the product of sick people (unintentional deviance) (Cockerham and Ritchley 1997). The seven deadly sins have been recast in the language of personality disorder (Best 2008). Those responsible for adverse conditions are said not to require punishment so much as help; they do not require reform so much as treatment. Audiences are encouraged to act not on the basis of right and wrong, but on the basis of healthy and unhealthy. Children once classed as unruly, disruptive, or otherwise displaying undesirable behaviors are routinely labeled with disorders for which a variety of drug treatments are available (Cockerham and Ritchley 1997). Solutions to social problems like poverty, inequality, or unemployment are increasingly sought not in deep underlying economic structures, but through, for example, the provision of therapy or the close medical or

other expert surveillance of behaviors and consumption habits of families and even pregnant women. It is worthwhile to mention, however, that the move of conceptualizations of social deviance from "badness" to "sickness" does not always entail blamelessness on the part of those classed as victims. Cultural appraisals of illness often reveal a great deal of ambivalence with regard to those labeled as ill – on the one hand describing them as victims and, on the other, as is sometimes the case with obesity and smoking as social problems, complicit in their own suffering.

Medicalization has arguably grown to become one of the leading means through which modern societies attempt to understand and address conditions they view as problematic (Best 2008). Its expansion is fostered both through the increased tendency to use the language of biology and medicine to understand social problems and the broadening of existing and new diagnostic categories, and particularly psychiatric categories, to encompass ever greater varieties of deviance (Conrad and Mcgann 2011). As with the expansion of the medicalization process in general, the success of this way of understanding social issues came at a time when advancements in science and medicine led to rising expectations that science would soon offer the answers to nearly all human problems, even those once considered solely moral, philosophical, or religious questions. Where these older belief systems seem to have atrophied, or at least lost their rhetorical power in the public sphere, the language of science seems to offer a value-neutral means of speaking to a diversity of audiences.

Note, however, that in the media and political arenas, where social problem claims mostly vie for attention, science operates differently to the way it (at least ideally) operates in other spheres. As Best (2008) describes, scientific developments are often slow and take a great deal of time to verify. By contrast, in social problem campaigns, uncertainty potentially undermines the credibility of claims-makers. Take, for example, the apparent tendency for particular foods and drinks to be associated on one occasion with health benefits and on another with health risks (coffee and wine being the most obvious examples). It is easy to see how this can lead to undermining the advice of dieticians if the facts seem to change from one day to the next. But studies must be repeated many times, mechanisms of causation identified, and alternative explanations invalidated before the latter can be definitively, if ever, put to rest. While conclusions may be drawn with a considerable degree of certainty in the physical sciences, this is less often the case in the biological sciences and least so in the social (Best 2008). However, in media claims-making about social issues, this ambiguity is often minimized.

Therefore, it may be more accurate to say that rather than "science" it is the rhetoric of science, or "scientism," that plays the greatest role in the medicalization of social problems. Scientism refers to the way that science can operate as a kind of secular religion by evoking a sense of ultimate truth, objectivity, and expert authority over personal and social matters (Freidson 1970). This is accomplished through the use of medical or technical language and models borrowed from the natural world in order to describe and understand human behavior and social life

(White 2009). With a bewildering array of potential social problems and limited space in newspapers, minutes in a television newscast, platforms in election campaigns, and now perhaps, characters in social media posts, it is essential that claims are effective in competing for attention and recognition among the cacophony of other competing claims. To do so, they must be compelling enough to move significant proportions of the population to act or at least not oppose the problem frame. Although performing many of the same functions, scientific and medical language has a greater potential to speak to, and avoid alienating, a far broader swathe of the public than moral, religious, or other nonscientific pronouncements.

Medical and scientific rhetoric also underpins the claims of particular classes of experts to assert authority and dominion over certain problems and their solutions. For example, through claims to specialized knowledge, psychiatrists were able to assert jurisdiction over a variety of deviant behaviors as symptoms of psychiatric disorders including juvenile delinquency, crime, homosexuality, and drug addiction (Best 2008). But specialisms and expertise need not come from medical training alone. It was not only psychiatrists, whose training placed them within the medical profession, but many others with more tenuous links to medicine including clinical psychologists, licensed social workers, and even those with little to no professional training (Best 2008). Professional "exes," survivors, and sufferers also steadily began to adopt the language of illness and disease to describe the problems they faced and to assert a special understanding over them. In addition to the previously cited example of the amateur campaigning for recognition of alcoholism as a medical condition by Alcoholics Anonymous, sufferers of contested conditions like myalgic encephalomyelitis (ME) and fibromyalgia have actively campaigned for medical recognition of their symptoms and have even been highly critical of doctors failing to recognize the medical origins of their difficulties (Furedi 2008). Collective action also played a key role in the movement of ADHD from a childhood condition to one diagnosed in adults. In these cases, diagnostic advocacy by laypeople who had largely self-diagnosed was decisive in legitimating new medical labels (Furedi 2008). In this way, the allusion of the early medicalization thesis to a "top-down" scapegoating of deviants has given way to a bottom-up movement, whereby an array of activists and citizen groups campaign for medical recognition (Furedi 2008).

Constructions of social problems benefit in a number of ways from adopting medical and scientific rhetoric. Most significantly, it allows advocates to eschew moral, aesthetic, or other overtly opinion- or value-based judgments in favor of an apparently value-free science that promises to act in the best interests of putative victims or populations as a whole. Medical claims, rooted as they are in the body, have the ability to bypass varied belief systems and political affiliations. Action is courted not on the basis of opinion, but on the basis of the true nature of (all) human beings, as uncovered by science. It thus encourages audiences to think about claims-makers' desired changes to bodies, behaviors, beliefs, or social policies not as moral injunctions, but as enhancements to personal health and well-being.

Causes and Contexts

Characterizations of the medicalization process have been subject to less debate than have its causes. While the thesis that "medical imperialism" ultimately underlies medicalization has waned, the expansion of expertise and professionalization of everyday life must not be entirely discounted. As the previous section described, the use of medical-scientific language is often a prelude to the assertion of authority and control over that domain by new or existing types of experts. By identifying new diseases and problems and offering expert solutions, many professions have created the needs they claim to satisfy, in turn justifying their existence and proliferation.

However, Peter Conrad has described a shift away from medicalization being primarily driven by physicians, social movements, and other interest groups toward a greater role played by pharmaceutical and biotech companies as well as consumers (Gabe 2013). In some countries, drug companies advertise new diseases and their cures direct to consumers who are encouraged to ask their doctors for specific drugs (Gabe 2013). Many parents actively seek out ADHD diagnosis and treatment for children who might otherwise have been labeled disruptive or unruly. Other parents fearing future discrimination suffered by short-statured children go to physicians seeking prescriptions for HGH (human growth hormone) (Conrad 2007). From this perspective, medicine has become a vehicle for broader projects of bodily enhancement and self-improvement, with consumers increasingly taking an active part. In some ways, this echoes earlier arguments made by Marxists and feminists that medicalization serves certain class and gender interests, but in this case those of large pharmaceutical and biotech companies (Gabe 2013).

Although perhaps instigated by professional or other interests, this does not entirely explain why, once the framework of health and illness was made available for making sense of daily life and social issues, it rapidly took off. Thus, medicalization studies have long rooted their analyses in more diffuse sociocultural processes rather than solely vested interests. Both Illich and Zola set their analyses against a backdrop of increasingly complex technological and bureaucratic structures, contexts that fostered the professionalization of everyday life and a reliance upon experts (Gabe 2013). Previous sections have described the ways by which additional forces such as secularization, the rise of science and technology, and the growing power of medical and scientific knowledge have all been bound up with the rise of medicalization. The institutional concern with health has also been suggested to have grown at a time when traditional politics and political affiliations were on the wane (Furedi 2008). The decline of a political focus on production and its succession by movements more concerned with consumption and lifestyles paved the way for ideals of health and well-being to pervade nearly all aspects of existence (O'Brien 1995). This expansion of the domain of health has meant that nearly any aspect of life and any social problem can be subsumed under its purview, from issues in interpersonal relationships to the structure of education and work. At the same time, health has come to be portrayed not as something one has, but something that must be promoted and actively pursued. As Wainwright (2008, p. 2) describes, health promotion has shifted the "clinical gaze from treatment of the sick

to regulation of the well. What we eat, drink and smoke, who we sleep with, how we relate to family members and friends, and the demands of working life, have all become subjects of professional advice in the pursuit of that elusive end point: "wellbeing"."

These broader social and cultural developments offer an explanation for why medicalization "from below," that is, by consumers and other members of the lay public, has become such an important driver. According to Furedi (2008), a key moment occurred in the 1980s when contestation of medical labels gave way to their embrace and even promotion by groups who once fiercely opposed them. For instance, where feminists have traditionally contested the medicalization of women's experiences, more recently there has been less opposition to, and even outright support for, the increased application of diagnostic labels like postnatal depression (Furedi 2008). For some people, illness confers a positive sense of identity and meaning as well as structures of support and kinship. In the face of a decline in other identities and the shared systems of meaning from which they stemmed, and in an increasingly atomized world, illness arguably offers a rare common ground on which people can unite and share common experiences. It also provides an explanation of problem behavior and a means of dealing with it in a way that attracts sympathy rather than condemnation or disdain (Furedi 2008). This has led to a situation in which diagnoses are not simply passively received, but are rather actively sought out or even demanded.

The rise of medicalization is likely multicausal in nature, but it is important to understand that this process would not be so successful outside of a sociopolitical context hospitable to claims framing social problems in medical language. It is clear that modern society is one in which a variety of parties now routinely seek out medical and "quasi-medical" solutions to an expanding range of social problems (Conrad 2007, p. 14). In this way, "medicalization of all sorts of life problems is now a common part of our professional, consumer, and market culture" (Conrad 2007, p. 14).

Consequences

Medicalization is generally recognized to have both positive and negative effects (Mcgann and Conrad 2011). Moving people into the sick role increases the likelihood that they will seek help and that solutions to their difficulties will be more humane. For instance, diagnosis and treatment for ADHD is probably a more humane way of treating children who do not seem to "fit" into traditional school systems than punishment and constant reprimand. Medicalization can provide coherence to people's lives and assurance that people are not at fault for their putative transgressions (Mcgann and Conrad 2011). People once classed as deviants in need of punishment, segregation, or even eradication are re-classed as requiring help, compassion, and care. In spite of this, the medicalization thesis is usually forwarded as critique. Indeed, in Zola's early development of the concept, he described medicalization as an "insidious" phenomenon, despairing of the creeping

pathologization of everyday life incurred through the attachment of medical labels (1972, p. 487). Thus, the concern is usually with "over-medicalization" rather than with a simple description of the movement of conceptualizations of phenomena from one category (nonmedical) to another (medical).

Medicalization theorists have long worried that medicalization encourages greater professionalization and disempowerment of ordinary people to deal with their problems. Both Illich and Zola worried that it increased people's reliance upon experts, granting undue authority to medical professionals over bodies, minds, and lives (Barker 2010). This professionalization of everyday life can transform previously mundane aspects of life into problems and everyday problems into subjects of professional knowledge and intervention, undermining existing means of living and coping. Moreover, social control is enacted through professional pronouncements on how to think and behave, allowing for its insidious expansion, its encroachment appearing unproblematic and therefore uncontested (Mcgann and Conrad 2011). Instead, as a "good citizen" one is simply expected to be "actively engaged with the advice of experts and lifestyle gurus, just as the recalcitrant citizen fails to adapt and adopt the identity and lifestyle moralities of psychology, health and medicine" (Back et al. 2012, p. 96).

The expansion of the sick role implied by medicalization also presents challenges to the moral autonomy of those classed as ill. Adoption of the sick role entails that the sick person is not responsible for his or her illness. While this potentially reduces stigma and culpability, it can also provide a "medical excuse" for deviance that diminishes individual responsibility for one's actions (Mcgann and Conrad 2011, p. 141). This acquires deeper significance as medical advancements, most notably in genetics and neuroscience, appear to locate the causes of various behaviors and dispositions ever deeper in human physiology, further deferring and displacing responsibility from the person to the body (Mcgann and Conrad 2011). The corollary of this can be a sort of "physiological determinism," which can encourage a fatalistic outlook and static vision of human potential. Further, the resultant construction of a human subject with a diminished capacity for rational action is one that also invites the increasing reliance upon experts.

In addition, the moral impulses that often guide social problem campaigns can be obscured through the use of medical and scientific language. Nazi Germany's early introduction of public smoking bans illustrates some of the ways that moral indignation and science can compound each other. Although Nazi tobacco epidemiology was among the most advanced in the world, smoking, like alcoholism, was antithetical to the ideology of racial hygiene (Cederström and Spicer 2015). Medical knowledge about harm and risk thus offered an underlying basis for moral judgments. In a similar way, while the apparent driver of contemporary debates about smoking has been ill-health and its associated economic costs, discussions are often underwritten by implicit moral indignation. While evidence of smoking's injuriousness to health may seem to straightforwardly lead to campaigns for its eradication, consider that many overtly risky behaviors are not widely considered de facto problematic including skydiving, mountain climbing, and other recreational activities that carry (or even court) high levels of risk.

Accepting a label is in some ways an admission of deviance. Although often not an admission that one is "bad" but merely "sick," there is nonetheless the implicit agreement that the condition is undesirable and, in many cases, needs to be rooted out. This implicit affirmation makes it difficult to propose alternative frameworks of meaning or otherwise challenge the medicalized frame. For example, the medical language of "obesity" guides campaigns primarily centered on threats to health and the economic costs of obesity-related illness. Even if the finger of blame is sometimes pointed at "obesogenic" environments and social structures, the focus of change is ultimately on the corpulent body. Medical language offers a guise of infallibility so that aesthetic and moral judgments about the social desirability of "fatness" seem irrelevant to the larger task of its eradication. However, even science requires moral judgment in terms of which questions are legitimately subject to scientific analysis, how questions are asked, results interpreted, and categories and classifications developed and decided. Indeed, body mass index (BMI) categories have been criticized for being culturally and aesthetically driven, particularly in terms of a lack of a significant relationship between the category "overweight" and adverse health outcomes (Best 2008). However, the language of science and measurement makes it difficult for alternative definitions of obesity, for example, a problem of discrimination, to compete against these medicalized definitions. Instead, the responsibility is placed upon those classed as deviant to change their behaviors and bodies (sometimes at great expense) rather than upon society at large.

Indeed, perhaps the most significant complaint forwarded by critics of medicalization is this tendency to deflect the focus of change from the social to the individual. By individualizing social problems, medicalization renders them apolitical, personal issues. For instance, the label of antisocial personality disorder is applied disproportionately to members of the lower social classes, often with experiences and histories heavily colored by poverty (Crews et al. 2007). Problems associated with poverty are therefore transformed into medical issues amenable to treatment rather than societal problems requiring political and/or economic solutions. Similarly, "female sexual dysfunction" locates within the female body issues that may well be rooted in broader cultural expectations associated with gender as well as problems within interpersonal relationships. Imbuing social problems with medical explanations also misses the opportunity to consider that phenomena classed as "deviant" may reflect rational adaptations to certain situations and contexts (Mcgann and Conrad 2011). It leads away from analysis of other potential explanations and encourages the view that even the most insoluble and perennial problems of society can at last be solved if only people would follow a given course of treatment, modify their behaviors, or otherwise learn to think and act in ways conducive to health.

Critical Perspectives

Particular case studies have been subject to a variety of critiques. However, the amalgamation of medicalization and a constructionist approach to social problems in particular has meant that such studies have been subject to many of the same

general criticisms launched against both of these perspectives. In terms of medicalization, it has been argued that studies overstate the degree to which particular cases have been medicalized and understate constraints on the process. For instance, welfare states may have less incentive to medicalize social problems and encourage expensive drug treatments for their amelioration than may be the case elsewhere. Claims about social control may also overlook the considerable role played by laypeople in the medicalization process as well as the benefits afforded by application of a medical framework. For instance, redefining a problem as an appropriate object of medical attention can reduce guilt and social stigma, lend legitimacy to sufferers, and make it available for research and possibly prevention (Gabe 2013). Moreover, even if operating as a form of social control, it arguably exerts this in a far more benign way than other institutions like religion or the law.

On the other hand, the constructionist outlook that colors medicalization studies of social problems means they tend to bracket assumptions about a preexisting reality, "out there," that can be accurately grasped in language. But this leaves open the question of why constructionist conceptualizations of issues should be considered any more valid than the medical ones being bracketed. However, what constructionist studies may offer are alternative constructions of reality that foreground the role of social processes in scientific and medical discoveries. They also bring forward the social significance of labels and definitions attached to social issues at a given time. As Conrad (2013) describes, this may or may not lead to undermining the categories under scrutiny. Albeit most often operating as critique, in the same way that researchers study industrialization, medicalization can be studied as a social process without necessarily entailing a judgment about its detriment. Indeed, there are many cases in which medicalization has arguably been beneficial including epilepsy and, in many ways, childbirth (Conrad 2013).

It is also sometimes suggested that medicalization studies simply replace the medical model with a similarly functioning social model. That is, medicine's disproportionate focus on the body is replaced with one that roots more and more diseases in society. This can have the perverse effect of expanding the purview of medical interventions deeper into individual and social life. From this perspective, it is social scientists more than doctors who have been complicit in this expansion. By moving social factors like inequality into the etiology of disease, the threats posed by these conditions come to be understood primarily as health risks (O'Brien 1995). Social problems come to be understood as health problems whose primary connection to the social is through the latter's illness-inducing capacity. However, rather than searching out the causes of problems, social or, otherwise, constructionist studies tend to be more interested in the etiology of definitions (Conrad 1992). Nonetheless, medicalization studies have at least partially fostered critiques of medicine's narrow focus on the body and call to expand definitions of health to include social factors. While this has demonstrable benefits, there is a danger that the movement toward a more socially aware and humane medicine may produce one with an increasingly broad remit of control.

Conclusion

Since its emergence in the 1960s and 1970s, medicalization has become a key concept in sociology. Its application to studies of social problems has broadened understandings of the diverse ways by which contemporary societies make sense of the issues that face them. Increasingly, this has been through the medicalized language of health and well-being, illness and disease, syndromes and disorders, and treatments and cures. This shift has had a number of positive effects including more humane treatment of those once classed as bad, mad, or criminal, the removal of stigma, and the provision of systems of meaning and social support. On the other hand, it can produce a number of negative consequences, most notably in its capacity to depoliticize social problems by transforming them into personal and technical matters closed off from debate and alternative definitions. It can also erode moral autonomy and call into question human rationality, fueling reliance on expertise and the ongoing professionalization of everyday life. However, the expertise fueling medicalization has broadened and expanded so that it no longer stems primarily from the medical profession but to a wide variety of parties claiming specialized knowledge over medical and technical definitions of social problems and how they are to be resolved. Indeed, one of the key questions animating medicalization studies has been the shifting causes of medicalization in the face of an apparent decline in the authority of the medical profession. Rather than being a straightforward result of professional interest, it has been suggested that the proliferation of medicalized understandings of social problems can best be understood by studying the broader contexts into which these claims emerge (Conrad 2007). In particular, the decline of shared meaning systems once proffered by religion and tradition and the rise of medicine, science, and technology have made the latter pivotal in legitimating claims. Today, those wishing to draw attention to social problems are more likely to look not just to medicine but also to neuroscience, psychology, psychiatry, and epidemiology for support over the word of God or the strength of tradition. In many ways, this marks a clear progress, but when applied to human problems, the rhetoric of science can serve to obscure the moral, aesthetic, or other value-based impulses that ignite such campaigns in favor of what appear to be purely technocratic interventions on behalf of the common good.

Studies of the medicalization of social problems have been subject to many of the same criticisms launched against medicalization on the one hand and constructionist studies of social problems on the other. In spite of criticisms, the medicalization thesis remains a useful framework for understanding the changing ways by which modern societies have come to understand and approach social problems. One need only flip through the pages of a newspaper or scroll through popular social media websites to see that defining problems in medical terms continues to hold a strong grasp over the public imagination. Making problems social can be a dangerous and difficult task (Crews, et al. 2007). Solutions are often complex and difficult to communicate in a compelling way; sometimes they can be too radical for

broad appeal or political expedience. Rendering them in medical language contains them and simplifies them and makes them seem less intractable and therefore treatable. But it is possible that some problems are unlikely to disappear even with the most advanced technological interventions; their solutions may lie in the deeper structures of society and may not be easily changed. Perhaps the biggest issue with medicalization is its tendency to deflect thinking from these deeper structures and from society as a whole, encouraging the view that nearly any problem is reducible to issues at the level of the individual body and mind.

Definitions of Key Terms

Medicalization	"[D]escribes a process by which non-medical problems become defined and treated as medical problems, usually in terms of illnesses or disorders" (Gabe 2013, p. 49).
Medical model	"A general framework for thinking about medical matters as diseases that require treatment" (Best 2008, p. 340).
Deviance	The violation of a social norm that may result in condemnation or punishment. Accounts of deviance can be subjective or objective. Objective accounts may consider the causes of deviant acts; subjective accounts consider how people and actions come to be defined as "deviant" (Goode 2011, p. 135).
Social problem	(1) Troubling conditions that affect society in some way (objectivist); (2) putative conditions defined as problematic by at least some individuals and groups in society (subjectivist).
Social construction	"The process by which people continually create – or construct – meaning" (Best 2008, p. 342).

Summary Points

- The medicalization of social problems refers to the process of applying medical definitions and descriptions to previously nonmedical issues.
- Medicalization is driven by a variety of professions, interest groups, and laypeople.
- Contextual factors fostering acceptance of medicalized social problems include secularization, growing power of medical and scientific knowledge, the decline of tradition, and the shift of political focus from production to consumption.
- Medicalizing social problems has benefits; it can remove social stigma, encourage people to seek help, and lead to more humane solutions.

- Medicalizing social problems has consequences; it can encourage reliance upon experts, undermine existing ways of coping, problematize notions of rationality and responsibility, close down other potential definitions, and lead to apolitical, individualized solutions.
- Studies have been criticized for overstating the consequences and scope of medicalization as well as the role of physicians in the process.
- Studies have been criticized for understating the benefits of medicalization and the role of laypeople in the process.
- Positing alternative social causes to medicalized social problems can ironically fuel medicalization as more and more aspects of social life come to be conceptualized as health risks.

References

Back L, Bennett A, Edles LD et al (2012) Cultural sociology. Wiley, Hoboken

Barker KK (2010) The social construction of illness: medicalization and contested illness. In: Bird CE, Conrad P, Fremont AM, Timmermans S (eds) Handbook of medical sociology. Vanderbilt University Press, Nashville, pp 147–162

Best J (2008) Social problems. Norton, New York

Busfield J (2006) Medicalisation. In: Scott J (ed) Sociology: the key concepts. Routledge, Abingdon, pp 99–103

Cederström C, Spicer A (2015) The wellness syndrome. Polity Press, Cambridge

Clarke A, Shim J (2011) Medicalization and biomedicalization revisited: technoscience and transformations of health, illness and american medicine. In: Pescosolido BA, Martin JK, McLeod JD, Rogers A (eds) Handbook of the sociology of health, illness, and healing: a blueprint for the 21st century. Springer, New York, pp 173–199

Cockerham WC, Ritchley FJ (1997) Dictionary of medical sociology. Greenwood Press, Westport

Conrad P (1992) Medicalization and social control. Annu Rev Sociol 18:209–232

Conrad P (2007) The medicalization of society. Johns Hopkins University Press, Baltimore

Conrad P (2013) Medicalization: changing contours, characteristics, and contexts. In: Cockerham WC (ed) Medical sociology on the move. Springer, New York, pp 195–214

Crews M, Moran P, Bhugra D (2007) Personality disorders and culture. In: Bhugra D, Bhui K (eds) Textbook of cultural psychiatry. Cambridge University Press, New York, pp 272–281

Foucault M (1965) Madness and civilization. Pantheon Books, New York

Freidson E (1970) Profession of medicine. Dodd, Mead, New York

Furedi F (2008) Medicalisation in a therapy culture. In: Wainwright D (ed) A sociology of health. Sage, London, pp 97–114

Gabe J (2013) Medicalization. In: Gabe J, Monaghan L (eds) Key concepts in medical sociology, 2nd edn. Sage, London, pp 49–53

Goode E, Ben-Yehuda N (2009) Moral panics: the social construction of deviance, 2nd edn. Wiley-Blackwell, Malden

Goode E, (2011) Deviance. In: Ritzer G, Ryan M (eds) The concise encyclopedia of sociology, 1st edn. Wiley-Blackwell, Malden, pp 135–136

Mcgann PJ, Conrad P (2011) The medicalization of deviance. In: Ritzer G, Ryan M (eds) The concise encyclopedia of sociology, 1st edn. Wiley-Blackwell, Malden, pp 141–142

Nettleton S (2013) The sociology of health and illness, 3rd edn. Polity Press, Cambridge

O'Brien M (1995) Health and lifestyle: a critical mess? Notes on the dedifferentiation of health. In: Bunton R, Nettleton S, Burrows R (eds) The sociology of health promotion: critical analyses of consumption lifestyle and risk. Routledge, London, pp 189–202
Sheaff M (2005) Sociology and health care. Open University Press, Maidenhead
Szasz T (1961) The myth of mental illness. Harper & Row, New York
Szasz T (1970) The manufacture of madness. Harper & Row, New York
Wainwright D (2008) The changing face of medical sociology. In: Wainwright D (ed) A sociology of health. Sage, London, pp 1–18
White K (2009) An introduction to the sociology of health and illness, 2nd edn. Sage, London
Zola IK (1972) Medicine as an institution of social control. Soc Rev 20:487–504

Changing Human Nature: The Ethical Challenge of Biotechnological Interventions on Humans

65

Jan-Christoph Heilinger, Oliver Müller, and Matthew Sample

Contents

Abstract

There is currently a prolific ethical debate about biotechnological interventions into human beings and the potential to alter the human organism, its functioning,, or genetic makeup. This article presents how such interventions can be seen as a challenge to concepts of "human nature" and reviews the different understandings

J.-C. Heilinger (✉)
University of Munich, Munich Center for Ethics, Munich, Germany
e-mail: heilinger@lmu.de

O. Müller
University of Freiburg, Department of Philosophy and BrainLinks-BrainTools, Freiburg, Germany
e-mail: oliver.mueller@philosophie.uni-freiburg.de

M. Sample
University of Washington, Department of Philosophy, Seattle, WA, USA
e-mail: sample@uw.edu

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_52

of this notion. Representative of this debate is the ethical concern about "human enhancement interventions" that aim to improve human functioning beyond what can be considered as healthy or normal. The different ethical stances toward enhancement are presented.

Introduction

One of the consequences of the "biotechnological revolution" is that "human nature" is increasingly claimed as a central topic of bioethical reflection. It is widely assumed that some of the currently available and emerging biomedical and biotechnological interventions into the human organism may alter human nature and our understanding of it, respectively. In order to understand and assess this assumption it will be necessary to specify which interventions into the human organism may have an altering effect on human nature (section "Medical and Biotechnological Interventions as a Challenge for Conceptions of Human Nature"). Next, it is important to be clear about how to understand human nature in the first place, seeing as it is a complex term with a long philosophical history. This conceptual investigation requires two steps. It is important, firstly, to identify the metaethical and conceptual problems, gaining an awareness of the intricate relation between human nature and normativity. Secondly, it is necessary to distinguish the main conceptions of human nature as they are employed in bioethical debate (section "Human Nature in Bioethics: Conceptions and Implications"). On this basis, the normative function of arguments that draw on concepts of human nature can be analyzed more precisely. Specifically, within the current bioethical debate, the different conceptions of human nature support three main positions: bioconservative, bioliberal, and transhumanist. Each position represents a certain set of attitudes toward biotechnological interventions and bioethical (and biopolitical) agendas. This is particularly evident in the debate about human enhancement interventions (section "The Ethical Debate on Enhancement: An Exemplary Clash of Views About Human Nature"). As will become clear, the debate about conceptions of human nature is ongoing, and the dispute between the different positions is not settled yet.

Medical and Biotechnological Interventions as a Challenge for Conceptions of Human Nature

Human nature has been a subject of debate for a broad range of biomedical and biotechnological interventions. These interventions can be assembled into four general groups based on their mode of action: genetic, surgical, pharmaceutical, and prosthetic. While each group presents unique ethical challenges, the worry that humans are changing human nature often occurs as a technology pushes the

boundary between therapy and enhancement or when a technology intervenes on a culturally significant part of the body (e.g., our genes or our brain).

Genetic interventions, here, refers to the set of technologies and procedures that center on the genetic material of humans or human cells. Notable examples include both indirect gene selection, such as preimplantation genetic diagnosis (PGD) and selective abortions, as well as more active interventions that directly modify genetic material, such as germline enhancement, gene therapy, and the creation of human-to-animal chimeras. Though the characteristics of an adult human are not fully determined by genetic endowment, the mere possibility of removing certain genes from the population or promoting the proliferation of a culturally desirable gene has incited scholars to weigh the importance of an unengineered genome (Habermas 2003).

Surgical interventions typically do not trigger the same concerns about tinkering with humanity's genetic "essence." But to the extent that they alter the body and its typical function or introduce foreign tissue into a person, they too can seem threatening to our nature. Cosmetic surgery, transplantation, xenotransplantation, body modification such as tongue splitting, and voluntary amputation of healthy limbs are some of the more striking examples. Taking a particular case, xenotransplants might provide much-needed organs for society, but the creation of adult human chimeras may problematically blur the line between human and nonhuman animals (Robert and Baylis 2003).

Psychopharmaceutical interventions aim to change an individual's cognitive or emotional state by acting on the brain. Their therapeutic and off-label use has been discussed extensively in the bioethics literature (see, e.g., Elliott 1998). Antidepressants, for instance, are marketed for a range of mood disorders beyond depression. Modafinil and other stimulants, though intended to treat narcolepsy and attention deficit disorders, are often perceived as multipurpose "steroids" for the mind. By aiming to change the brain via chemical rather than cognitive means, psychopharmacology tends to challenge many deep-seated cultural intuitions regarding authenticity and self-control. The concept of human nature, however, becomes increasingly relevant as ethical debate is confronted with a future in which the nontherapeutic ("cosmetic") use of psychopharmaceuticals is commonplace (Parens 2000).

Prosthetic interventions, lastly, augment the human form with additional "hardware," often involving both structural support and complex circuitry. Therapeutically, prosthetics can replace missing limbs, restore neural function, and support muscles weakened by age. To this end, engineers and scientists have developed a range of smart prosthetics, brain-computer interfaces ("neuroprosthetics") that reconnect the motor cortex to a limb or a muscle, and robotic exoskeletons. But even in therapeutic contexts, scholars have focused on the potential for these technologies to stretch current conceptual categories. Deep brain stimulators, for example, may be able to replace cognitive functions lost as a result of brain disease or damage; it is an open question whether or not a brain composed of both tissue and machine constitutes an unprecedented way of being (Schermer 2009).

Human Nature in Bioethics: Conceptions and Implications

If biomedical technologies seem to intervene in our nature in ways that might change our understanding of "human nature," then this intuition presupposes, on the one hand, that we have a solid knowledge of what human nature is, including the implication that shifts in the conceptions of human nature are or should be of ethical concern. On the other hand, attempts to bring "human nature" – a concept that has played most often no important role in the major theories of medical ethics – into the bioethical debate may be seen as an effective strategy to introduce specific – and neglected – normative frameworks into the debate. Although the question "what is 'human nature'?" has been part of ethical reflection since the ancient Greeks, in a sense it is true that ethical theories have systematically tended to avoid drawing on concepts of human nature at least since Kant; the term "human nature" was considered either too vague or metaphysically ballasted. This is one of the reasons why the concept of "personhood" made its career in ethical theories. However, the biotechnological revolution entails a certain discomfort in view of the constraints that are carried with the concept of the person – and forces bioethicists to rethink the relevance of "human nature" for an adequate ethical evaluation of new technologies.

Such concern for "anthropological" issues includes two interconnected questions, both the classical philosophical question "what actually is 'human nature'?" and the ethical question "what kind of normative framework (which set of values) follows from a robust concept of human nature?" Obviously, these questions lead to conceptual and methodological problems, which will be mentioned briefly (section "Conceptual and Meta-Ethical Aspects"). In distinguishing the main arguments that draw on specific concepts of "human nature" within a bioethical argument, one can group them into a highly schematic "pool" of prominent positions (section "A "Pool" of Prominent Positions").

Conceptual and Metaethical Aspects

Among the wide range of conceptual problems that arise when referring to "human nature," there are two aspects of particular importance for bioethical debate: firstly, the conceptual and semantic problems regarding the notion of "nature" generally; secondly, the problems of drawing normative conclusions from a description of human nature that are often discussed under the label "naturalistic fallacy."

Firstly, the term "human nature" carries an intricate ambiguity: Nature can mean not only "essence," in some cases, but also can refer to concepts or background theories of "naturalness." Ambitious philosophical approaches try to combine "essence" and "naturalness" systematically (as prominent "classical" position Scheler 1928/2007), but in bioethical contexts one often finds a certain focus on the mere "naturalness" of human beings – and this changes both the perspective on "human nature" and the reflection on the normative implications (see Birnbacher 2014 for a discussion). But often "personhood" is invoked to capture the "essence"

of human nature (Singer 1979 may count as a prominent and controversial example). The specific qualities and capabilities of persons – such as self-consciousness, rationality, responsibility, etc. – then count as the core aspects that should be included in an ethical reflection or even foundation. Given the focus on "naturalness," however, one tends to discuss aspects that are not necessarily covered by the concept of a person. Unsurprisingly, naturalness could be conceptualized in very different ways, from reductionist positions that accept only the biological nature of homo sapiens to Aristotelian approaches that are based on a natural striving for certain goods to theories of embodiment that include the vulnerability and fragility of human existence to metaphysical theories of a God-given "pure" nature – to name but a few distinctive positions.

Secondly, on a metaethical level it has to be discussed whether justifying a norm or value by reference to a concept of human nature is plausible. David Hume was right to raise awareness for the is-ought gap in ethical theories because there is indeed a certain temptation to step in an unreflective way from a descriptive "is" to a normative and prescriptive "ought." This, however, could be a severe methodological problem when dealing with human nature in ethics (Hume 1739/2011). It is rather easy to claim that human nature or a certain aspect of human nature should be preserved or should be the benchmark for a value. But it is difficult to substantiate these kinds of intuitions and provide good reasons for the normativity of human nature or parts thereof. So, Hume's formulation of the is-ought problem as well as Moore's "naturalistic fallacy" and his "argument of the open question" raise concerns about the problematic relation between natural and moral properties (Moore 1903/1971). Yet the charge of committing a naturalistic fallacy can be abused, too. Stipulating, for example, that natural human fragility or vulnerability has some moral significance or even moral value is not necessarily based on that fallacy. Nor is referring to a normatively charged Aristotelian understanding of nature per se fallacious. The role of human nature stands or falls by the normative framework that explicates certain values – which, however, always has to avoid simplistic inferences of norms from its understanding of human nature.

A "Pool" of Prominent Positions

During recent decades, different conceptions of human nature have been employed to evaluate the ethical implications of biomedical interventions. The following "pool" of prominent positions tries to categorize the main concepts and arguments. These positions are ideal types, with a focus on select examples; they stand for a range of partly overlapping variants that cannot be fully spelled out here. The present focus will be on conceptions of human nature that are promoted as alternatives or additions to normative conceptions of "personhood." Consequently, the long Western tradition of humans as persons or of defining humans as *animal rationale* is missing in the following compilation. While the demand for "informed consent" with its underlying understanding of a human being as a rational, autonomous agent could count as the most prominent example for a bioethical principle that is based on

a conception of human nature in this respect, too, drawing on human nature in the bioethical debate usually goes beyond a mere guarantee of such rational, informed consent in difficult decisions.

Human Nature as Pure and Untouched by Culture and Technology

In some contexts human nature is understood in an essentialist way which is based on the conviction that there is a "pure" human nature before all culture and technology. This is Rousseau's idea of an "authentic" nature that becomes subverted by technology ("L'Homme naît naturellement bon, c'est la société qui le corrompt."). More prominent in the debate are conceptions that understand human nature as somehow "given" and "untouched" before technological interventions. Leon R. Kass is a well-known proponent of this position. In his book *Life, Liberty and the Defense of Dignity* he understands medical technologies as challenging human nature as such: "Human nature itself lies on the operation table, ready for alteration, for eugenic and neurophysic 'enhancement' for wholesale redesign. In leading laboratories, academic and industrial, new creators are confidently amassing their powers and quietly honing their skills, while on the street their evangelists are zealously prophesying a posthuman future. For anyone who cares about preserving our humanity, the time has come to pay attention." (Kass 2002, p. 4). Kass underlines that his reference to human nature means more than referring to the concept of a person (and a "liberal" framework, respectively): "The account of human dignity we seek goes beyond the said dignity of 'persons' to reflect and embrace the worthiness of embodied human life, and therewith of our natural desires and passions, our natural origins and attachments, our sentiments and aversions, our loves and longings. . . It is a life that will use our awareness of need, limitation and mortality to craft a way of being that has engagement, depth, beauty, virtue and meaning – not despite our embodiment but because of it." (ibid., pp. 17–18) It is typical that such an understanding of human nature perceives technology as a threat. With their potential changes, new biotechnologies undermine this concept of human nature and consequently put "human dignity" in danger. Kass' human nature has a pretechnological status; it is understood as something that is technologically untouched and has to be treated as a "given" – by God or nature itself. This "givenness" has a normative dimension insofar as Kass and others argue to restrict biotechnological interventions. Speaking of the untouched status of human nature implicates that it is "fixed." In this respect, Kass' position is close to approaches that refer to the "biostatistical" normalcy of human nature (see section "Biological Conceptions of Human Nature").

Kass explicitly uses religious and Judeo-Christian intuitions about human nature, but one can find the idea of human nature as "untouched" in texts by secular authors as well. A famous example is Jürgen Habermas, who speaks of "das Gewachsene" or "das Gewordene" ("the grown," cf. Habermas 2001, p. 80). Habermas' notion is, albeit on a different foundation, quite close to the idea of the "untouched" as he contrasts "das Gewachsene"/"Gewordene" with "das Gemachte" ("the made"/"the produced"/"the fabricated"). Habermas explicitly refers to the Aristotelian distinction between nature and technology. In doing so, Habermas bases his theory on the presupposition that human nature and technology could be thought of as separated

and that there are at least some important pretechnological elements of human nature. Consequently, the "Gewachsene" has normative implications; in the Habermasian sense it should be respected as something valuable per se that is not at our unrestricted disposal.

Human Nature as Contingent

Some authors assert that humans should strive to accept natural bodily differences that they might call disabilities or imperfections. By this reasoning, one should hesitate to promote medical interventions that promise to remove such differences, like selective abortions or genetic enhancements. Garland-Thomson (2012) is a useful example of these positions; she calls for a "conservation" of disability as a natural and valuable part of the quite diverse human condition. For her, "honoring the 'is' rather than the 'ought'" of human nature is based on a conviction that differences coming along with disability and imperfection are valuable. The existence of bodily differences, she argues, is not a liability that needs to be tolerated and protected but is rather a source of narrative, ethical, and epistemic resources. The empirical understanding of the world, for instance, might be impoverished if society erased individuals who cultivate alternative sensory modalities.

Sandel (2009) fits within this framework of valuing difference and disability. He claims that an "openness to the unbidden," a recognition that we humans are not in complete control of our lives, is crucial for healthy parent–child relationships and encourages kindness and generosity toward the less-fortunate in society. Similarly, Parens (1995) stresses that human "fragility" is a source of diversity and meaning for us. For these authors, the existence of disability or imperfection is a resource for humanity, a chance to appreciate good fortune and to exercise virtuous behavior toward others. Medical technologies that promise to remove human vulnerability, with their "Promethean striving," can represent the opposite attitude, the expectation that individuals master their own fates such that virtuous behavior is not necessary.

Though these authors avoid subjecting themselves to the charge of committing a naturalistic fallacy, their arguments do rely on some factual claim about human nature; their prescriptions for acceptance presuppose that human nature is essentially contingent and entails some inescapable tendency toward difference and disability. As Garland-Thomson emphasizes, the human body transforms as it ages, pulling even the most able-bodied into a radically different mode of being. And it is this inevitability, she notes, that makes disability a "generative" concept. To the extent that society has the biotechnological or medical means necessary to erase such bodily differences and prevent economic or biological misfortune, arguments based on the capriciousness of fate might seem less compelling, and the value of accepting the "unbidden" might wane (see section "Human Nature as Flawed" for an elaboration of this line of thought).

Aristotelian Understandings of Human Nature

Several positions in the debate stand in an Aristotelian tradition: as more or less "essentialist" these positions tend to understand human nature as striving for a good life. In her contribution to "Human dignity and Bioethics" (one of the essays

commissioned by the President's Council of Bioethics), Martha C. Nussbaum underlines that we humans have to include more aspects of human nature than only rationality. This is perfectly in line with the approaches of human nature that try to complement the concept of personhood: "In general, when we select a political conception of the person we ought to choose one that does not exalt rationality as the single good thing and that does not denigrate forms of need and striving that are parts of our animality. Indeed, it is crucial to situate rationality squarely within animality, and to insist that it is one capacity of a type of animal who is also characterized by growth, maturity, and decline, and by a wide range of disabilities, some more common and some less common. There is dignity not only in rationality but in human need itself and in the varied forms of striving that emerge from human need." (Nussbaum 2008, p. 363) On this basic assumption about human nature, Nussbaum developed a "capabilities approach" that tries to list the main capabilities of human beings, based on a series of dimensions of the essence of human nature (such as being able to live to the end of a human life of normal length, have good health, have adequate shelter, form a conception of the good, and engage in critical reflection about the planning of one's life).

Nussbaum's view is only one example for an Aristotelian approach to human nature. Foot (2001) and Thompson (1995) are others. The latter introduced "Aristotelian categoricals" in the debate that should help to describe human nature in a nonreductionist way. Typical for such positions is the idea that the "human flourishing" should be respected. In contrast to "human nature as untouched," more sophisticated Aristotelian positions do not perceive biotechnological interventions as per se against human nature because the concepts of human flourishing and of human capabilities allow the integration of biomedical techniques into the "good life." This is a political question as Nussbaum points out: "It means that the respectful government promotes health capabilities, not healthy functioning. That is, it should make sure that all citizens have adequate health insurance and access to good medical facilities... In short, respecting human dignity requires informing people about their choices, restricting dangerous choices for children, but permitting adults to make a full range of choices, including unhealthy ones – with the proviso that competitive sports need to set reasonably safety conditions so that unwilling participants are not dragooned into taking a health risk that they don't want to take." (Nussbaum 2008, p. 370). The capability approach and its conclusions, however, are not paternalistic. Buchanan (2009), in discussing the President's Council for Bioethics, notes that arguments based on human nature can be interpreted as either a form of normative essentialism or an assertion about the good life. Strengthening the capabilities means strengthening the responsibility to lead a good life – hence to fulfill the intrinsic goodness of human nature.

Biological Conceptions of Human Nature

Human nature can also be understood in terms of the characteristics of the human species, as disclosed by biological and other scientific investigation. Fukuyama (2003) defines human nature as "the sum of the behavior and characteristics that are typical of the human species, arising from genetic rather than environmental

factors." In drawing on a population-based understanding of our nature, Fukuyama follows current scientific categorizations of the human species. According to these "biologically respectable" understandings, a species is more like a particular individual with relative properties than a "kind" with some definable essence; unlike a chemical element, a species has no intrinsic properties and refers to a population that comes into existence, can change over time, and may even comprise members that are genetically different (Lewens 2012). Authors differ, however, in what they choose to draw from this scientific insight.

Fukuyama, for example, takes the complex features that are shared by members of the population to justify an account of human rights. He suggests that biotechnological alteration of the species risks upsetting "Factor X," his term for a multicausal human quality that justifies basic rights, gives humans a common form, and grounds liberal social structures. For Fukuyama, then, the cost of altering our nature is nothing less than endangering "human dignity." But for practical reasons some do not follow Fukuyama to this conclusion. Daniels (2009), for example, who also advocates a biostatistical understanding of human nature, stresses that population-scale interventions are highly unlikely, given the current and near-future state of medical technology. And if we humans are unable to actually change human nature, then the associated ethical worries are merely hypothetical.

Simultaneously, other authors deny that a biological understanding of human nature can contribute to ethical debates. Buchanan (2009) highlights the fact that if humans have a nature, it likely consists of both desirable and undesirable characteristics. It is an empirical question, he argues, whether specific interventions on human biology will inevitably disrupt the desirable tendencies. Thus, the question of altering our nature can seem too broad to be helpful. Along these lines, Lewens (2012) and Buchanan (2009) both stress the fact that invocations of human nature – biologically understood – often disguise normative commitments in seemingly descriptive scientific language. In this way, they highlight the possibility for nature-based arguments to preempt discussions of value by relying on the cultural authority of science or by relying on the naturalistic fallacy.

The *Homo Faber* View of Human Nature

In the philosophy of technology there is a certain consensus that human beings are "by nature artificial" and because of their natural disposition "technicians" (cf. Plessner 1928/1975). What is called human nature is the nature of *homo faber* who is constantly transforming the world and, in doing so, himself. In this context, technology is often compared to language. Technology is then a way of self-exploration and self-understanding (cf. Cassirer 1930/2004). Human nature is not thinkable without technology. There is no pretechnological human nature. This assumption is widely spread in the debate although it is not often made explicit. A controversial exception is Andy Clark, who argues explicitly that human beings always tend to transform themselves technologically. Human nature cannot be understood unless we think of it as being technologically shaped; Rousseau's "pure" human nature simply does not exist. To encapsulate his position he coined the expression of human beings as "natural-born cyborgs," creatures in transition.

Clark's main work "is the story of that transition and of its roots in some of the most basic characteristic facts about human nature. For human beings, I want to convince you, are natural-born cyborgs. . . What makes us distinctively human is our capacity to continually restructure and rebuild our own mental circuitry, courtesy of an empowering web of culture, education, technology, and artifacts." (Clark 2003, pp. 3, 10) This *homo faber* position includes openness for any transformation of human nature. Yet it is no "transhumanist" or "posthumanist" position (see below) because it does not call for overcoming human nature as it is.

Biotechnological interventions may allow misuse, but as human nature is per se and always technologically transformed, humans only have to be careful, not abstinent: "And we do need to be cautious, for to recognize the deeply transformative nature of our biotechnological unions is at once to see that not all such unions will be for the better. But if I am right – if it is our basic human nature to annex, exploit, and incorporate nonbiological stuff deep into our mental profiles – then the question is not whether we go that route, but in what ways we actively sculpt and shape it. By seeing ourselves as we truly are, we increase the chances that our future biotechnological unions will be good ones." (Clark 2003, p. 198) Normatively, this position promotes openness to biomedical interventions and is insofar close to a bioliberal attitude in general. Consequently, the *homo faber* view is unable to develop concrete advice about *which* biotechnological transformation may be worthwhile or not, because of the imprecision of its underlying concept of human nature.

Human Nature as Flawed

The vision of humans as *homo faber*, as natural technicians, takes on a different meaning when it is placed within a narrative of progress. Transhumanism has provided one such narrative in the recent bioethical literature. Bostrom (2003) asserts that the transhumanist sees human nature as "as a work-in-progress, a half-baked beginning that we can learn to remold in desirable ways." (p. 493) Senescence, limited memory, and vulnerability to disease may be part of the human condition, as it exists, but the transhumanist does not afford this fact any special significance, as do the authors (in section "Human Nature as Pure and Untouched by Culture and Technology" and section "Human Nature as Contingent"). Transhumanists see humanity's undesirable characteristics as flaws waiting to be fixed via technoscientific or even social interventions. Accordingly, Bostrom and Sandberg (2009) suggest the use of an "evolutionary heuristic" to sort out which traits are well suited for the modern world from those which are only holdovers from humanity's hunter-gatherer past, such as metabolic constraints on cognitive capacity or immune activation. In this way, transhumanists hope to plan a deliberate transition to the "posthuman."

By calling for the redesign of human nature, the transhumanist perspective does more than merely imply that tool use and self-modification is part of our nature. Beyond the invocation of *homo faber*, transhumanist literature presents a theme of progressive change, a persistent yearning for a better nature. It is an impulse in keeping with H.G. Wells' vision for the dawning twentieth century; he asks, "why should things cease with man?" (Wells 1902). Like Wells, transhumanists stress both

the promise and inevitability of "posthumanity." The implication of "posthuman" is not that humanity will completely disappear but that the particularities of our current nature, especially the negative ones, are transient (Birnbacher 2009). The transhumanist understanding of human nature, thus, can be interpreted as a negative point; human nature is not what it could (and perhaps should) be. This idea will be explicated further in the next section.

The Ethical Debate on Enhancement: An Exemplary Clash of Views About Human Nature

In the current ethical debate about the legitimacy of altering human nature, the variety of views presented above becomes visible. As indicated above, the constellation of positions regarding human nature resists easy categorization as "pro-" and "anti-" biotechnology. Even when scholars agree on a feature or a definition of human nature, they can leverage that understanding to justify radically different attitudes toward biotechnological or biomedical interventions. Biostatistical or biological understandings of human nature, as mentioned above, have been used to both forbid alteration of human nature and – somewhat paradoxically – to dismiss the very possibility of such alteration. This flexibility in the argumentative use of human nature demands that we carefully disentangle the subtly different ways in which scholars ground their conclusions. One can, nevertheless, place scholars within a broader framework based on their general attitude toward changing human nature, whether enthusiastic, worried, or otherwise. This last section is devoted to that task.

The debate on human enhancement can be understood as a forceful and significant clash of different concepts of human nature. For a decade or so there has been a thriving international debate on the legitimacy of enhancement technologies, on biotechnological interventions "beyond therapy." Typical arguments in this debate refer to justice, fairness, individual responsibility, etc. It is evident that beside these normative concepts human nature is seen as relevant for the ethical evaluation (Heilinger 2014). And it is characteristic for this debate that different concepts of human nature are subsumed in the named broader frameworks that "bundle" different concepts of human nature.

One such general framework has emerged organically within the bioethics literature and deserves consideration here. One can distinguish three main views with regard to such alterations: a bioconservative, a bioliberal, and a transhumanist one. Each of these views draws on different understandings of human nature in order to reach ethical conclusions about the legitimacy of enhancement interventions. These views are distinguished by whether human nature is to be seen as fixed or malleable and as intrinsically or instrumentally valuable. The three views pertain to the use of the various technologies that aim at altering human nature, including therapies, but become particularly visible in the debate about human enhancement interventions.

Those who see alterations of human nature as morally problematic and highlight the dangers of technology are often called bioconservatives. In stressing the importance of preserving human nature as it is, they often cite an "untouched" human

nature that is fixed and intrinsically valuable. Those who are willing to endorse changes in human nature if it is conducive to a greater good are often labeled transhumanists or posthumanists, since they wish to expand the current boundaries of human existence. This position tends to stress the malleability of human nature, as a means to achieve other instrinsically valuable states. Between these two extremes, one can find bioliberals, who do not deny that changes in human nature occur and that some of them may be very worthwhile, while holding that human nature may have both some intrinsic and instrumental value. These three views will be briefly presented.

Still exemplary for the bioconservative position is the influential report "Beyond Therapy. Biotechnology and the Pursuit of Happiness" (2003) by the former President's Council on Bioethics (which Barack Obama ended when he came into office). In this text the negative consequences of attempting to alter human nature are vividly illustrated, sometimes with a clearly evident religious background. The authors of this report – among them Leon Kass, Francis Fukuyama, and Michael Sandel – direct attention first and foremost to several "essential sources of concern." Kass' concept of human nature that is explicated in this context was mentioned above (section "Human Nature as Pure and Untouched by Culture and Technology"). It is one possible background assumption about human nature which may lead to a bioconservative position that can be described as reacting to certain "dangers." There is (1) the danger of lacking humility and "respect for the given" when playing God and trying to alter human nature with its intrinsic value, (2) the danger that unnatural means will threaten the dignity of the naturally human way of activity, because the valued process of effort, success, and merit is cut short, (3) the danger that individuals will lose their personal identity and individuality when undergoing alterations of their human nature and (4) the danger that the pursuit of perfection in some domains of human existence may, when it accompanies alterations of human nature, ultimately lead to an impoverished life, not a flourishing one. On the basis of these concerns, the bioconservative position tries to find arguments against enhancement technologies, referring to the aforementioned dangers and the supposed threat to human nature.

As mentioned above (section "Human Nature as Flawed"), trans- or posthumanists fully endorse the options provided by using biotechnologies to enhance human beings. Transhumanists see themselves as "extending the liberal democratic humanist tradition to a defense of our right to control our own bodies and minds, even if our choices make us something other than 'human'" (Hughes 2004, p. xv). Even if, in the end, humans will turn into transhuman beings, this is not to be regretted. Quite to the contrary, humans should strive to bring about beings that are capable to reach these higher states of mind.

Transhumans can be understood to be still human beings, but those humans that are already on their way to becoming posthuman. Bostrom, for example, argues that some distinctively posthuman modes of being are intrinsically valuable, even though they are not only gradually but substantially different from standard human modes. Hence it could be very good for human beings to become posthuman. By a posthuman, he understands a being that has at least one "posthuman capacity,"

understood as "a general central capacity greatly exceeding the maximum attainable by any current human being without recourse to new technological means" (Bostrom 2008, our emphasis). These capacities are, following Bostrom, a healthy life span, cognition, or emotion.

The posthumanist argument for alterations of human nature shifts the burden of proof from those who want to strive for posthuman capacities to those who want to deny persons the pursuit of those capacities. If there are means available to improve the human lot in realizing obvious, widely shared goods – like living longer, healthier lives, becoming smarter, or emotionally better off – those who are opposing such changes stand in need to justify their opposition. One should not hold back because of concern for some abstract idea of the nature of human beings. What matters is that lives go well, not that they match some disputed idea of a fixed human nature, or so transhumanists argue.

Between bioconservatives and transhumanists one can identify bioliberals as holding an intermediate position. Despite all differences, these intermediate positions share the conviction that there are no principled objections against alterations of human nature. They hold this view either because they doubt the existence of something fixed that can be called human nature or because they doubt that human nature is intrinsically valuable and must hence never be altered. Instead, they are willing to consider alterations on a case-by-case basis. Buchanan has rightly pointed out that the dispute in the ethical debate about human enhancement is not so much between those who object to and those who promote enhancements (contra- and proenhancement) but between those who fundamentally object to enhancements (antienhancement) and those who do not share this general objection (anti-antienhancement). Those being anti-antienhancement hence are not obliged to promote or call for enhancements. Their position is determined rather by a willingness to consider enhancement interventions if they meet certain other criteria (Buchanan 2011).

The bioliberal ethical assessment of possible interventions draws the main attention not to "anthropological" arguments about human nature. Instead, it focuses on assessments of risks and benefits and considerations about justice. Furthermore, as liberals they hold the individual, informed, and autonomous decision for or against some intervention with a potentially human nature–altering effect particularly important.

Conclusion

The understanding of human nature and its ethical implications in the context of medical, biotechnological, and enhancement interventions in the human organism is a widely debated topic in contemporary philosophy of medicine and bioethics. A lot of academic attention has been paid to the challenge to phrase in philosophical and secular terms the widely shared intuition that there are some limits to what we may ethically do with the biological underpinnings of our human existence. But the debate is not settled yet. Transhumanists often seem overly optimistic about the

possibility of improving the human lot while neglecting dangers associated with it. Bioconservatives often seem overly pessimistic and sometimes endorse religiously motivated views or intuitions that are difficult to justify in secular terms. This makes the bioliberal view attractive, at least prima facie. However, being generally open to allow for alterations of human nature is rather the statement of the problem and offers no solution yet. Hence, no single approach presented above can claim to have provided a solution to the challenge yet, but the bioliberal view, when it integrates both the concern and the promises connected to human enhancements, should, it seems, be pursued further in order to determine a wise way of dealing with the novel technologies.

Definition of Key Terms

Human nature	A complex, multivalent concept used to, among other things, justify various and sometimes contradictory attitudes toward biotechnological interventions on humans
Human enhancement interventions	Biological, technical, or medical interventions in the healthy human organism that aim to improve the human organism or its functioning beyond a level of normalcy
Biotechnologies	Intentional biological, technical, or medical interventions on living organisms
Bioconservatives	Those holding that human enhancements are morally prohibited, since they alter human nature
Bioliberals	Those holding that some forms of human enhancements may be morally acceptable, even if they come along with changes of human nature
Transhumanists	Those holding that human enhancements are morally desirable, because they help overcoming some limitations of human nature

Summary Points

- Biotechnological interventions that have the potential to alter human nature can be genetic, surgical, psychopharmacological, or prosthetic.
- The notion "human nature" plays a complex and disputed role in contemporary bioethics. It has to be employed carefully, due to its possibly normative content.
- The concept of human nature is contested, and there are different attempts to spell it out: some contrast human nature with culture; others understand it as consisting essentially in culture. Some define an ultimate and fixed essence of human existence, while others stress the contingency of human traits.

- Normatively, these views can support diverse evaluations of human nature as perfect and in need of protection on the one hand and of human nature as flawed and worthy of improvements on the other.
- With regard to the ethical debate about human enhancement interventions, three prominent attitudes are evident. Bioconservative views clash with bioliberal or even transhumanist views over whether enhancement interventions that may change human nature are morally problematic, acceptable, or desirable.

References

Birnbacher D (2009) Posthumanity, transhumanism and human nature. In: Chadwick R, Gordijn B (eds) Medical enhancement and posthumanity. Springer, Dordrecht, pp 95–106

Birnbacher D (2014) Naturalness. Is the "natural" preferable to the "artificial"? (trans: Carus D). University Press of America, Lanham

Bostrom N (2003) Human genetic enhancements: a transhumanist perspective. J Value Inq 37 (4):493–506

Bostrom N (2008) Why I want to be a posthuman when I grow up. In: Gordijn B, Chadwick R (eds) Medical enhancement and posthumanity. Springer Netherlands, Berlin, pp 107–136

Bostrom N, Sandberg A (2009) The wisdom of nature: an evolutionary heuristic for human enhancement. In: Human enhancement. Oxford University Press, Oxford, pp 375–416

Buchanan A (2009) Human nature and enhancement. Bioethics 23(3):141–150

Buchanan A (2011) Beyond humanity? The ethics of biomedical enhancement. Oxford University Press, Oxford

Cassirer E (2004) Form und Technik. In: Gesammelte Werke, vol 17. Felix Meiner Verlag, Hamburg, pp 139–183

Clark A (2003) Natural-born cyborgs. Minds, technologies, and the future of human intelligence. Oxford University Press, Oxford

Daniels N (2009) Can anyone really be talking about ethically modifying human nature. In: Savulescu J, Bostrom N (eds) Human enhancement. Oxford University Press, Oxford, pp 25–42

Elliott C (1998) The tyranny of happiness: ethics and cosmetic psychopharmacology. In: Parens E (ed) Enhancing human traits: ethical and social implications. Georgetown University Press, Washington, DC, pp 177–188

Foot P (2001) Natural goodness. Clarendon, Oxford

Fukuyama F (2003) Our posthuman future: consequences of the biotechnology revolution. Macmillan, Picador, New York, NY

Garland-Thomson R (2012) The case for conserving disability. J Bioeth Inq 9(3):339–355

Habermas J (2001) Die Zukunft der menschlichen Natur. Suhrkamp, Frankfurt am Main

Habermas J (2003) The future of human nature (trans: Beister H, Rehg W). Polity Press, Cambridge, UK

Heilinger J-C (2014) Anthropological arguments in the ethical debate about human enhancement. Hum Mente J Philos Stud 26(2014):95–116

Hughes J (2004) Citizen Cyborg. Why democratic societies must respond to the redesigned human of the future. Westview, Cambridge, MA

Hume D (1739/2011) A treatise of human nature. Oxford University Press, Oxford

Kass L (2002) Life, Liberty and the Defense of Dignity: The Challenge for Bioethics. Encounter Books, New York

Lewens T (2012) Human nature: the very idea. Philos Technol 25(4):459–474

Moore E (1903/1971) Principia ethica. Cambridge University Press, Cambridge

Nussbaum M (2008) Human dignity and political entitlements. In: Human dignity and bioethics: essays commissioned by the President's Council on Bioethics, President's Council of Bioethics, Washington, DC
Parens E (1995) The goodness of fragility: on the prospect of genetic technologies aimed at the enhancement of human capacities. Kennedy Inst Ethics J 5(2):141–153
Parens E (2000) Enhancing human traits: ethical and social implications. Georgetown University Press, Washington, DC
Plessner H (1975) Die Stufen des Organischen und der Mensch. Einleitung in die philosophische Anthropologie. De Gruyter, Berlin/New York
President's Council on Bioethics (2003) Beyond therapy. Biotechnology and the pursuit of happiness. Dana Press, New York
Robert JS, Baylis F (2003) Crossing species boundaries. Am J Bioeth 3(3):1–13
Sandel MJ (2009) The case against perfection. Harvard University Press, Cambridge, MA
Scheler M (2007) Die Stellung des Menschen im Kosmos. In: Gesammelte Werke, vol 9. Bouvier, Bonn, pp 7–71
Schermer M (2009) The mind and the machine: on the conceptual and moral implications of brain-machine interaction. Nanoethics 3:217–230
Singer P (1979) Practical ethics. Cambridge University Press, Cambridge
Thompson M (1995) The representation of life. In: Hursthouse R, Lawrence G, Quinn W (eds) Virtues and reasons. Oxford University Press, Oxford, pp 247–297
Wells H (1902) The discovery of the future. Nature 65(1684):326–331

Social Determinants of Health

66

Sridhar Venkatapuram

Contents

Abstract

This chapter discusses social determinants of health, an area of research and health policy initially coming out of epidemiology. Two categories of philosophical issues are presented including epistemological issues related to casual explanations as well as ethical issues related to health inequalities and social justice. In pursuing better explanations of causation and distribution of disease, social epidemiology expands the scope of causal chain outward beyond factors on or within the body as well as upward in terms of nested spaces such as family, neighborhood, region, country, and global system. New thinking about the ethical value of health and well-being and the causal role of social factors in producing inequalities in health raise questions of social justice and require drawing on disciplines such as political philosophy that evaluate conceptions of a good or just society.

S. Venkatapuram (✉)
Global Health and Social Medicine, King's College London, London, UK
e-mail: Sridhar.venkatapuram@kcl.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_72

Introduction

"Social determinants of health" (SDH) is a phrase that has emerged from the discipline of epidemiology, the science that identifies the determinants and distribution of morbidity and mortality. However, while epidemiological and related literature use the phrase "social determinants of health," in actuality, the language refers to the social determinants of disease, disability, illness, mortality, and other such negatively valued states of biological and mental functioning. Some chapters in the present book as well as other authors discuss how the concept of health is contested and in flux (Nordenfelt et al. 2001; Blaxter 2010; Cribb 2005; Venkatapuram 2011). And, indeed, there have been important advances recently in our knowledge about various dimensions of "positive health" (e.g., well-being, longevity, resilience, happiness, life satisfaction) (Huppert et al. 2005; Diener et al. 1999; Ryan and Deci 2001). Nevertheless, despite being directly relevant, the debates on conceptions of health as well as research on positive health are going to be set aside in the following discussion. Unless stated otherwise, SDH from here on refers to the social determinants of states of ill-health and premature mortality because that is what is most often meant.

There is a long history of identifying the role of the social environment in the causal pathways to disease and mortality. It was central to the epidemiological work of Louis-René Villermé and Rudolph Virchow in the nineteenth century (Virchow and Rather 1985; Julia and Valleron 2011). And it was very much a visible part of community and social medicine that began to flourish in the mid-twentieth century (Trostle 2004). However, as epidemiology developed into a distinct and sophisticated scientific discipline in the twentieth century in the United States, its research paradigm or scope was increasingly narrowed down to individual-level factors (Krieger 1994). That is, disease was understood to be caused by the independent or interactive result of individual-level factors such as individual biology, behaviors, and harmful proximate exposures to the body.

In contrast, contemporary social epidemiology harnesses state-of-the-art epidemiological tools and methodologies combined with sociological analysis to explicitly identify supra-individual social phenomena that affect both the causation and distribution of ill-health across individuals and social groups, within and across countries, and over time (Marmot and Wilkinson 1999; Berkman et al. 2014; Krieger 2011). One the of many interesting aspects of SDH research is that while the vast majority of epidemiological research is undertaken and published in the United States, much of the SDH research has been done by researchers outside the United States, mainly in Europe (Braveman et al. 2011). The explicit focal points of social epidemiological research – social phenomenon as causes and inequalities in health across social groups – are the initial entry points for philosophical investigations that are both intellectually challenging and have profound real-world implications.

The present chapter will first introduce social epidemiology and some of its insights and then discuss the implications for the philosophy of epidemiology. The subsequent sections discuss the social justice implications of SDH, and then the chapter concludes.

Social Epidemiology

The science of epidemiology is the informational engine of medicine, public health, health research, and policy. Despite the foundational role it has in these fields, epidemiologists themselves consider it to be a relatively new discipline and still in an early stage of development (Rothman et al. 2008, p. v). While there has been enormous growth starting in the 1960s in research outputs as well as in the understanding of epidemiological concepts, there are still some strong disagreements about fundamental concepts. One of the central controversies is whether epidemiology should include research on social determinants (Rothman et al. 1998; Zielhuis and Kiemeney 2001; Susser 1999; Krieger 2011). To put it simply, the debate is about whether epidemiology is and should be a natural science or a social science. Including the study of social determinants would make epidemiology more like a social science. And, in the opinion of some epidemiologists, this would make epidemiology less objective, authoritative, and scientific (Rothman et al. 1998; Zielhuis and Kiemeney 2001; Marmot 1976).

A common view understands scientific research and moral reasoning as belonging to two separate spheres, and scientific research is often treated as factual and objective inputs into the sphere of moral reasoning ("evidence-based policy"). In line with such a view, some epidemiologists see themselves as pure natural scientists discovering natural facts about biological processes, which then become inputs into the separate sphere of moral reasoning regarding which health policies to pursue through social action. Contrary to such a view which upholds a strong fact-value distinction, a "science is social" perspective recognizes scientific research as occurring within a social context. Whether in epidemiology or other scientific fields, social values and intellectual virtues of individuals are recognized as influencing scientific practice starting from which questions are researched, how hypotheses are framed, the scope of observations, how data and hypotheses are adjusted, how causal inferences are made, how findings are disseminated, and so forth. The conflict between these two perspectives on what science is and how it should be done is real and cannot be overstated. It is at the center of the acrimonious debates occurring within epidemiology for well over a decade. For a good reflection of the debates, see the article by Zielhuis et al. and commentaries in the same journal issue (Zielhuis and Kiemeney 2001).

What initiated and sustains this debate about the type of science epidemiology should be is that social epidemiologists have been producing a growing base of significant findings. Among the many productive insights from social epidemiology over the last four decades, one discovery has been particularly revolutionary. Initially discovered by the researchers of the Whitehall studies in the late 1970s, epidemiologists have been producing compelling evidence that health outcomes (e.g., life expectancy, mortality rates, obesity, cognitive development, etc.) are distributed along a stepwise, social gradient; each socioeconomic class – defined by income, occupational grade, educational attainment, etc. – has worse health outcomes than the one above it (Kawachi et al. 2002; Macintyre 1997; Marmot et al. 1997). Health and disease are not simply divided between the haves and have-

nots; there is a health/illness gradient from top to bottom of the social hierarchy within all societies. Such a stepwise gradient in health outcomes suggests a "dose-response" pathway between social determinants related to social hierarchy/inequality and health outcomes. Research also shows that the steeper the socioeconomic gradient (i.e., the more social inequality there is in a society), the lower the health of the entire population overall. Everyone in a given society is worse off in the domain of health and many other life domains than they could be otherwise – if there was less social inequality (Wilkinson and Pickett 2009; Deaton 2003).

Prior to the identification of the health gradient, social epidemiology was mainly focused on including social, economic, and cultural factors in the individual-level exposure category and seeing if there was a causal inference to be made with disease. The remarkable findings on the social distribution patterns of ill-health across groups now motivate research that seeks to explain both causation of disease in individuals and distribution across social groups. That is, unlike most epidemiological studies that try to identify risks or what causes a disease in one individual rather than other, post-gradient social epidemiology aims to identify what causes a disease in certain individuals and differing amounts of that disease in different social groups.

Researchers have so far identified a whole range of social determinants (discrete factors and pathways) to ill-health throughout the entire life cycle, starting from the social conditions surrounding the mother while the child is still in utero all the way to the quality of social relationships in old age. To be clear, health care is still recognized as being crucial to treating or mitigating ill-health, but social epidemiologists argue that the more influential *causal* determinants of health and disease include such things as early infant care and stimulation, safe and secure employment, housing conditions, discrimination, self-respect, personal relationships, community cohesion, and income inequality (Marmot and Wilkinson 1999; Berkman et al. 2014). And along with the rapid growth in knowledge about discrete social factors and pathways, a variety of explanatory theories have been proposed. While keeping in mind that most social epidemiological research so far has been done in high-income countries, Mackenbach presents a good review of the extant theories (Mackenbach 2012). The World Health Organization's Commission on the Social Determinants of Health presented an explanatory model for all societies and human beings (Commission on Social Determinants of Health 2008). More recently, a Lancet commission considered the transnational social factors that affect health and health inequalities (Ottersen et al. 2011).

Philosophy of Epidemiology

For most of the twentieth century, SDH were largely thought of in terms of material deprivations affecting the poor or as an additional factor to the proximate individual-level causal factors of biology, behavior, and exposures to harmful agents. However, in light of the many research findings and the identification of the social gradient in health in every society and across societies, SDH are now argued to be more dominant than proximate factors; SDH, in fact, shape the proximate causes. And

where one stands on the social gradient determines the types and levels of harmful exposures and protective factors in pathways to ill-health and mortality. The significance of this is that social epidemiology has the potential to produce a more general explanatory paradigm for epidemiology than the currently prevailing explanatory paradigm that focuses only on individual-level factors.

In philosophical terms, both the epistemology of causation of ill-health and the ontology of causal factors have been affected. The methods used to acquire knowledge, the causal processes we acquire knowledge about, and the qualities of the things being observed are now more expansive than before. On the one hand, we now know that the number of links in the causal chain from exposures to the onset of ill-health is larger than previously thought. We are confident of this finding even though the specific causal links and processes are just beginning to be more specified. On the other hand, multilevel analysis has opened up new dimensions in the causal chain beyond individual-level exposures. Such multilevel analysis attempts to identify the independent and interactive effects on the causal chain by determinants operating at various social levels (Kawachi et al. 2002; Subramanian and Kawachi 2004). These supra-individual levels can be that of the family, work environment, neighborhood, state, region, country, and so forth.

Such analysis of the impact of phenomena at multiple levels on individual biological functioning has motivated the use of the metaphor of "Chinese boxes" (Susser and Susser 1996a, b). Though it has limitations, the metaphor helps to visualize an etiological model of ill-health where different levels of determinants are nested within each other with the individual's biological processes in the center. The metaphor is particularly helpful in illuminating the tension between discounting the effects of determinants at each level as it becomes more distal from the individual and at the same time recognizing that each distal level significantly defines and/or constrains the determinants operating at levels nested within. This opening-up of epidemiological analysis outward and upward to include supra-individual social phenomena or contexts that influence individual biological functioning has been labeled "macro-epidemiology"(Rockett 1999). However, to put things into perspective, exponentially more resources are being channeled into research identifying determinants going in the other direction, at the molecular level. In facing persistent limitations to effective or complete knowledge of the causation of chronic diseases, there is great optimism that genetic "risk factors" are the missing pieces of the "causal pie" or the hidden links in the "web of causation" of individual impairments and mortality. The continued focus on individual-level factors and the more concerted effort to dig deeper down into the biological makeup of the individual are referred to as "micro-epidemiology" (Rockett 1999).

What is currently at play in the field of epidemiology is whether micro-epidemiology, the dominant explanatory paradigm during the second half of the twentieth century, can continue to survive as a general theory of epidemiology in the twenty-first century. In order for micro-epidemiology to survive, it must at least be able to integrate macroanalysis. The productivity of SDH research over the last few decades compels both intellectually and ethically pursuing further SDH research and the construction of an explanatory paradigm with less blind spots or "slippage." As it

now stands, the individual-level multifactorial framework, whether metaphorically described as the web of causation or a causal pie, does not recognize "nonnatural" determinants of disease and mortality. The model allocates relative responsibility for the causation of ill-health across three categories of determinants consisting of individual biological factors, individual behaviors, and proximate exposures to harmful substances. While the model does not limit the number of different links in the web of causation or pieces in the causal pie, the directions of interactions, or of timescales, all determinants must come from within the three categories. Such a causal model clearly excludes social phenomena that influence the three proximate categories of causal factors.

For example, individual biological endowments, the category that seems to be the most natural of the causal factors, in fact, can be significantly affected by social factors. Prior to an individual's birth, social phenomena can profoundly affect an individual's parents' sexual behavior, reproduction, and the quality of pregnancy, which then directly determine an individual's biological endowments and functioning (Posner 1992; Bauman 2003; Barker 2001). So, all three categories in the micro-epidemiology model can clearly be subject to social influence. Furthermore, by only recognizing individual-level factors, the model only recognizes individual variations and, therefore, cannot recognize or evaluate the distribution of health outcomes across social groups within a population or explain differences across populations.

The inability of micro-epidemiology's explanatory framework to recognize the influences of social phenomena on the three individual-level "natural" causal categories yields incomplete explanations. Alternatively, we can say that such explanations are useful only for specific kinds of causal determinants and pathways. Indeed, many effective health interventions have been based on such individual-level analyses. However, only when the causal links beyond individual-level factors – the causes of proximate causes – are allowed into the frame are we able to perceive other types of proximate natural and social causes as well as social distribution patterns (Rose 1985). If micro-epidemiology cannot integrate macroanalysis, a new general theory or explanatory paradigm for epidemiology must be found that can account for the independent and interactive effects of determinants that work at the molecular level all the way up to the global social environment (March and Susser 2006).

Health Inequality and Ethics

The link between SDH and social group inequalities in health raises ethical questions in the follow way. The interest in identifying inequalities in health across social groups for their own sake as well as to identify the social determinants of such inequalities both directly intersect concern for social ethics and justice. Any practical policy deliberations striving to identify the right social response to ill-health in individuals or groups unavoidably confront ethical questions. Health policies are profoundly political because they distribute significant and diverse benefits and burdens across individuals and groups. In contemporary health policy debates, ethical ideas are often used to justify how limited resources are distributed across

individuals and groups or for constraining individual rights. But just beyond these familiar and immediate policy questions about distribution of resources or individual liberties, there exist more fundamental questions and "wicked" problems regarding how and why there should be social interventions to address ill-health in the first place. What is it about health or ill-health that compels a social response or makes it a concern for social justice? Is it the types of causes of ill-health, the absolute levels of health achievements, their relative inequalities, or the consequences of ill-health that must be addressed as a matter of social justice? There are good reasons to believe that all of these multiple dimensions of health should matter for realizing social equity and justice (Sen 2002). Even so, how do we then morally evaluate the different dimensions of the types of causes, levels of ill-health, and consequences of ill-health in relation to each other? Which dimension should social action address first, second, and so on? Furthermore, how does the understanding of what matters about these different dimensions change when the moral concern for individuals is supplemented by concern for groups?

SDH research complicates these numerous and difficult ethical questions even further by showing how improving absolute and/or relative health inequalities requires making changes to a range of basic social practices and institutions. In light of SDH research, the scope of social intervention to address health concerns has now become much larger than just providing health care or addressing individual-level material causal factors. In fact, SDH research explodes the scope of social intervention to encompass all social environments as it strives to identify and address any and all possible social determinants of impairments and mortality. While some social determinants are such things as the social bases of autonomy, freedom, dignity, or respect, interventions to transform such determinants could mean redistributing economic resources and opportunities, material goods, as well as choices and duties of individuals and institutions. What this means is that addressing inequalities in the realm of individual or group health achievements will have to manipulate or, indeed, create inequalities in other realms of individual lives and societal functioning.

In the language of distributive justice debates, mitigating or manipulating social determinants of ill-health and mortality means that there must be a redistribution of some valued goods or "things" in different social spheres. While SDH research has provided information on some social bases of causal pathways to impairments and mortality, the literature has given little attention to the possible consequences in other non-health social spheres that would follow from transforming such causal pathways. It is often implicit in the SDH literature that the logical social response to the identification of social determinants of ill-health is to transform them. Ideally, transforming or redistributing a particular social determinant will improve health achievements which, in turn, will create even more positive social determinants. For example, engendering the social bases of dignity through creating opportunities for income and wealth could improve health achievements. Individuals who take advantage of those opportunities could in turn create more opportunities for income and wealth and thus, also, more social bases of dignity for themselves and others. Where such a virtuous circle does not exist, however, what sort of criteria shall we

use to evaluate if, when, and how trade-offs are made between improving absolute levels and relative inequalities in health functioning, and how things function or are distributed in other social realms?

In conjunction with evaluating such multiple dimensions such as causes, distribution, and consequences of ill-health, the identification of SDH means that reasoning about the right social response to health concerns must occur across multiple disciplines. Multidisciplinary reasoning is necessary in order to both identify the variety of social bases of the causal factors of ill-health and identify the potential non-health consequences in other social realms of possible interventions addressing SDH. It is important to identify how addressing various kinds of SDH will affect their respective social spheres because avoiding ill-health is only one among other goals valued by individuals and societies (Preda and Voigt 2015).

When standing within the health sector, it seems self-evident that the primary goals of health interventions are to transform the causes, levels, and consequences of ill-health. All things being equal, it may be a good thing to lessen health inequalities. Yet, as is now made more obvious by SDH research, health policies must also be cross-sector social policies. Thus, determining the right social response will require reasoning about how the moral concern for the multiple dimensions of health of individuals and groups relates to the right and just functioning of a variety of social spheres. Ideally, a general theory of social justice would provide a clear framework which would help guide social action by identifying why and how to address health concerns in relation to pursuing other social goals. However, there is no general theory of social justice that is commonly accepted within a society or across societies.

Social Justice Theory

Throughout the nineteenth and most of twentieth centuries, the dominant conception of social justice in liberal societies was framed by the philosophy of utilitarianism. Simplifying greatly, an action or society was considered to be just if it produced the greatest happiness or welfare for the greatest number of individuals. However, since the 1970s, due to the profound critiques of utilitarianism and a meaningful alternative proposed by the philosopher John Rawls, debates on alternative conceptions of social and, indeed, global justice are flourishing again (Kymlicka 2002). Utilitarian thought, however, continues to profoundly shape public policy making around the world and particularly public health policy.

All liberal theories of social justice begin from the premise of the individual as the primary unit of analysis or moral agent, and that every individual has equal moral worth. The equal moral worth of individuals is seen to arise from the capacity of human beings to reason and thereby conceive and purse a plan of life. Equal moral worth and the freedom to conceive and pursue one's life plans are seen as interrelated concepts. From this common starting point, different theories go on to articulate what that means for how society must treat the individual. This central question of how individuals should be treated by society has been transformed into the question

of what should be distributed to individuals. The reason why treatment has turned into distribution is because social contract theories have had profound influence on liberal conceptions of social justice, of which John Rawls's theory is the most recent (Rawls and Kelly 2001; Rawls 1971). Such influence has meant that liberal social justice, or how society should justly treat its members, is predominantly understood as being a conception of how to distribute the benefits and burdens of social cooperation fairly across individuals (Brighouse and Robeyns 2010).

In reviewing the range of alternative theories, Amartya Sen has argued that the various modern conceptions of liberal social justice can be understood to differ most fundamentally according to the "thing" that is valued and how the theory distributes that thing across individuals (Sen 1992). Among the range of different theories of social justice, the things to be distributed include welfare (preferences, objective welfare), resources (income, primary goods, personal and impersonal resources, negative liberties), or capabilities (basic capabilities, ten central human capabilities). Underlying both the identification of the things and the distribution schemes is the profound concern for inequality. Each of the different theories provides reasoning as to how the equal moral worth of individuals allows or disallows inequalities in different aspects of lives of individuals thought to be relevant to social justice. Importantly, what has come to be accepted is that equal respect and concern for every individual does not necessarily mean the distribution of things equally to individuals (Daniels 1996; Sen 1992; Clayton and Williams 2002).

Despite the resurgence of philosophizing about social justice over the past five decades, only within the last two decades have the concerns for health, health inequalities, and SDH been given significant attention by social justice philosophers. One explanation may be that the philosophers like most others also thought ill-health was caused by the natural lottery of biology, personal behaviors, and proximate exposures. SDH shifts both the causal story and the moral responsibility of ill-health from the individual and nature squarely onto social institutions and choices. Various philosophers have sought to rise to the challenge of developing a theory of social justice or health justice that takes account of SDH (Sen 1999; Daniels 2008; Powers and Faden 2008; Venkatapuram 2011; Weinstock 2015). The comparative evaluation of these theories is just beginning.

Conclusion

This chapter has outlined two kinds of philosophical issues raised by SDH. One set relates to the philosophy of science and epidemiology, and the other relates to social justice theorizing. Given the emerging nature of the debates, the chapter aimed to present an introduction to the major philosophical issues rather than specific issues. It was argued that the starting points of the study of the social and the concern for inequality in the SDH research lead immediately to a rich and complex set of philosophical questions that are only just beginning to be given concerted attention. There is much to be done.

Definitions of Key Terms

Social determinants of health	The causes of the proximate causes of disease.
Health gradient/social gradient in health	Health outcomes follow the social gradient. The higher the social position of individuals and groups, the better the health outcomes.
Multilevel analysis of health determinants	Analysis of factors that operate at different social levels such as family, neighborhood, state, country, etc.
Distributive justice	An area of social justice philosophy that identifies and values things related to human well-being and rules for their distribution.

Summary Points

- Social determinants of health are causes of the causes on or within the body that causes disease.
- These factors challenge the existing scope, methodologies, and purpose of the science of epidemiology.
- These factors raise questions about inequality and social justice.
- Philosophical reasoning is needed both to improve the science of epidemiology and to identify the appropriate social responses.

Acknowledgements I would like to thank John Wiley and Sons for permission to draw on an article published in Bioethics journal titled Epidemiology and Social Justice in Light of Social Determinants of Health.

References

Barker DJP (2001) A new model for the origins of chronic disease. Med Health Care Philos Eur J 4:31–35

Bauman Z (2003) Liquid love: on the frailty of human bonds. Polity Press/Distributed in the USA by Blackwell Pub, Cambridge, UK/Malden

Berkman LF, Kawachi I, Glymour MM (2014) Social epidemiology. Oxford University Press, Oxford

Blaxter M (2010) Health. Polity, Cambridge

Braveman P, Egerter S, Williams DR (2011) The social determinants of health: coming of age. Annu Rev Public Health 32:381–398

Brighouse H, Robeyns I (eds) (2010) Measuring justice: primary goods and capabilities. Cambridge University Press, Cambridge

Clayton M, Williams A (2002) The ideal of equality. Palgrave, Basingstoke
Commission on Social Determinants of Health (2008) Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health. World Health Organization, Geneva
Cribb A (2005) Health and the good society: setting healthcare ethics in social context. Clarendon, Oxford
Daniels N (1996) Equality of what: welfare, resources, or capabilities? In: Justice and justification: reflective equilibrium in theory and practice. Cambridge University Press, Cambridge, UK/New York
Daniels N (2008) Just health: meeting health needs fairly. Cambridge University Press, Cambridge/New York
Deaton A (2003) Health, inequality, and economic development. J Econ Lit 41:113–158
Diener E, Suh EM, Lucas RE, Smith HL (1999) Subjective well-being: three decades of progress. Psychol Bull 125:276–302
Huppert FA, Baylis N, Keverne B (2005) The science of well-being. Oxford University Press, Oxford
Julia C, Valleron AJ (2011) Louis-Rene Villerme (1782–1863), a pioneer in social epidemiology: re-analysis of his data on comparative mortality in Paris in the early 19th century. J Epidemiol Community Health 65:666–670
Kawachi I, Subramanian SV, Almeida-Filho N (2002) A glossary for health inequalities. J Epidemiol Community Health 56:647–652
Krieger N (1994) Epidemiology and the web of causation: has anyone seen the spider? Soc Sci Med 39:887–903
Krieger N (2011) Epidemiology and the people's health: theory and context. Oxford University Press, New York
Kymlicka W (2002) Contemporary political philosophy: an introduction. Oxford University Press, Oxford/New York
Macintyre S (1997) The Black Report and beyond: what are the issues? Soc Sci Med 44:723–745
Mackenbach JP (2012) The persistence of health inequalities in modern welfare states: the explanation of a paradox. Soc Sci Med 75:761–769
March D, Susser E (2006) The eco- in eco-epidemiology. Int J Epidemiol 35:1379–1383
Marmot M (1976) Facts, opinions and affaires du coeur. Am J Epidemiol 103:519–526
Marmot MG, Wilkinson RG (1999) Social determinants of health. Oxford University Press, Oxford/New York
Marmot M, Ryff CD, Bumpass LL, Shipley M, Marks NF (1997) Social inequalities in health: next questions and converging evidence. Soc Sci Med 44:901–910
Nordenfelt L, Khushf G, Fulford KWM (2001) Health, science, and ordinary language. Rodopi, Amsterdam
Ottersen OP, Frenk J, Horton R (2011) The Lancet-University of Oslo Commission on global governance for health, in collaboration with the Harvard Global Health Institute. Lancet 378:1612–1613
Posner RA (1992) Sex and reason. Harvard University Press, Cambridge, MA/London
Powers M, Faden RR (2008) Social justice: the moral foundations of public health and health policy. Oxford University Press, New York/Oxford
Preda A, Voigt K (2015) The social determinants of health: why should we care? Am J Bioeth 15:25–36
Rawls J (1971) A theory of justice. Harvard University Press, Cambridge
Rawls J, Kelly E (2001) Justice as fairness: a restatement. Harvard University Press, Cambridge, MA
Rockett IRH (1999) Population and health: an introduction to epidemiology. Population Reference Bureau, Washington, DC
Rose G (1985) Sick individuals and sick populations. Int J Epidemiol 14:32–38
Rothman KJ, Adami HO, Trichopoulos D (1998) Should the mission of epidemiology include the eradication of poverty? Lancet 352:810–813

Rothman KJ, Greenland S, Lash TL (2008) Modern epidemiology. Wolters Kluwer Health/ Lippincott Williams & Wilkins, Philadelphia
Ryan RM, Deci EL (2001) On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 52:141–166
Sen A (1992) Inequality reexamined. Harvard University Press, Cambridge
Sen A (1999) Development as freedom. Knopf, New York
Sen A (2002) Why health equity? Health Econ 11:659–666
Subramanian SV, Kawachi I (2004) Income inequality and health: what have we learned so far? Epidemiol Rev 26:78–91
Susser M (1999) Should the epidemiologist be a social scientist or a molecular biologist? Int J Epidemiol 28:S1019–S1022
Susser M, Susser E (1996a) Choosing a future for epidemiology: I. Eras and paradigms. Am J Public Health 86:668–673
Susser M, Susser E (1996b) Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. Am J Public Health 86:674–677
Trostle JA (2004) Epidemiology and culture. Cambridge University Press, Cambridge, UK/New York
Venkatapuram S (2011) Health justice. An argument from the capabilities approach. Polity Press, Cambridge
Virchow RLK, Rather LJ (1985) Collected essays on public health and epidemiology. Science History Publications, Canton
Weinstock DM (2015) Health justice after the social determinants of health revolution. Social Theory Health 13:437–453
Wilkinson RG, Pickett K (2009) The spirit level: why more equal societies almost always do better. Allen Lane, London
Zielhuis G, Kiemeney L (2001) Social epidemiology? No way. Int J Epidemiol 30:43–44

Further Reading

A popular and accessible overview of the science of SDH is presented in Wilkinson R, Marmot MG (1998) Social determinants of health: the solid facts. Centre for Urban Health/World Health Organization/Regional Office for Europe, Copenhagen. For an economist's view on SDH across time and geography seen Deaton AA (2013) The great escape: health, wealth, and the origins of inequality, Princeton University. The journals of Public Health Ethics, Bioethics and International Journal of Epidemiology are good sources for the most recent literature on the issues presented.

Health Promotion in Public Health: Philosophical Analysis

67

Peter Allmark

Contents

Abstract

Health promotion can reasonably be viewed as a major element in public health work. The latter was defined around a century ago as "the science and art of preventing disease, prolonging life and promoting health through the organized efforts of society." Health promotion involves (i) health education, such as advertising; (ii) illness prevention, such as screening; and (iii) legislation, such as banning smoking in public places. Although it has older roots, it is largely a phenomenon of the mid-twentieth century and beyond. Three factors stimulated

P. Allmark (✉)
Centre for Health and Social Care Research, Sheffield Hallam University, Sheffield, UK
e-mail: p.allmark@shu.ac.uk

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_53

its development. The first was the development of epidemiology and in particular the work showing the link between smoking and illness followed by success in reducing smoking in the population. The second was the increased cost of health care in its standard form of illness treatment. And the third was a concern that despite improvements in health and, in the UK, the inception of a National Health Service, the inequality in health status between rich and poor remained and even grew. The philosophical questions concerning health promotion fall into three categories: the philosophy of science, ethics, and political philosophy.

Introduction

This chapter sets out the major areas of philosophical interest and controversy in the practice of health promotion and, by extension, of public health. It begins with the definition and delineation of the central concepts before setting out their historical roots. The rest of the chapter is concerned with the main task. The areas of philosophical interest it discusses concern the philosophy of science, ethics, and political philosophy.

Definition and Delineation

In the *Bangkok Charter for Health Promotion in a Globalized World*, the World Health Organization (WHO) defines health promotion as

> the process of enabling people to increase control over their health and its determinants, and thereby improve their health. It is a core function of public health and contributes to the work of tackling communicable and non-communicable diseases and other threats to health. (p. 1)

Note that this situates health promotion as a subset or "core function" of public health. In turn, public health work is defined in the Acheson Report as "the science and art of preventing disease, prolonging life, and promoting health through the organised efforts of society" (Winslow 1920). Commonly, public health work is divided into the three categories:

- Health improvement – for example, by improving housing stock or encouraging healthy lifestyles
- Improving services – for example, by reducing waiting lists
- Health protection – for example, by reducing environmental health hazards or planning for pandemics (Griffiths et al. 2005)

Acheson's definition needs an additional limitation, which is that the "organized efforts" must in some way be directed at preventing disease and so on. Without that limitation, almost any activity could constitute public health work; for example, a television station showing a good comedy might improve the nation's health, but

this is not part of its intention and as such it is not public health work. The WHO definition of health promotion also looks broad, and it would arguably seem to cover all three categories of public health work. Indeed, a widely used three-way categorization of health promotion activity seems to cover similar areas to that given of public health work:

- Health education – such as putting health messages in public places
- Illness prevention – such as screening and vaccination
- Legislation – such as banning environmental tobacco smoke in public places (Doxiadis 1987)

The distinction between public health and health promotion is only of import where it is used, say, to mark out areas of responsibility for different groups of professionals. In itself, not too much rests on it philosophically. As such, philosophical problems relating to health promotion will most likely relate also to public health. From here on, therefore, this chapter will use both terms with little distinction between them.

Roots of Health Promotion

The idea that one can affect people's health by altering factors in their behavior or in the environment is ancient. Aristotle talks of the virtuous agent indulging in pleasures only to the extent that they are "conducive to health and vigour" (Aristotle 2000, 1119a). The Greeks also were aware of a distinction in the activity of health carers between prevention and cure (Kleisiaris et al. 2014). And both the Romans and Greeks were aware of the possibility of biological warfare through, say, poisoning a water supply (Roffey et al. 2002). In the Victorian era, Dr. Snow's using of epidemiological research to advocate closing of the Broad Street pump in London's Soho area to combat an outbreak of cholera is an early and often cited example of the involvement of health professionals in health promotion (Smith and Ebrahim 2001). However, the idea of the systematic involvement of health professionals is probably best seen as a postwar phenomenon.

A central driver in this development was Doll and Bradford Hill's research showing the strong link between smoking and lung cancer (in 1950) and subsequently between smoking and heart disease (in 1954). The results were "compelling and unexpected" (Richmond 2005); a widespread social activity that was not obviously noxious (in the way that excessive drinking or drug-taking is) was shown nonetheless to be seriously harmful to the population's health. It took some time for the research to affect public behavior, but as it did so health improved; thus was manifest the possibility of preventing illness through population behavior change. And the search for other health-affecting behavior that could be altered began in earnest.

Alongside this, at least two other drivers can be detected. The first relates to the cost of health care. Maynard coins the term the "Nye Bevan fallacy" to indicate the idea that as a nation spends more on health care, so the health of its population improves, and over time the demand for and cost of health care fall (Maynard and Sheldon 2001).

It is Bevan's fallacy as it was one of the beliefs that lay behind the setting up of the UK National Health Service in 1947 (when Aneurin Bevan was the Health Secretary). Green and Kreuter suggest health promotion came to the fore in the 1960s in what they term the era of cost containment (Green and Kreuter 1991). Governments recognized that health care was increasingly expensive. The problem seemed to be that control of contagious diseases led to the emergence of other diseases (the so-called diseases of affluence) such as cancer and heart disease. Often these could be traced to people's behavior. Thus, in 1976 a UK Government paper, produced under the direction of the then Health Secretary David Owen, says,

> Much ill-health in Britain today arises from over-indulgence and unwise behaviour. Not surprisingly, the greatest potential and perhaps the greatest problem for preventive medicine now lies in changing behaviour and attitudes to health. The individual can do much to help himself, his family and the community by accepting more direct responsibility for his own health and wellbeing. (Department of Health and Social Security 1977, p.39)

This theme of personal responsibility is repeated in Government documents and beyond throughout the 1980s.

At the same time, a second and quite distinct driver of health promotion emerged; this was associated more with the political left where the personal responsibility driver was associated with the political right. This driver is related to health inequality. The Black Report into health inequality was commissioned under a Labour Government and was famously ignored when it was published under a new Conservative Government in 1978 (Black et al. 1982). It was followed by other reports showing the same phenomenon, most recently, the Marmot Review (Marmot 2010). This phenomenon is that in relatively prosperous Western countries, the so-called diseases of affluence fall disproportionately on the poorest. In fact, the phenomenon can be generalized further: in any society with wealth inequalities, there will be health inequalities; whatever harms people in such a society will harm the poor most. For those concerned to reduce health inequality, health promotion was seen as a potential tool. Thus, both political right and left saw value in health promotion: the right emphasizing personal responsibility and the left social determinants of health such as poor housing and education.

The critics of health promotion were primarily from the political libertarian right. This critique is considered below in the sections on the philosophical issues associated with health promotion. These issues are discussed under three broad headings: philosophy of science, ethics, and political philosophy. Each section is discrete such that readers can skip, say, the philosophy of science if their interest is in ethics or political philosophy.

Philosophy of Science

One of the central questions in philosophy of science is whether and if so how scientific method delivers knowledge. Health promotion is grounded in epidemiology, the science concerned with patterns of health and illness in the population.

For example, epidemiology showed that lung cancer was linked to smoking. The job then for practitioners of health promotion is to find interventions that successfully reduce smoking in the population, such as health education and bans on smoking in public places. As with other health interventions, these should be judged for effectiveness using the principles of evidence-based medicine, for example, and, in particular, by a randomized controlled trial (RCT). If shown to be successful, the intervention should then be further judged for cost-effectiveness before being implemented. Thus, while health promotion begins with epidemiology, it also uses social sciences such as those relating to education and psychology. This picture could be said to mirror other areas of health care. For example, the treatment of cancer is grounded in the biology of cancer and oncology but also in the sciences of pharmacology, surgery, and radiation, which tell us how it might be treated, and of statistics, which tells us how these treatments might be tested. And both are then subject to evaluation by the science of health economics.

Despite this similarity in the structure of science and practice between cancer treatment and health promotion, it is notable that there is a difference in what might be termed the hardness of the sciences involved. In the case of cancer treatment, almost all the science involved is hard, natural science. There is room for social science, particularly in relation to health economics and also to the question of compliance with treatment, but at the core is natural science.

With health promotion, almost the opposite applies. In epidemiology, the data are often woolly and unreliable, for example, people are inclined to lie or deceive themselves about their intake of tobacco, alcohol, and food (Smith and Ebrahim 2001). Furthermore, the data for epidemiology are situated in society, which is an open system. An open system is one in which there are interactions between the internal elements and the outside environment. For example, if you seek to isolate smoking as a cause of lung cancer, you are faced with numerous problems of confounders; compared to nonsmokers, smokers might live in more polluted areas, have different diets, drink more coffee, and so on. Any one of these, or a combination, might be the true cause of lung cancer rather than smoking. Davey-Smith and Ebrahim (2001, p. 5) suggest that statistical adjustment in population studies for a few potential confounders "fails to recognise the complexity of the reasons why people differ with regard to particular and general characteristics of their lives." A recent example is hormone replacement therapy (HRT) which was repeatedly shown to be cardioprotective in epidemiological studies but which randomized controlled trials (RCTs) have shown to be the reverse (hence, Davey-Smith and Ebrahim ask "Is this the death of observational epidemiology?"). By contrast, research in oncology is largely done in a relatively closed environment, such as in cellular research or animal models.

One response to the HRT example is to suggest that health promotion needs to be more like the rest of evidence-based medicine, adopting the RCT as the gold standard for evaluating interventions. This is problematic. Open systems tend to be resistant to control. Indeed, one criticism of RCTs in standard medical treatment is that they show only that a treatment works (or not) in carefully controlled situations, not in the open environment in which they will actually be given.

An alternative to RCTs is to embrace methods that are designed for open systems, such as logic models (Allmark et al. 2013; Davies et al. 2006; Murphy et al. 1998). The work of philosophical realists, such as Pawson, has been influential in the development of methodology for open systems (Pawson and Tilley 1997; Pawson 2013). To the extent that health promotion takes on this approach, it is set apart from the more empiricist RCT-focused methods in other parts of health care. However, there is a case for introducing realist methods more widely: RCTs produce results that are often not replicated in practice perhaps because the control involved in them renders them inapplicable to the real world (Craig et al. 2012).

Ethics

The most widely discussed ethical issue in health promotion concerns the balance of liberty against intervention; this is covered in the next section under the heading of "Political Philosophy." This section discusses two related issues that appear in the literature and which are more tightly focused on specific health promotion strategies. The first concerns hidden harms in treatment and, the second, the problem of treating populations rather than individuals.

Hidden Harms

Skrabanek, a central libertarian critic of health promotion, asks "Why is preventive medicine exempted from ethical constraints?" (Skrabanek 1990). His question is prompted in part by the perception that health promotion interventions are judged only to be of potential benefit and hence are not tested rigorously for harm. Unfortunately, health promotion can harm. Allmark et al. looked at how public health education initiatives relating to smoking had been evaluated between 1992 and 2004 (Allmark et al. 2010). They found that the evaluation was done purely on behavior, namely, whether people stopped smoking or did not take it up; if so, the initiative was judged successful. However, in their own separate research, they examined people who had shown signs of lung cancer but had been late to present to health-care professionals. At that time, the UK had a poor record of late presentation with lung cancer. In qualitative interviews, patients reported a number of ways in which information they had gathered from health education had influenced their late presentation. For example, ex-smokers reported being told and believing that their risk of lung cancer would revert to that of a nonsmoker if they gave up; nonsmokers believed they could not get lung cancer; and smokers believed they could get lung cancer but that nothing could be done if they did. All these beliefs are false, had developed in the light of health education, and had influenced late presentation. Clearly the harm done to these few individuals might be outweighed by massive health benefits to those persuaded to behave differently, but for any other treatment such harm would be looked for in evaluation – hence Skrabanek's

question. Allmark et al.'s (2010) finding is supported in a review of unintended harm associated with public health interventions (Allen-Scott et al. 2014).

Treating Populations

The previous example might be characterized as a concern about utilitarian reasoning; that harm to a few is justified by good to many. The concern carries over into the notion of treating whole populations rather than individuals. A particular example is Rose's population strategy (Rose 2001). Rose points out that "a large number of people at a small risk may give rise to more cases of disease than the small number who are at high risk" (p. 431). Thus, the most effective health promotion intervention will target the whole population and try to get everyone to, say, reduce their fat intake, even those whose intake does not put them at high risk of coronary heart disease. In this way, the whole Poisson (bell curve) distribution curve will shift to the left; those at high risk will be at lower risk; those at low risk will be at virtually no risk. There are at least two criticisms of this. The first is that risk may sometimes rise at the low end. For example, those who reduce their alcohol, fat, and BMI levels beyond a certain point may increase their risk of illness (Adams and White 2005). The second is that it is wrong to ask people to change their behavior when it is not high-risk behavior: if a man drinks 21 units of alcohol per week, which is not associated with risk to health, one should not ask him to reduce it for the sake of a population goal. Similar questions of balancing individual good against that of the population arise in other areas of health promotion such as vaccination, fluoridation of the water supply, and screening (Nuffield Council on Bioethics 2007).

Political Philosophy

Libertarianism

The central philosophical debates in health promotion and public health have been in the area of political philosophy and, in particular, the legitimate role of the state or government. Broadly the division is between those who view the role to be to protect citizens' liberty, largely through protecting the operation of the free market, and those who view the role as greater than this, perhaps to create the conditions in which citizens can flourish. This division is in turn based upon a difference in view between those who believe that individual flourishing is best assured through individual liberty and those who believe more is required, such as the provision of a minimum level of external goods. In 2004, the UK Department of Health characterized the discussion in the following way:

> While there were many notable successful public efforts … too often work to tackle longstanding, intractable or emerging problems was increasingly caught up in a sterile

> national debate . . . that created a false dichotomy between those proposing a heavy handed nanny state on one hand, and those supporting inactivity bordering on neglect in the name of individual freedom on the other. (Department of Health 2004) (Paragraph 6 of Executive Summary)

Arguably, however, there were few on the "nanny state" side while the "bordering on neglect" side set the agenda; discussions tended to be either statements of or responses to libertarian criticism of health promotion. Critics of health promotion largely emanate from the political right, especially the libertarian element. The right-wing think-tank the Social Affairs Unit and the pro-smoking lobby group FOREST were both involved in documents and campaigns against health promotion. Well-known right-wing authors are James Le Fanu and Petr Skrabanek (Skrabanek 1990, 1992; Le Fanu 2011). A small left-wing group in the UK also criticized health promotion. Their origins were in the Revolutionary Communist Party but are now linked to the Institute of Ideas, in London, and the Spiked website. Its key representative in this area is Michael Fitzpatrick, a London GP and author (Fitzpatrick 2001). It too has a libertarian agenda. However, Fitzpatrick also argues that health promotion is being used as a Trojan horse for equality and socialism but that it is ineffective in that role.

The starting point for the libertarian critique of health promotion is Mill's Liberty (or Harm) Principle:

> the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right. . . The only part of the conduct of anyone, for which he is amenable to society, is that which concerns others. In the part which merely concerns himself, his independence is, of right, absolute. Over himself, over his own body and mind, the individual is sovereign. (Mill 1972, p.78)

This powerful and compelling principle seems to have important implications for health promotion. The most obvious are in relation to legislation, the third category of health promotion in Doxiadis's list (see above). Some health promotion legislation prevents individuals from harming others, that relating to infection, pollution, and work safety, for example. But much seems to be there to prevent individuals harming themselves; examples include legislation against taking recreational drugs and enforcing the wearing of seat belts and crash helmets. Some legislation rests between these; the ban on environmental tobacco smoke was defended for its effect both on the health of passive smokers and also on smokers themselves.

One response to this has been to say that many self-harming activities do in fact harm others. Motorcyclists without crash helmets endanger their own lives but also endanger those who witness or attend the accidents, their families who endure the outcome, and the carers who look after the survivors. The small infringement of liberty is justified to avoid this overwhelming harm, particularly if one factors in also the cost of treating survivors. Feinberg in a work that follows Mill's Principle

through to a number of conclusions about the role of the State considers the crash helmet issue in detail (Feinberg 1986). The problem lies in part with the loose nature of the terms employed in the discussion, particularly harm. Feinberg warns against taking an overgenerous definition of harm such that bans become justified if, for example, they harm people by offending or upsetting them. As a consistent libertarian, he finds it hard to conclude in favour of the ban although he expresses unease (see also Dworkin 1988).

However, consistent libertarianism is a minority view, particularly outside the USA, and is based in part at least on a misreading of Mill (Crisp 1997; Nuffield Council on Bioethics 2007). Arguably it has had undue influence on health promotion discussion, perhaps in part due to the influence of the tobacco industry. A consistent libertarian will find all state-funded health promotion objectionable as it involves coercing people to spend money (via taxes) on a product they do not choose. But such a libertarian will also find government health spending in general objectionable for the same reasons. For a libertarian, individuals should be free to harm themselves but also must make their own decisions about preparing for the consequences; if they do not insure themselves for health care, they will not get it when they become ill. Anyone who believes the government should provide some level of health care is not a full-blown libertarian. And once health-care provision is accepted, there seems no reason to believe the government might justifiably also have a role in protecting the health of the population.

Autonomy and Positive and Negative Liberty

For non-libertarians, then, the question is not whether some kind of state interference in the population health is justified but rather how much and of what type. Broadly, the scope of opinion ranges from the Liberal (Seedhouse 1997) to the communitarian and Aristotelian (Allmark 2005; Buchanan 2006; Eriksson and Lindström 2008). Liberals are committed to something like Mill's view of the good life for human beings being constituted in "experiments in living" through which an individual finds happiness. The State should ensure that such experiments are possible but should interfere little in them. By contrast, communitarians have what is sometimes termed a "thick" view of the good. This is that the good life for all human beings has a large amount that is shared both as and between individuals. In other words, most people need similar things to live well and all need to do so as part of a well-functioning community. In order to achieve this, State interference to some degree is justified. Applied to health promotion, this means that Liberals would tend to disallow State control or limitation of self-harming activities and communitarians and Aristotelians to allow it to some degree. The stewardship model of the state adopted by the Nuffield Council on Bioethics discussed below probably errs on the side of the communitarian.

One way of addressing the issue of limiting or trying to alter people's unhealthy choices has been to raise doubts about the idea of liberty and, in particular, the notion of autonomy, or self-rule (Allmark 2008; Cheung and Yam 2005; Dworkin 1988;

Lindley 1986; May 1994; Mcknight 1993; O'Neill 2001). Some people's self-harming decisions can be badly informed; Mill finds it acceptable to, for example, prevent someone from taking a route that he does not know is dangerous because a bridge might collapse. Once that person is informed, however, should the decision be left to him? This might depend on his state of mind; if suffering a psychosis he believes he can fly over the gap, again it would be reasonable to prevent him. What if he wants to die? What if he wants to impress his friends with his bravery (or foolhardiness) in stepping right up to the edge? Some authors say that the respect due to people's free choices should be proportional to the extent to which those choices are truly their own; they should be unencumbered by excessive emotion (as when a heartbroken teenager wants to harm himself) and by undue influences from others (as when someone refuses a medical treatment to comply with his parents' but not his own religious beliefs). On such an account, banning recreational drugs or enforcing crash helmets is justified on the basis that these are the right choices for individuals which they would make if they were fully rational, or fully themselves, unencumbered by emotions and other influences.

Isaiah Berlin calls this an "inner citadel" view of people, the idea that inside the person who is actually thinking and choosing is their true self, what they would be if fully unencumbered (Berlin 1969). He uses this notion in a contrast between what he terms positive and negative liberty. Negative liberty is what people are or should be allowed to do without interference. Positive liberty is more like self-determination, free of controlling influences that manipulate, pressure, or misinform you to have the desires you do (Christman 1991). Berlin accepts that positive liberty is part of being free; it is no good being able to make a wide range of choices if the choice you actually make is the product of others' manipulation of you. However, he sees danger in the idea. Berlin has in mind examples from politics such as Marx's concept of false consciousness; this is roughly the idea that the nonrevolutionary beliefs most working class people actually have do not reflect the beliefs they would have once enlightened by Marxism. The political outcome of this is that governments may oppress people's real choices on behalf of their hypothetical ideal ones.

The implication for health promotion philosophy is that caution should be exercised in any use of the idea of a divided self as the justification for the inhibition of negative liberty, such as bans of self-harming activities or the use of techniques to "bring out" positive liberty. It seems unlikely there is an easy solution here. Berlin is not opposed to positive liberty, but he is aware of its dangers. Helping people to overcome alcohol or drug addiction looks like health promotion that promotes positive liberty without compromising negative liberty. By contrast, some notions of so-called empowerment, such as that whereby people are judged to be empowered only once they make the right choices, are more questionable (Allmark and Tod 2006).

Nudge

The recent turn to the use of behavioral intervention techniques (or "nudge") presents related issues (Department of Health 2010). Psychological study shows

that people's choice-making is often nonrational, that in decision making, people often use heuristics that bypass reasoning. For example, people opt for the status quo, or follow the herd, or overly discount the future in favour of the present (Thaler and Sunstein 2009). How we choose, therefore, is often more a product of our environment than our reasoning. Advertisers and sellers make use of this notion of choice architecture in manipulating consumers; a supermarket plays fairly fast music to encourage rapid buying, and has a bakery to put out the scents that make people hungry, for example. This being so, the proponents of nudge suggest that the same techniques could be used for more beneficial ends, such as encouraging people to save more for their pensions, to pay their tax on time, and to eat more fruits and vegetables. This is something that has been taken up in a number of Western countries, with some success (see examples on the website of the Behavioural Insights Team). The concern of course is whether using unethical methods for a good end is nonetheless unethical. For instance, if we view health promotion interventions as akin to health treatments, then it looks as though these are being undertaken without informed consent, something usually considered objectionable except in emergencies. The Lords Committee set up to look at nudges suggested two questions by which to judge the acceptability of a nudge style intervention (Science and Technology Select Committee 2011):

(i) Is it visible in principle? Earlier it was noted that supermarkets sometimes introduce smells, such as coffee or baking, to increase customers' spending. Although usually unnoticed, it is not hidden. In the same way, environmental changes designed to encourage people to, say, take the stairs or buy more vegetables are not hidden. If an intervention is visible in principle, it is less worrying than one that is not, such as subliminal advertising (if such a thing exists).
(ii) Is it proportionate? This question is one asked of all health-care interventions and is usually answered in terms of risk of harm versus chance of benefit set against cost. However, there is an additional issue here which is whether the bypassing or manipulation of people's choice is proportionate to the gain made. The Nuffield ladder (discussed below) offers a tool that can be used in making this judgment.

To these two questions, Allmark and Tod (2014, p. 114) add:

(iii) "Is the end unequivocal or disputed? The ends sought through some nudges are unequivocal; no reasonable person would prefer environments in which, say, they were more likely to insert their credit card in the wrong way, or forget to turn off the gas when leaving home. Where this is so, it counts in favour of the nudge. Other ends are disputed. Some who smoke, drink or overeat might object to being manipulated towards not doing so. Other ends may be highly disputed; it seems unlikely that all young people would value the avoidance of drug taking, binge drinking and unsafe sex. The more disputed the ends, the less justified the nudge."

(iv) "Is choice-architectural design required? Doors must have handles; pension schemes must have default contribution levels; supermarkets have to put their shelves in some order; organ donation schemes have to be opt-out or opt-in. In contrast, there is no requirement to have posters informing youngsters that drug-taking is a minority pursuit, or that binge drinking exposes you to danger and ridicule. Where choice-architectural design is required it seems reasonable that the design would favour choices all or most people would prefer to make. Where there is no immediate need to change choice architecture this would seem to require a slightly higher level of justification. For example, building new housing that is naturally warm seems perfectly acceptable whilst insulating the house of someone who stoically prefers to be cold does not."

These questions are useful once it is accepted that the State has a justifiable role in public health; however, the problem of consent remains – nudges, and other health promotion interventions, look like health measures that are undertaken without the informed consent of those who receive them and, as such, seem to be unethical.

Stewardship

This issue is addressed by the Nuffield Council on Bioethics in its report "Public health: ethical issues" (Nuffield Council on Bioethics 2007). The principle it suggests is that consent is only required for interventions where health or other risks are involved. Advertisers do not seek consent before putting up posters; supermarkets do not seek consent before baking bread or playing background music. Similarly, if the State is justified in intervening to affect public health, then explicit consent is required only when the interventions are intrusive or risky. O'Neill says that

> An adequate ethics of public health needs to set aside debates about informed consent and to consider the permissible limits of just compulsion for various types of public good. (O'Neill 2004, p. 1133)

As to what these permissible limits are, the Nuffield report advocates a stewardship model of the state in which it has a responsibility to look after the important needs of its citizens. One of those needs is health. The report also discusses the types of public health intervention a steward state might undertake from, at one end, low-level provision of information to, at the other, complete enforcement of behavior by law. It develops a tool called the Nuffield Intervention Ladder (Nuffield Council on Bioethics 2007; see Fig. 1).

This is a simple and useful device although it is noteworthy that some types of nudge policy, akin to the smells in supermarkets, do not fall into any of the categories. The device also says nothing about how to decide when a State is justified in stepping up the ladder from no intervention to quite restrictive intervention. However, the Nuffield report provides principles and examples that the policy maker might find useful.

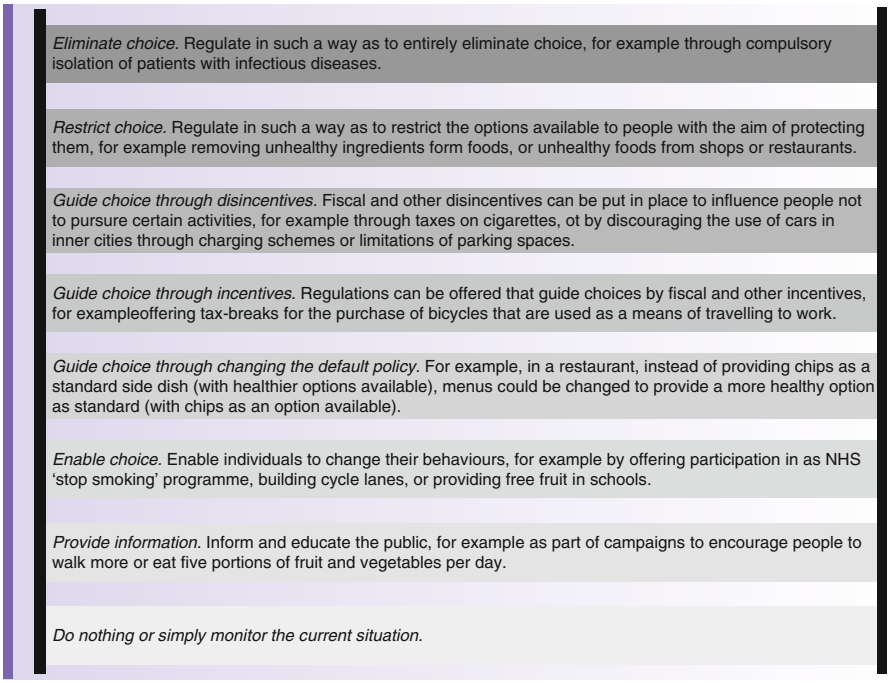


Fig. 1 The Nuffield Intervention Ladder. Source: Nuffield Council on Bioethics (2007) [with permission]

Inequality

In a section above, it was noted that the observation of health inequality was a root of the development of health promotion as an area of health practice. The hope was that, for example, education in healthy lifestyle would reduce the impact of lifestyle illnesses on those groups who suffered most, particularly the poor. Unfortunately this has not happened, and, as the Marmot Review shows, health inequality in the UK is increasing (as it is in other Western nations) (Marmot 2010). Two philosophical questions arise:

1. Why should we be concerned with health inequality at all? The thought here is that provided health is improving to all sections of the population, it is not of concern that it improves for some faster than others (Le Fanu 2011). There are at least two lines of response to this. The first is empirical: it might be suggested that health inequality has negative effects on wider social well-being. This thesis is suggested and backed up by a large range of statistics in relation to various types of inequality by Wilkinson and Pickett (2008). There may also be a concern about connectedness; infectious illness, in particular, if allowed to take hold in some sections of society, will eventually pass on. The second line

Table 1 Alternative ten tips for better health

Ten tips for better health	Alternative ten tips for better health
Don't smoke. If you can, stop. If you can't, cut down	Don't be poor. If you can, stop. If you can't, try not to be poor for long
Follow a balanced diet with plenty of fruits and vegetables	Don't have poor parents
Keep physically active	Own a car
Manage stress by, for example, talking things through and making time to relax	Don't work in a stressful, low-paid manual job
If you drink alcohol, do so in moderation	Don't live in damp, low-quality housing
Cover up in the sun, and protect children from sunburn	Be able to afford to go on a foreign holiday and sunbathe
Practice safer sex	Practice not losing your job and don't become unemployed
Take up cancer screening opportunities	Take up all benefits you are entitled to if you are unemployed, retired, or sick or disabled
Be safe on the roads: follow the highway code	Don't live next to a busy major road or near a polluting factory
Learn the first aid ABC – airways, breathing, circulation	Learn how to fill in the complex housing benefit/ asylum application forms before you become homeless and destitute

Raphael (2000, p. 362) [with permission]

of response is philosophical and is essentially that it is unjust for health to be unevenly distributed in this way; health and health care is an important human good and its distribution is at least to some extent within human control; it is therefore a matter of justice how it is distributed. Such an argument might draw on the capability approach to justice of Sen (2010) and Nussbaum (2011); see, for example, Venkatapuram (2011).

2. Is health promotion the right way to tackle health inequality? One problem with health promotion interventions is that they may even increase health inequality. Typically, well-off and educated people gain most from, for example, health education initiatives. One response to this is to invoke the social model of health. This locates the primary determinants of health in social and environmental factors rather than individual behavior. It was amusingly illustrated in a table that compares the behavior-focused recommendations from a UK Government report with some suggested alternatives based on social determinants of health (Raphael 2000) (Table 1).

The practical problem is that it is easier to address individuals' behavior rather than social issues such as those in the right-hand column. Thus, despite homage paid to the social model of health and to health inequality, much health promotion still ends up being about trying to get poor people to behave differently rather than to stop being poor. If health promotion is to succeed in tackling social determinants of health, it needs a stronger remit. This leads us to the issue of the scope of health promotion.

Scope

Consider first the well-known WHO definition of health taken from its 1946 constitution:

> Health is a state of complete physical, social and mental well-being, and not merely the absence of disease or infirmity. WHO

Then, add to this the social model of health illustrated in Dahlgren and Whitehead's famous image – see Fig. 2 (Dahlgren and Whitehead 1993).

In the light of these, it seems difficult to imagine life-enhancing measures that a State or others might undertake which could not also be described as health promotion. The provision of schools, the improvement of transport and the housing stock, and improvements in wages, couldn't these all be described as health promotion measures? Some health bodies have funded the provision of information about welfare rights and benefits as a health promotion measure, for example (Allmark et al. 2013). The UK has recently (2014) seen much of the budget for public health moved from the health service to the local government; would local government be justified in spending this money on, say, improving the condition of weather-damaged roads? In this way, health promotion may become wide and nebulous to the point of meaningless.

By contrast, Seedhouse perceives the danger that the remit of health promotion could become wide and oppressive; it could become what he terms well-being promotion (Seedhouse 1997). The broad range of factors that might improve our

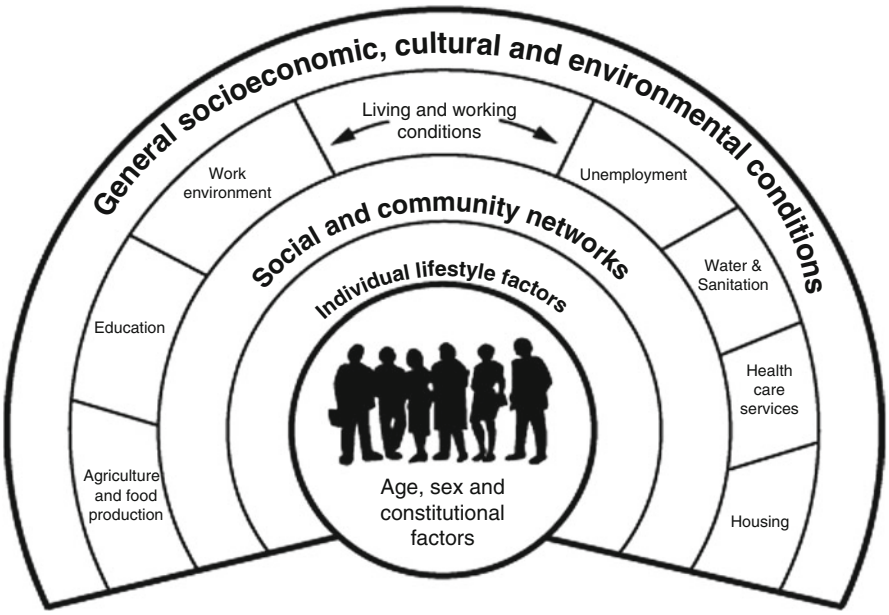


Fig. 2 The Social Model of Health. Source: Dahlgren and Whitehead (1993) [with permission]

health ranges from social equality to healthy walks in the country; is all this to be deemed suitable grounds for health promotion intervention? Seedhouse's concern is that of a Liberal faced with a communitarian or Aristotelian agenda discussed earlier. Communitarians are likely to be untroubled by the State's involvement in promoting the well-being of its population; Liberals are likely to view it as either wrong in principle or likely to fail in practice. As this chapter has attempted to demonstrate, the issue of the role of the State is at the core of much of the controversy in health promotion.

Definition of Key Terms

The definitions offered here are brief and cover only their use in relation to this chapter. Clearly, for example, there is much more that could be said in defining Aristotelian philosophy.

Aristotelian – Aristotle's ethical and political philosophy has two points of importance for this chapter. First, he views the human good (also called inter alia happiness, flourishing, eudaimonia) as having many shared elements across all people; for example, a life of intemperance or of inactivity is not good for anyone even if that is what they desire. This is sometimes called a thick view of the good and is contrasted with liberal and libertarian "thin" views of the good, which emphasize the differences between people in what constitutes a flourishing life and the role of choice in allowing people freedom to discover what works best for them. Second, he views the state's purpose as to enable human flourishing. This is liable to give it a far more interventionist role than is acceptable from a liberal or libertarian viewpoint.

Communitarian – This emphasizes the role of the community in human flourishing. It grew up in opposition to liberal (and now libertarian) viewpoints; where these see human flourishing as consisting in the autonomous actions of individuals, communitarians see it as existing where individuals are part of a flourishing community. It resembles Aristotelianism in this because Aristotle too sees people as essentially social and therefore flourishing also as having a major social element. However, some Aristotelians have distanced themselves from elements of the philosophy.

Liberal – This has been called "the most confusing term in the world" (Chang 2014, p. 68); but, for the purpose of this chapter, liberal views can be seen as moderate libertarianism. Liberals have a thin view of the good, emphasizing people's own experiments in living as the basis for their flourishing. The State has some role in this, for example, in the education of children; what should be the extent of this role is a matter of debate within liberalism.

Libertarian – Libertarian views in the context of this chapter are the antithesis of Aristotelian ones. Libertarians see the human good as comprised primarily in the freedom of people to act as they wish without interference. The State has at most a minimal role, to protect this freedom. It has no role, for example, in enforcing or

encouraging certain behavior for people's own good or in protecting them from their, perhaps foolish, decisions.

Health promotion and public health – See definitions in the chapter, under the heading "Definition and Delineation."

Summary Points

The philosophical questions concerning health promotion fall into three categories: the philosophy of science, ethics, and political philosophy.

(A) *Philosophy of Science*
The complexity of roots of causation in public health makes it resistant to scientific inquiry of an empiricist or positivist nature and more amenable to something like a realist approach, using, for example, logic modeling.

(B) *Ethics*
Hidden harms: There are concerns about the quality of evidence behind health promotion, and in particular that potential harmful effects are ignored in its evaluation.
Treating populations: Health promotion is often focused on the health of the population rather than of individuals; this can give rise to ethical concerns when treatment is based more on the health of the population than of an individual person such as in vaccination and mass medication.

(C) *Political Philosophy*
Libertarianism: Much health promotion appears to violate Mill's Harm (or Liberty) Principle that "the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good . . . is not sufficient warrant" (*On Liberty* 1.9).
Autonomy and positive and negative liberty: It might be said that a decision is autonomous only to the extent that it is consistent with the individual's personality or is rational to some degree. The danger of this argument is that it can be used to justify preventing people from doing what they want on the basis that if only they were, say, fully rational, they would not want to do so.
Nudge: Advertisers have long exploited nonrational and environmental factors to bypass our reason and get us to buy. The idea of nudge is to use the same techniques for social good, such as promoting healthy behavior. There are concerns as to whether this is acceptable given that it bypasses consent and involves manipulating people.
Stewardship: In response to libertarianism, it has been suggested that the State should have a stewardship role in which it takes some responsibility for the important needs of its citizens, including health and health care. In order to help judge what level of health promotion intervention the state should take in response to a health issue, the Council offers a tool, the Nuffield Intervention Ladder.

Inequality: Health inequality is one driver of health promotion policy. Two philosophical questions arise. (1) Why is health inequality a health problem? (2) Is health promotion the right tool to tackle health inequality?

Scope: There are two concerns about the scope of health promotion. (1) The first is that almost all social action that is aimed at improving some element in society could be deemed health promotion, for example, improving the roads. (2) Might not health promotion become well-being promotion? To some extent, those concerned about the second question are more likely to be of liberal or libertarian ilk.

References

Adams J, White M (2005) When the population approach to prevention puts the health of individuals at risk. Int J Epidemiol 34(1):40–43

Allen-Scott L, Hatfield J, McIntyre L (2014) A scoping review of unintended harm associated with public health interventions: towards a typology and an understanding of underlying factors. Int J Public Health 59(1):3–14

Allmark P (2005) Health, happiness and health promotion. J Appl Philos 22(1):1–15

Allmark P (2008) An Aristotelian account of autonomy. The J Values Inq 42(1):41–53

Allmark P, Tod A (2006) How should public health professionals engage with lay epidemiology? J Med Ethics 32(8):460–463

Allmark P, Tod A (2014) Can a nudge keep you warm? Using nudges to reduce excess winter deaths: insight from the Keeping Warm in Later Life Project (KWILLT). J Public Health 36(1):111–116

Allmark P, Tod A, Abbott J (2010) The evaluation of public health education initiatives on smoking and lung cancer: an ethical critique. In: Peckham S, Hann A (eds) Public health ethics and practice. Policy Press, Bristol, pp 65–82

Allmark P, Baxter S, Goyder E et al (2013) Assessing the health benefits of advice services: using research evidence and logic model methods to explore complex pathways. Health Soc Care Community 21(1):59–68

Aristotle (2000) In: Crisp R (ed) Nicomachean ethics. Cambridge Univ Press, Cambridge

Berlin I (1969) Four essays on liberty. Oxford University Press, Oxford

Black D, Morris J, Smith C, Townsend P (1982) The black report. Pelican, London

Buchanan D (2006) Perspective: a new ethic for health promotion: reflections on a philosophy of health education for the 21st century. Health Educ Behav 33(3):290–304

Chang H-J (2014) Economics: a user's guide. Pelican, Harmondsworth

Cheung P, Yam B (2005) Patient autonomy in physical restraint. J Clin Nurs 14(3A):34–40

Christman J (1991) Liberalism and individual positive freedom. Ethics 101(2):343–359

Craig P, Cooper C, Gunnell D et al (2012) Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health 66 (12):1182–1186

Crisp R (1997) Mill on utilitarianism. Routledge, London

Dahlgren G, Whitehead M (1993) Tackling inequalities in health: what can we learn from what has been tried? Working paper prepared for The King's Fund International Seminar on Tackling Inequalities in Health, September 1993, Ditchley Park, Oxfordshire. The King's Fund, London. Accessible In: Dahlgren G, Whitehead M (2007). European strategies for tackling social inequities in health: levelling up Part 2. WHO Regional office for Europe, Copenhagen. http://www.euro.who.int/__data/assets/pdf_file/0018/103824/E89384.pdf

Davies M, Macdowall W, Black N, Raine R (2006) Health promotion theory. Open University Press, Maidenhead

Department of Health (2004) Choosing health: making healthy choices easier. London. Available via http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/Browsable/DH_4880358
Department of Health (2010) Healthy lives, healthy people: our strategy for public health in England. Department of Health. Available via http://www.dh.gov.uk/en/index.htm
Department of Health and Social Security (1977) Prevention and health: everybody's business. Department of Health, London
Doxiadis S (1987) Ethical dilemmas in health promotion. Wiley, New York
Dworkin G (1988) The theory and practice of autonomy. Cambridge University Press, Cambridge
Eriksson M, Lindström B (2008) A salutogenic interpretation of the Ottawa charter. Health Promot Int 23(2):190–199
Feinberg J (1986) The moral limits of the criminal law: volume 3: harm to self. Oxford University Press, New York
Fitzpatrick M (2001) The tyranny of health: doctors and the regulation of lifestyle. Routledge, London
Green L, Kreuter M (1991) Health promotion planning: an educational and environmental approach, 2nd edn. Mayfield, Mountain View
Griffiths S, Jewell T, Donnelly P (2005) Public health in practice: the three domains of public health. Public Health 119:907–913
Kleisiaris C, Sfakianakis C, Papathanasiou I (2014) Health care practices in ancient Greece: the Hippocratic ideal. J Med Ethics Hist Med 7:3–7
Le Fanu J (2011) The rise & fall of modern medicine, 2nd edn. Abacus, London
Lindley R (1986) Autonomy. Macmillan, London
Marmot M (2010) Fair society, healthy lives: strategic review of health inequalities in England post 2010. Department of Health, London, Available via www.marmotreview.org
May T (1994) The concept of autonomy. Am Philos Q 31(2):133–144
Maynard A, Sheldon T (2001) Limits to demand for health care [Letter]. BMJ (Br Med J) 322:734
Mcknight C (1993) Autonomy and the akratic patient. J Med Ethics 19(4):206–210
Mill JS (1972) On liberty. Dent, London
Murphy M, Black N, Lamping D et al (1998) Consensus development methods, and their use in clinical guideline development. Health Technol Assess 2(3):1–88
Nuffield Council on Bioethics (2007) Public health: ethical issues. Nuffield, London, Available via http://nuffieldbioethics.org/
Nussbaum M (2011) Creating capabilities: the human development approach. Belknap/Harvard University Press, Cambridge, MA
O'Neill O (2001) Autonomy and trust in bioethics. Cambridge University Press, Cambridge
O'Neill O (2004) Informed consent and public health. Philos Trans R Soc London Ser B-Biol Sci 359(1447):1133–1136
Pawson R (2013) The science of evaluation. Sage, London
Pawson R, Tilley N (1997) Realistic evaluation. Sage, London
Raphael D (2000) The question of evidence in health promotion. Health Promot Int 14(4):355–367
Richmond C (2005) Sir Richard Doll obituary: epidemiologist who showed that smoking caused cancer and heart disease. BMJ (Br Med J) 331:295
Roffey R, Tegnell A, Elgh F (2002) Biological warfare in a historical perspective. Clin Microbiol Infect 8(8):450–454
Rose G (2001) Sick individuals and sick populations. Int J Epidemiol 30(3):427–432
Science and Technology Select Committee (2011) Behaviour change. House of Lords, London, Available via http://www.publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/179.pdf
Seedhouse D (1997) Health promotion: philosophy, prejudice and practice. Wiley, Chichester
Sen A (2010) The idea of justice. Penguin, Harmondsworth
Skrabanek P (1990) Why is preventive medicine exempted from ethical constraints? J Med Ethics 16(4):187–190

Skrabanek P (1992) Smoking and statistical overkill. Lancet 340(8829):1208–1209
Smith GD, Ebrahim S (2001) Epidemiology – is it time to call it a day? Int J Epidemiol 30(1):1–11
Thaler R, Sunstein C (2009) Nudge: improving decisions about health, wealth and happiness. Penguin, Harmondsworth
Venkatapuram S (2011) Health justice: an argument from the capabilities approach. Polity, Cambridge
Wilkinson R, Pickett K (2008) The spirit level: why more equal societies almost always do better. Allen Lane/The Penguin Press, Harmondsworth
Winslow C (1920) The untilled fields of public health. Science 51(1306):23–33

Psychopathy: Morally Incapacitated Persons

68

Heidi Maibom

Contents

Abstract

After describing the disorder of psychopathy, I examine the theories and the evidence concerning the psychopaths' deficient moral capacities. I first examine whether or not psychopaths can pass tests of moral knowledge. Most of the evidence suggests that they can. If there is a lack of moral understanding, then it has to be due to an incapacity that affects not their declarative knowledge of moral

H. Maibom (✉)
University of Cincinnati, Cincinnati, OH, USA
e-mail: heidimaibom@gmail.com

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1_56

norms, but their deeper understanding of them. I then examine two suggestions: it is their deficient practical reason or their stunted emotions that are at fault. The evidence supports both explanations. I conclude with an overview of the debate concerning whether they are morally or legally responsible for their actions.

Introduction

The psychopath lacks a conscience, regards others as mere means to his selfish and unscrupulous ends, experiences no empathy for others and no regret at his harmful actions, and is incapable of seeing his or others' actions as wrong or right, good or bad. These and other equally sensational claims surround psychopathy. This has led to a spirited philosophical debate about whether psychopaths even understand right and wrong or good and bad, what role their emotional and rational deficits play in such understanding, and whether they are morally or legally responsible. But what does the evidence actually show about this intriguing disorder? Below, I give an outline of the current debates and present the evidence from psychology and neuroscience as it stands.

What Is Psychopathy?

Psychopathy is a mental disorder characterized by deficient emotionality, interpersonal dysfunction, behavioral disinhibition, and antisocial behavior. Between 1 % and 2 % of the population suffer from psychopathy. Males are disproportionately affected, with an estimated four males for each female sufferer. Having psychopathy is one of the best predictors of criminal offending and reoffending; psychopaths are three times as likely as other offenders to recidivate. The average North American psychopath will have four convictions for violent crime by the age of 40. Since the prison population is estimated to contain around 20 % psychopaths, roughly 90 % of all psychopaths are either incarcerated, on probation, or on parole (Kiehl and Lushing 2014). Consequently, most of the research on psychopathy is conducted with male criminals. Two of the most common measures of psychopathy are *The Psychopathy Checklist-Revised* and *Levenson's Self-Report Psychopathy Scale*. The former is more commonly used in forensic settings, whereas the latter is an easier measure to use outside such settings.

Psychopathy is a dimensional construct, meaning that there is a relatively arbitrary cutoff point, above which someone is classified as a psychopath and below which he is not. However, the individual who does not make the cut will share many of the features of the psychopath. There is some debate about the ideal cutoff on the various scales. There is also debate about whether psychopathy should be regarded as a mental disorder at all or whether it should be thought of as an adaptation characterizing a subgroup of human beings (Hare 2004).

Deficient Affect

According to Robert Hare (2004), psychopaths have shallow affect, lack remorse, guilt, and empathy. Hervey Cleckley (1976) thought they had no shame and David Lykken (1957) that they lacked fear. Shallow affect describes an inability to experience the full range or depth of normal emotions. Whereas people normally recognize emotional words quicker than nonemotional words, psychopaths do not. They are sometimes confused about whether an event is positive or negative, and although they sometimes engage in dramatic displays of emotion, particularly anger, they often appear cold and unemotional (Hare 2004). Psychopaths trivialize the harms they do, which is one reason they are thought to lack empathy or sympathy. They tend to blame others for their own actions or failings, rarely take responsibility, and appear to experience little, or no, guilt or remorse. The prospect of pain or punishment seems not to deter them. Psychopaths do not experience stress, anxiety, or fear in the types of situations where people normally feel them, or when they do, these emotional reactions do not affect them as they would others (Lykken 1957).

Relating to Others

Psychopaths tend to think that they are better than other people and consequently that their needs and desires have priority. They do not shy away from manipulating others to get what they want whether by flattery, deception, or coercion. They are often fanciful liars, seemingly taking great pleasure in telling tall tales about their experiences and accomplishments. When caught in a lie, they appear unfazed. Psychopaths may be quite charming, but usually in rather superficial and shallow ways. Although they may speak authoritatively about technical matters, they typically do not possess the relevant knowledge although they have a certain ability to mimic experts. Psychopaths are known for having many short-term marital relationships and not to put much stock in being faithful to their lovers.

Psychopathic Lifestyle

Psychopaths commonly engage in irresponsible and impulsive behavior. For instance, they may leave infants unattended while going on a weekend bender, fail to show up at work or simply quit because they are bored, bludgeon a shop attendant to avoid paying for beer, or force a woman to have sex with them because the opportunity suddenly arises. They tend not to plan ahead, and when they do, their goals are often unrealistic. They crave stimulation, and so drug and alcohol addiction are quite common in psychopaths. They also prefer to live off others compared to making their own way.

Antisociality

Psychopaths engage in antisocial and harmful conduct mostly from an early age. They may torture defenseless animals, such a puppies, coerce other children to perform sexual acts, steal from others, frame others for their misconduct, and so on. This conduct continues or worsens in adolescence. Their criminal conduct tends to be extremely diverse compared to other criminals, and they are more likely to reoffend than are other criminals and to violate conditional release or escape from prison.

Psychopathic Subtypes

Some people argue that psychopathy is composed of clinically distinct subtypes. People distinguish between the primary, low-anxious, or callous-unemotional psychopath and the secondary, high-anxious psychopath. Many regard secondary psychopathy as a sort of hodgepodge category, which likely contains many distinct types of antisocial and emotionally dysregulated individuals. Primary psychopaths are characterized predominantly by deficient affect and a callous interpersonal style of relating to others. Many now assume that there is a genetic component to psychopathy (Blair et al. 2005), though most assume that psychopathy is a result of an interaction between genetic predisposition and a problematic early environment, often characterized by neglect or violence (Porter 1996). Deficient affect is typically used to explain why psychopaths have few qualms about harming others. But interestingly, evidence suggests that secondary psychopaths, who have relatively spared affective abilities, are more violent than primary psychopaths (Hicks et al. 2004).

Judging Right and Wrong, Good and Bad

Do psychopaths *understand* moral categories? Can they comprehend right and wrong? Are they capable of seeing their own or other people's actions as good or bad? Psychopaths can certainly *say* that it is wrong to lie, steal, murder, etc. But although the legal stance is that psychopaths have sufficient understanding of right and wrong to be held responsible for their actions, moral philosophers have long questioned whether psychopaths *really* understand right and wrong. Recently, James Blair's (1995) study of psychopaths' performance on the moral-conventional distinction has been thought to show that psychopaths lack moral understanding. As we shall see, however, the evidence is much more complex and perplexing.

The better known tests of moral competence are the Kohlberg moral stages test and Turiel's moral-conventional distinction. Kohlberg's stages are meant to measure reasoning about wrongs and rights, ranging from the so-called preconventional to postconventional. Preconventional reasoning concerns mainly how to avoid morally motivated aggression from others, such as punishment of wrongs. At the

conventional stage, people come to appreciate the importance of meeting the expectations of others, upholding the law, and fulfilling one's social obligations. At the most advanced postconventional stage, individuals' reasoning about moral rights and wrongs issue from an autonomous internalized conscience which may or may not accord with society's principles and which focuses on the application of abstract and universal moral principles. Perhaps surprisingly, one study of psychopaths' performance on the moral stages shows their performance to be superior (Link et al. 1977) and another that any deficient performance is accounted for by differences in IQ (O'Kane et al. 1996).

Turiel's moral-conventional distinction is a largely instrumental measure of moral competence, although he thinks morality mainly concerns harms, rights, and justice. The degree to which a transgression is thought to be serious, impermissible, and subject to change by authority is a mark of its being moral or conventional. Moral transgressions are judged to be more serious, less permissible, and less subject to change by a relevant authority than conventional transgressions. Blair's well-known experiment shows that psychopaths do not make a distinction between moral and conventional norms on any of these dimensions (Blair 1995). However, Blair himself failed to replicate the result (Blair 1997). Others found that psychopaths perform as well as controls on all moral transgressions except accidents (Young et al. 2012), rate moral transgressions as severe as controls do (Harenski et al. 2010), and judge actions to be wrong even if there are no rules prohibiting them (Aharoni et al. 2012).

Because both Kohlberg's and Turiel's moral tests represent disputed conceptualizations of the moral realm, other tests have been proposed. For instance, the Moral Foundations Questionnaire reflects Jonathan Haidt's more eclectic view of the moral domain. In two studies – one conducted with criminal psychopaths and another with a subclinical population with psychopathic tendencies – psychopaths were found to perform as well as nonpsychopaths on measures relating to authority, in-group loyalty, and purity. Where they were lacking in both studies were in their harm and fairness ratings; here, they performed significantly below the norm (Aharoni et al. 2011; Glen et al. 2009b). The results, however, contrast with those of other studies testing psychopaths' tendencies to make welfare justifications. Although Blair (1995) found that his psychopathic subjects made fewer welfare justifications than his other subjects, he and others have failed to replicate this result (Blair 1997; Aharoni et al. 2012).

Other areas of moral competence are thought to be affected in psychopaths, for instance their judgments concerning the permissible tradeoff between the good of the many and the harm done to the one. Here, too, the evidence is mixed. A couple of studies find increased tendency to judge that the one should be sacrificed to save the many particularly in low-anxious psychopaths (Bartels and Pizarro 2011; Koenigs et al. 2012), whereas others do not (Glen et al. 2009a; Cima et al. 2010) (for more detail, see section "Reason and Emotion").

When it comes to the behavioral data, then the findings are again mixed. Some show intact performance, some do not, but often there are more of the former than the latter. Some of the more promising studies suggesting a moral deficit are Abigail

Marsh's, and they indicate that psychopaths think it slightly less wrong to cause fear in others than nonpsychopaths do (Marsh and Cardinale 2012). Such results, however, are pretty bland compared to the rather sensational literature on psychopaths' amorality! If so many studies of psychopaths' ability to make moral judgments show that they perform as well as nonpsychopaths, why should we think that they are not able to make moral judgments?

Here is a diagnosis of the problem. Most philosophers believe that if you judge that harming others is wrong, you are thereby motivated not to harm others. This is also known as internalism about moral judgment and motivation. If this idea is right, we can use lack of motivation as a sign of lack of the corresponding judgment. Someone who regularly abuses and hurts others is very unlikely to actually believe that it is wrong to abuse or hurt others. This seems to be the case with psychopaths. Psychopaths *say* that harming others is wrong, and they justify harm norms in terms of victim welfare. Nonetheless, they are involved in an awful lot of harm (see section "What Is Psychopathy?"). This suggests that there is something about right and wrong that these individuals fail to grasp! The suspicion is intensified by the observation that psychopaths don't evince the right kind of guilt or remorse about the harm that they have caused. So whereas it is possible to imagine someone who acts against her better judgment more often than others do, it is hard to imagine that she would experience neither guilt nor remorse as a consequence. But this is exactly what the psychopath fails to do. The conclusion appears ineluctable: psychopaths are not motivated to do right and not to do wrong like others are. If they are not, then they cannot really believe that what they are doing is wrong, when they do wrong. Consequently, psychopaths do not have the ability to *truly* judge that something is right or wrong; they cannot make *true* moral judgments.

Some philosophers object to this way of thinking about morality. They think that judgments of right and wrong are independent of any motivation to act in accordance with such judgments (Brink 1989). That is not to say that they do not agree that people are *typically* motivated to act in accordance with their moral judgments, only that there is no need to doubt that someone makes a true moral judgment if he or she appears unmotivated to act in accordance with it at the same time. Some people argue that the curious constellation of relatively intact declarative moral knowledge and immoral conduct in the psychopath provides evidence in favor of such externalism about moral motivation (Aharoni et al. 2012).

Affect and Reason in Moral Judgment

If psychopaths are not able to make true moral judgments, the most obvious interpretation of their deficit is that they fail to *understand* moral demands or restrictions. There is therefore something deficient about the way that they think or reason. Others retort that psychopaths do not have any obvious reasoning deficits, wherefore it must be their lack of emotion that causes the problem. Both positions have empirical support.

Deficient Practical Reason

The balance of evidence concerning intelligence – as measured by standard intelligence tests – suggests that psychopaths have as high, or higher, intelligence as matched controls (Salekin 2006). This finding is important as low IQ is correlated with deficient performance on tests of moral aptitude and with criminality generally. Since psychopaths do not appear to suffer from obvious rational impairments, some conclude that reasoning impairments cannot be responsible for their problems telling right from wrong (Nichols 2004). That is too quick. Although psychologists and social scientists often think of reason purely as theoretical reason, the sort of reasoning ability that is relevant for most philosophers when it comes to morality is *practical*. And here psychopaths have demonstrated deficits.

In this debate, the centrality of reason to morality is typically pitched in terms of Immanuel Kant (1785/1993), whose work has had a profound impact on theories of morality. Kant thought that what he called pure practical reason must be the driving force behind categorical judgments of right and wrong. Think of practical reason as decision-making. Practically rational agents subject their reasoning to the so-called categorical imperative. Its most famous forms include the injunction not to make exception to oneself when reasoning about the permissibility of performing an action and not to use others merely as means to an end. To be able to reason in this way, however, requires being able to comprehend what it is to have ends in the first place. Psychopaths, some argue, are incapable of comprehending what an end is (Duff 1977; Kennett 2002). Others argue that while psychopaths may comprehend the notion of ends, their ability to reason practically is so impaired that they are unlikely to consider to the moral value of many of their actions (Maibom 2005).

Psychopaths do, in fact, have decision-making deficits. This is abundantly evident in the anecdotal evidence. Hare (1993), for instance, tells the story of a psychopath who decided to bludgeon a shop attendant so that he didn't have to pay for a six-pack of beers. He was going to a party and did not want to show up empty-handed, but had forgotten his wallet at home. Instead of going back to retrieve it, he assaulted the attendant who subsequently suffered severe brain damage. Psychopaths often decide to represent themselves in court – take Ted Bundy, for instance – often with disastrous consequences to themselves. The long checkered criminal record of the average psychopath also suggests a lack of long-term planning.

It is not just anecdotal evidence that suggests that psychopaths have difficulties making good decisions. There is good experimental evidence too. Psychopaths have extensive attention and inhibition deficits. When it comes to attention, psychopaths are relatively insensitive to contextual and other information that is not the focus of their attention (Hiatt and Newman 2006). They focus narrowly and exclusively on what they are attending to and are comparatively blind to other, potentially relevant features of the situation. Whereas people generally attend to multiple features of objects, actions, and situations, psychopaths typically attend to only a subclass of those. Part of the problem is that they have difficulties shifting attention from one feature of the situation to another in response to relevant contextual cues (Hiatt and Newman 2006).

Psychopaths are also relatively insensitive to punishment. If asked to navigate a maze where at each choice point, 1 of 4 choices is reinforced with an electrical shock, psychopaths are as likely to choose the option associated with shock as they are to choose the other options. Needless to say, that contrasts with the choices of ordinary people (Lykken 1957). Psychopaths are not altogether insensitive to punishment, pain, or fear of punishment. They do as well as nonpsychopaths on simple negative reinforcement tasks. It is specifically when avoiding punishment interferes with their standing pursuit of a goal that psychopaths have problems (Hiatt and Newman 2006). However, if given clear information about the reward-punishment contingencies of a task or when forced to pause before each new choice, these problems disappear.

Although psychopaths can learn simple reward-punishment contingencies, as we have seen, once they learn to respond in a certain way, they have difficulties adjusting once their behavior is no longer adaptive. For instance, switching the reward-punishment contingencies in an experiment leads to reduced performance by psychopaths. Ordinary subjects adjust relatively easily, learning to respond to previously punished stimuli and to cease to respond to previously rewarded stimuli; psychopaths do not (Newman and Kosson 1986; Blair et al. 2001).

What does this have to do with one's ability to understand right and wrong or good and bad? According to Kant, such judgments are permissibility judgments and form part of our decision-making ability. One rule of decision-making is that if you want to achieve or possess something, you must also want to do what it takes. Philosophers talk of willing the necessary means to one's ends. But the rule or rules that specifically concern an actions' moral status are about the very structure of decision-making. What is permissible for one agent in her situation is permissible for another agent in the same situation. We can understand the situation broadly so as to consist in not only the environmental features but also the role the agent plays in that situation. For instance, what is permissible for a policeman qua policeman need not be permissible for a chimney sweeper qua chimney sweeper. A policeman can arrest somebody, for instance. However, what is permissible for one policeman is also permissible for another policeman assuming that the two situations are sufficiently similar. This rule is known as the categorical imperative. When you apply it, you think something like this: "could I will that anyone in my situation do what I am considering doing?"

It is easy to see that such thinking immediately rules out acts like cutting in line and making false promises. Someone who cuts in line expects to get where he or she is going faster. However, this will only be true if *other* people do not cut in line. To consistently will to cut in line, the agent must both will that she cuts in line and that nobody else does. But if she applies the categorical imperative to her proposed action, she sees that she must will at one and the same time that anybody in her position cuts in line and that nobody *other than herself* does so. That is inconsistent. False promising leads to other problems. Let us say that I get you to lend me money by promising that I will pay you back next month. I know, however, that my finances are not going to be any better next month or the month after that, etc., and so I will not be able to pay you back. To go through with this action, I will have

to intend that anyone in need of money can falsely promise to pay them back. To make this act of lying successful, I am relying on the practice that people keep their promises. However, if no one were to keep their promises, the very practice that I am relying on would cease to exist. So I would be willing an action that depends on a practice whose very existence my action would undermine. That is not consistent either.

Another version of the categorical imperatives states that we should never only use others only as a means to our ends but also always regard them as ends in themselves. Other agents are not ours to use *merely* for our own projects. This idea seems to underlie our practice of seeking consent for various activities involving others, e.g., sexual intercourse. Clearly, psychopaths do not adhere to either of these principles. They appear to play little or no role in their reasoning. In fact, the modus operandi of psychopaths is one that relies on others adhering to moral and social rules and they do what they please. Whereas psychopaths are simply outraged by being treated badly by others and sometimes go on about it at considerable length, they tend trivialize their own bad behavior (Hare 1993). They also regard others as mere means to their ends; they are extraordinarily manipulative and exploitative.

Some argue that psychopaths do not understand what ends are or the reasons they generate (Kennett 2002). Now to understand what having an end involves, you must understand that if you want to achieve something, you must also want what is necessary and sufficient to achieving that something and that is within your power. You must also be careful that you do not foil your own achievement by adopting a course of action that will ultimately prevent you from obtaining your goal. If this understanding also involves a comprehension of the demands of the various formulations of the categorical imperative, it is easy to see how this would go over the head of most psychopaths. Others argue that since psychopaths' reasoning deficits are, in the end, relatively subtle and context-specific, we don't have evidence that they have *no* conception of ends. However, we do know that if their attention is focused on something they want, they tend not to pay a lot of attention to other features of the situation, presumably including the moral features of it (Maibom 2005). They are unlikely to subject their decision-making to the categorical imperative in any of its formulations and thus to notice how their false promising or lying conceptually undercuts the very intentions they adopt in their actions.

Deficient Emotion

It is now more popular to think that the *real* problem with psychopaths is their deficient emotionality: their lack of guilt, remorse, empathy, and so on. The idea here is that true moral judgment is infused with affect. Take that affect away, and all we are left with are hollowed out thoughts or empty words. To understand this idea, think about the difference between so-called conventional norms and moral norms. We adhere to a number of relatively arbitrary conventions because doing so makes things run more smoothly. In England and many of its former colonies, you drive on the left side of the road; mostly you drive on the right side in the rest of the world.

The decision of what side to drive on is not driven by deeper or more profound concerns. There are two options; you choose one and stick to it. Of course, once the convention holds, individuals cannot choose, willy-nilly, which side to drive on. But coming to England, you are not morally outraged by this practice, though it might be the cause of some concern when crossing the road. Contrast this with prohibitions against harming others. Such norms are near universal – although who may be harmed and under what circumstances vary considerably – and are not arbitrary in the way which side of the road to drive on is. Violations of such norms often give rise to considerable outrage. Take, for instance, the practice of family members killing rape victims in parts of the world. Contemplating such acts, and the people who perform them, typically causes a strong emotional reaction we might describe as anger or outrage. For sentimentalists the ability to experience such emotional reactions – whether they be anger, guilt, shame, sadness, and so on – in response to agents performing certain kinds of actions is necessary for, or a constituent part of, our understanding such actions as being right or wrong, good or bad.

David Hume said that when he considered people and their actions, he did not see goodness or badness there, but only inside his own chest, as it were. What he meant was that an action or agent's goodness or badness rests on how it or she makes us feel. He talked broadly of the sense of approbation for judgments of right or good and the sense of disapprobation for judgments of wrong or bad. Hume thought that our basic propensity to feel with our fellow human beings was foundational to all our moral sentiments. In other words, the basic ability to empathize with another is the source of all the other sentiments one experiences in reaction to what people do to others. It is easy to see why this idea is appealing. Why do we think it is wrong to harm others? Because when we contemplate such harm, we feel some of the pain of the victim in our own bodies by means of the empathic affective reaction. This basic propensity, then, gives rise to other, more recognizable moral emotions, such as anger at injustice.

Lack of empathy or sympathy is understandably the top candidate for the affective deficit that is responsible for psychopaths failing to understand moral right and wrong. Various suggestions have been made as to how lack of empathy leads to the full-scale moral deficits psychopaths seem to have. It has been suggested that psychopaths are born with a deficient violence inhibition mechanism, which prevents them from developing empathy, moral emotions, and moral understanding (Blair 1995). This idea takes its departure from the evidence that psychopaths have deficient physiological responses to others' pain and distress. Another proposal has it that moral judgment has two components: one which involves knowledge of norms and another which involves the capacity to have concern for the well-being of others (Nichols 2004). Psychopaths may have the former, but lack the latter. Without concern for others, psychopaths are unable to make *true* moral judgments. They know, in a discursive sense, what is right or wrong. However, they fail to appreciate the wrongness of wrong because they lack the ability to experience the requisite emotional reaction to such wrong. Or so the story goes, at any rate.

Lack of empathy is one of the diagnostic criteria for psychopathy, but what exactly does it come to? As defined in the *PCL-R*, lack of empathy may mean anything from lack of concern for the well-being or rights of others to deficient ability to imagine being in their position. It may even include the inability to relate to others emotionally as other agents or a failure to appreciate the reality of other agents as agents. On the background of this characterization, it is rather extraordinary that so many studies of empathy in psychopaths show no deficits. One of the most used tests of empathy is the Interpersonal Reactivity Index. Four studies show intact performance on the empathic concern component (Shamay-Tsoory et al. 2010; Domes et al. 2013; Lishner et al. 2012; von Borries et al. 2012), and another study only has secondary or high-anxious psychopaths underperforming, but not primary or low-anxious psychopaths (Mullins-Nelson et al. 2006). Two studies also show psychopaths experience normal personal distress (Shamay-Tsoory et al. 2010; von Borries et al. 2012). Furthermore, studies showing pictures of people in distress elicit ratings of unpleasantness similar to those of ordinary subjects (Herpertz et al. 2001; Birbaumer et al. 2005; Levenston et al. 2000). Considering how psychopaths act, this is rather puzzling.

Physiological measures and, to some degree, brain scans are more revealing and support the common assumption that psychopaths *do* have empathy deficits. Although the evidence is still somewhat mixed, more studies suggest that psychopaths have reduced skin conductance and attenuated fear-potentiated startle to others in distress compared to nonpsychopaths that do not (Herpertz et al. 2001; Birbaumer et al. 2005; Patrick et al. 1994; Verona et al. 2013). Skin conductance measures arousal. Increased skin conductance is associated with stress, fear, anxiety, and pain. It would appear that others' pain and distress cause a generalized stress response in observers. This is not true of psychopaths, however. Fear-potentiated startle indicates that the organism is on high alert, ready to initiate defensive action. It would therefore seem that psychopaths do not regard others' pain and suffering as a threat, but that normal people do.

The fMRI data is also suggestive, but more mixed. Psychopaths have reduced orbitofrontal cortex and ventromedial prefrontal cortex activation compared to nonpsychopaths in response to pain or distress in others (Decety et al. 2013a, b, 2014). However, the evidence is mixed when it comes to the activation of the areas most consistently associated with empathy: the anterior insula (AI), the anterior cingulate cortex (ACC), and the inferior frontal gyrus (IFG). When specifically asked to empathize with a person experiencing social rejection, for instance, psychopaths show intact activation in all these areas. However, if given no instructions, they activate these areas less than nonpsychopaths do (Meffert et al. 2013). This suggests that psychopaths *are* capable of empathizing with others, but they tend not to do so spontaneously. The idea that there might be intact empathic capacity in psychopaths after all is also supported by studies by Jean Decety and his group. If a psychopath is shown a picture of a person in a painful situation and asked to imagine that this is happening to himself, he has intact activation in AI, ACC, and IFG. If, on the other hand, he is asked to imagine it happening to someone else, he does not (Decety et al. 2013a). The same discrepancy in activation to

explicit instructions to feel with and imagine-self in pain and no instructions or imagine-other in pain is evident in the amygdala. The two first sets of instructions are associated with intact activation, the latter ones with deficient activation compared to controls (Meffert et al. 2013; Decety et al. 2013a). Again, it does seem as if the psychopath is capable of intact empathic responding to pictures of people in pain; he just has to imagine that it is himself who is in pain. This speaks less to a pervasive empathy deficit than to a more selective impairment in spontaneous empathizing with others.

Other researchers have alternative explanations of the morally relevant emotional deficits in psychopaths. Jesse Prinz, for instance, suggests that their behavioral inhibition system is impaired, and this expresses itself in the lack of emotions that are supposed to inhibit actions, such as fear and sadness (Prinz 2007). As we have seen, it is plausible that deficient fear is at the core of psychopaths' deficient affective response to others' pain and distress. It also explains the range of risky behaviors psychopaths regularly engage in. Prinz's account may link up less elegantly to moral concerns than theories that focus on deficient concern for the well-being of others. Fear is rarely regarded as a morally relevant emotion by philosophers, who instead tend to focus on such emotions as resentment or guilt (Greenspan 1995; Strawson 1962; Wallace 1996). Interestingly, at least one test of guilt in psychopaths found that primary or low-anxious psychopaths reported as much guilt as nonpsychopaths; only secondary or high-anxious psychopaths reported less guilt than controls (Mullins-Nelson et al. 2006).

Even if psychopaths do not altogether lack emotional reactions to the types of situations that typically give rise to moral affect in nonpsychopaths, they clearly have deficient emotional responses to them. It is therefore easy to see how such a deficit would impair their ability to make true moral judgments. When they see harm done to another, for instance, they are not gripped with the fear and horror that ordinary people are, and so the wrongness of the act may seem like a relatively superficial property of that act, like what clothes the agent was wearing or what knife was being used. Extend this general line of thinking to all other types of moral transgression – though probably not those committed toward the psychopath himself – and it should be clear what a sentimentalist picture of the psychopaths' moral deficit might look like.

Reason and Emotion

Some philosophers have taken a more ecumenical approach to moral judgment, arguing that some moral judgments are more based in reason, others more based in emotion (Greene et al. 2001). Much of the literature on so-called utilitarian versus deontological judgment reflects this idea. Utilitarianism and deontology constitute two philosophical theories about morality. For utilitarians the basic good is happiness, and the more happiness is created by an action, the better it is. Moral agents can either calculate the foreseeable happiness that their actions will produce before

they act (act utilitarianism) or they may adopt general rules, which typically have happiness-maximizing effects (rule utilitarianism). Typically, the happiness of the many will outweigh the happiness of the few. One need not only focus on maximizing happiness; one can also be concerned with minimizing suffering. If the latter, the best act would be the one that causes the least amount of suffering. By contrast, deontological moral theory rests on the idea that people have certain inviolable rights that no amount of optimizing happiness elsewhere can trump. There are certain things that are simply wrong, no matter how good the consequences. For instance, we may not kill an innocent person even if doing so would save many other innocent people. Killing (the innocent) is simply wrong and is therefore absolutely impermissible. The rightness or wrongness of an action rests in the kind of action that it is, not in its good or bad consequences. For instance, an action is wrong if it violates basic rights that a person has in virtue of being an agent or if it violates the categorical imperative (see above).

Most people have both utilitarian and deontological intuitions, though perhaps not at the same time. These intuitions are often brought out in moral dilemmas. Imagine that you are hiking next to a rail track. You reach an interchange. You look up the track and spot an out-of-control trolley bearing down the tracks. You look in the other direction and see hikers on both tracks. However, whereas one set of tracks only has one hiker on it, the other has six. The rails are set so that the trolley will go down the track with six hikers on it, almost certainly killing all of them. Should you pull the switch at the interchange, thereby causing the train to go down the track with just the one hiker on it, almost certainly killing her? Most people think it would be ok for you to do so. After all, you are saving six people even if you cause one to die. This thinking seems to be based on the utilitarian-sounding principle that it is better to sacrifice the few to save the many. But this principle does not hold in other situations. Imagine that you are a doctor at a hospital. You have in your care six patients in urgent need of organ transplants: liver, heart, kidneys, and lungs. You also have a routine meeting with a healthy patient. If you put that patient down, you can harvest his organs and thereby save six of your other patients. Should you kill your one patient to save six of your other patients? Most people say such an act would be impermissible. It seems just plain wrong to kill a healthy person even if doing so might save the lives of six others. Whatever the principle is here, it sounds more deontological. There is something wrong in the act of killing the patient itself, no matter what its positive consequences. However, the two situations seem similar: they both involve sacrificing the one to save the many.

Joshua Greene and colleagues (2001) scanned the brains of people while they were making decisions about moral dilemmas of the sort discussed above. They found that when people make deontological decisions – typically refusing to sacrifice the one to save the many – they engage more emotional areas of the brain than when they make utilitarian ones – roughly sacrificing the one to save the many. Greene concludes that deontological reasoning is based in affect. This is surprising since deontological reasoning is associated with Kant, who was a rationalist. The problem with affect, though, is that it often skews our opinions.

Consequently, we should be very skeptical about our deontological intuitions. There is considerable debate both about the appropriateness of labeling the decisions utilitarian or deontological and about whether deontological reasoning really is as affect laden as Greene seems to think. However that may be, some studies suggest that psychopaths make more utilitarian judgments than do controls. This has caused some people to argue that utilitarianism is associated with not caring much for people (Bartels and Pizarro 2011). This is certainly a bit of an odd conclusion, since utilitarian reasoning is based on a concern for others' well-being.

As before, the evidence is not as clean as one would hope. Where one study shows impaired performance (Bartels and Pizarro 2011), another shows that psychopaths make as many utilitarian-type judgments as do nonpsychopaths (Glen et al. 2009b). It is sometimes thought that the difference comes out most clearly in so-called personal dilemmas, where sacrificing the one involves physical contact with the victim (pushing, smothering, and so on). People with damage to their prefrontal cortex make more utilitarian-type judgments on such dilemmas (Koenigs et al. 2007), and such patients are often compared with psychopaths because of their partially overlapping symptomatology. However, at least one study shows no tendency of psychopaths to make more utilitarian-type judgments on this restricted range of moral dilemmas (Cima et al. 2010). Another study shows that only low-anxious psychopaths make more utilitarian judgments on personal moral dilemmas compared to controls, whereas high-anxious psychopaths perform at norm (Koenigs et al. 2012). This latter study does support the idea that affect backs deontological-type reasoning, but the results from other studies are so mixed that it would be premature to conclude anything very definite about the affective-moral capacities of psychopaths. There is also the added difficulty that even if low-anxious psychopaths make more utilitarian judgments on personal moral dilemmas, this hardly shows that they have a moral *deficit*. For perhaps they are making the *right* decision, they just happen not to be swayed by morally irrelevant affect.

Are Psychopaths Responsible?

The question of whether psychopaths are responsible for their actions is typically addressed either in moral or legal terms. Although legal responsibility typically tracks moral responsibility, the two can come apart, as in cases of strict liability. So, someone may not be morally responsible, yet be legally responsible. Someone may be morally responsible for something and not legally responsible because there are no laws prohibiting the kind of behavior. Concerns about psychopaths' standing as responsible agents derive from two sources: their emotional and rational deficits. Whether or not theorists think psychopaths are responsible for their actions comes down to how pervasive they think their deficits are in the areas that most affect understanding right and wrong.

Legal Responsibility

Historically, psychopaths have not been exempted from legal responsibility. This is largely due to the fact that they are aware of what they are doing when they commit a wrong, they know it is wrong, and they are able to control their actions. Furthermore, they do not suffer from delusions or hallucinations of the sort that usually exculpate other mentally ill defendants. They appear neither cognitively nor volitionally impaired in ways relevant to criminal responsibility. A number of philosophers (Maibom 2008), psychopathy researchers (Hare 1993; Cleckley 1976), and legal theorists (Reznek 1997) argue that they are therefore legally responsible, whereas others maintain that psychopaths do not have sufficient understanding of right and wrong to be held responsible (Duff 1977) or at least not fully responsible (Glannon 1997).

In order to be held legally responsible for a wrongdoing, a person must have what is called "a guilty mind," or *mens rea*. She must understand, or be capable of understanding, what she is doing and she must know, or be capable of knowing, that what she is doing is wrong. There is little argument that psychopaths know what they are doing. The question is whether they know that what they are doing is wrong. The law distinguishes between two types of wrong: *malum prohibitum* and *malum in se*. *Malum prohibitum* is a legally enforced conventional wrong, such as double-parking or nude bathing. *Malum in se*, on the other hand, refers to something that is wrong in itself – i.e., something that has a deeper justification outside its being legally culpable – such as murder, rape, or theft. It seems, therefore, in order to have *mens rea* of murder, say, one must be capable of a deeper understanding of right and wrong.

As we have seen, there are some reasons to think that psychopaths may lack a deeper understanding of right and wrong, either because they have deficient rational capacities or deficient or absent affectivity. Accordingly, there are philosophers who maintain that psychopaths are no more responsible for their wrongdoings than other people who suffer from mental disorder (Duff 1977; Wallace 1996; Glannon 1997). We noted before that psychopaths have both deficient practical reason or deficient affective responses to the pain and distress of others. It should be stressed, however, that such capacities are impaired, not absent.

Those who argue that psychopaths have sufficient moral knowledge to be legally responsible typically focus on the fact that their deficits appear to be just that: deficits not inabilities. They have enough understanding to appreciate the wrongness of their actions – namely, that they are causing pain and disability to others, that they are acting against their wishes, and so on – so that they can be held legally responsible for them. This is the view of two more prominent psychopathy researchers, Hervey Cleckley and Robert Hare. Others argue that there is a fundamental difference between the moral disability experienced by people judged to be insane and psychopaths. Whereas the former have lacunae in an overall intact moral capacity, what seems to characterize psychopaths is a distinct *lack* of moral concern. We excuse the insane for committing wrongs often because had their mistaken beliefs been right, their action would have been justified (Reznek 1997). By contrast, psychopaths' "mistake" is to think of others as having value only to the

extent that they can help further their own goals, and to believe that the suffering of others comes second to their own interests. In a way, their mistake is to think that nothing that they do is impermissible. That is tantamount to being immoral or bad. And thus, it is only if you think that one cannot be bad without being mad that you will be convinced that psychopaths are not legally responsible because they are insane (Maibom 2008).

Another way to argue that psychopaths are not legally responsible would be to focus on their volition. Are they really capable of controlling their actions in the way the law requires? This line of argument is typically not explored because of the difficulty of showing that a defendant could not have done otherwise.

Moral Responsibility

Many consider legal responsibility to depend on moral responsibility, so that only if someone is morally responsible, can she also be legally responsible. Theories of moral responsibility are rather diverse. They typically all require either a capacity for practical rationality and/or for affective resonance or reactivity to others and their plight. As such, most of what was discussed above can be applied here. Typically, the debate is put in terms of whether psychopaths are responsive to moral *reasons*. This raises the difficult issues of what reasons for actions are and what counts as being responsive to them. A reason for you not to run to work is that you are going to show up all sweaty and put your colleagues off. For that to count as a reason for you, you must also want to not put off your colleagues. Moral reasons are somewhat different from such simple practical reasons in that they are supposed to give you reasons to act or refrain from acting no matter what your desires or plans are. Let's say the only way for you to get the job you want is to arrange the early demise of another candidate who has just been offered the job. The fact that you would have to kill another to get the job is reason enough for you not to do it period. In less extreme cases, the fact that you will hurt someone's feelings at least counts against the action that you are contemplating even if you end up performing it anyway. Psychopaths act in rather disorganized ways – even so-called successful psychopaths are typically found in community settings such as short-term work centers – that raise questions even about their ability to act in their own best interests. It is therefore understandable how they might be thought to not really grasp the nature of reasons (Kennett 2002).

As always the debate is about whether the deficits psychopaths have in the area of practical reasoning are sufficient to make it the case that they do not understand the nature of reasons. Whereas as a number of philosophers argue that they fail to understand the nature of reasons (Kennett 2002; Duff 1977), others argue that it is questionable that psychopaths are *that* deficient in their reasoning (cf. Maibom 2005). What is often definitive in this debate is how thick a reading one gives of "reasons." Those who think that psychopaths do not understand reasons tend to have rather thick notions of what such understanding consists in. A thinner notion characterizes the thoughts of those who think psychopaths have the ability to comprehend the nature of reasons. The worry about the thicker notion, of course,

is that many people other than psychopaths may also fail to understand the nature of reasons, rendering a potentially large number of people incapable of being morally responsible for their actions.

Not all ways of fleshing out the ability to be responsive to reasons are in terms of cognitive or rational capacities. Some argue that to understand moral reasons, we must be capable of understanding that other people's preferences are reason-giving *for us*. But such an understanding flows from our ability to empathize with them, some say (Shoemaker 2015). In a similar vein, it has been argued that without empathy we cannot appreciate the value of other people's goals and projects and therefore they cannot move us (Deigh 1995). Others maintain that the capacity to feel concern for others is foundational to moral judgments (Nichols 2004), yet others that sadness and (moral) anger are (Prinz 2007). And since psychopaths are deficient in these areas, these thinkers conclude that they are not morally responsible for their actions. Lastly, one could argue that psychopaths are undermotivated to do what is right and refrain from doing what is wrong. Again, one can reference either their practical reasoning deficits or their deficient emotions. This possibility is relatively underexplored.

Conclusion

It is evident that psychopaths have a very different attitude to moral and legal demands on their actions than ordinary people do. They frequently engage in immoral or illegal activities, and they demonstrate a curious lack of empathy for their victims and guilt or remorse for their actions. They have deficits in a number of areas that theorists have identified as central to moral capacities, such as practical reason or emotion. Such deficits may account for their deficient or lacking moral sense. Whether they also render them exempt from moral and legal responsibility depends on what is required for such responsibility and how pervasive you understand these deficits to be.

Definitions of Key Terms

Practical reason	The capacity for decision-making.
Moral understanding	Understanding that, and possibly why, certain acts are right/wrong, good/bad.
Moral judgment	The determination of the moral quality of an agent or an action; can be a thought or a verbal act.
Utilitarianism	The theory that the ultimate moral good is happiness and the ultimate moral evil is suffering. Thus, we ought to strive to increase happiness and reduce suffering. It is the total amount of happiness and suffering that counts. The moral value of our actions lie in the *consequences* they produce.

Deontology	The idea that certain actions are permissible or impermissible in virtue of the sorts of actions that they are, not due to their good or bad consequences.
Welfare justification	A justification of why some act is morally wrong that refers to the harm, suffering, or reduced welfare of the subject that is the patient of that action.
Purity	Concern with purity is one of Jonathan Haidt's moral domains. Purity concerns are very prominent in many religious cultures, where certain animals (pigs), certain people (women menstruating, untouchables), or certain actions (masturbating) are impure. The justification of why purity violations are wrong is not that there is harm involved. One may substitute "pure" for "natural" to get a sense of how the category is most commonly applied in the West.
Internalism about moral judgment	In its conceptual variant, it holds that it is part of our concept of what it is to judge that something is wrong (or: right) that you are also motivated not to (or: to) perform it. In its empirical form, it claims that as psychological matter of fact, if you judge an act to be wrong, you are thereby also motivated not to perform it.
Moral motivation	Motivation to act in accordance with what is thought to be right/good or to avoid doing what is thought to be wrong/bad.
Externalism about moral judgment	The opposite of internalism. Either it is not part of the concept of a moral judgment that if you make it, you are thereby motivated to act in accordance with it, or it is not an empirical fact about human psychology that if you make a moral judgment, you are thereby motivated to act in accordance with it.

Summary Points

- Psychopaths can distinguish between right/good and wrong/bad actions.
- Psychopaths are capable of giving moral justifications.

- Psychopaths do not react to pain and suffering in others as ordinary people do.
- Psychopaths have impaired ability to make good decisions.
- Psychopaths may lack a deeper understanding of right/wrong, good/bad due to their practical reasoning deficits and/or their impaired emotionality.
- It is debatable whether psychopaths' deficits are sufficient to make them not responsible for their actions, legally or morally.

References

Aharoni E, Atanenko O, Kiehl K (2011) Disparities in the moral intuitions of offenders: the role of psychopathy. J Res Pers 45:322–327

Aharoni E, Sinnott-Armstrong W, Kiehl K (2012) Can psychopathic offender discern moral wrong? A new look at the moral/conventional distinction. J Abnorm Psychol 121:484–497

Bartels D, Pizarro D (2011) The mismeasure of morals: antisocial personality traits predict utilitarian responses to moral dilemmas. Cognition 121:154–161

Birbaumer N, Veit R, Lotze M et al (2005) Deficient fear conditioning in psychopathy. Arch Gen Psychiatry 62:799–805

Blair RJR (1995) A cognitive developmental approach to morality: investigating the psychopath. Cognition 57:1–29

Blair RJR (1997) Moral reasoning and the child with psychopathic tendencies. Personal Individ Differ 22:731–739

Blair RJR, Colledge E, Mitchell D (2001) Somatic markers and response reversal: is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? J Abnorm Child Psychol 29:499–511

Blair RJR, Mitchell D, Blair K (2005) The psychopath: emotion and the brain. Blackwell, Oxford

Brink D (1989) Moral realism and the foundations of ethics. Cambridge University Press, Cambridge

Cima M, Tonnaer F, Hauser MD (2010) Psychopaths know right from wrong but don't care. Soc Cogn Affect Neurosci 5:59–67

Cleckley H (1976) The mask of sanity, 5th edn. Mosby, St. Louis

Decety J, Chen C, Harenski C et al (2013a) An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. Front Hum Neurosci 7:489. doi:10.3389/ fnhum.2013.00489

Decety J, Skelly L, Kiehl K (2013b) Brain responses to empathy-eliciting scenarios involving pain in incarcerated individuals with psychopathy. JAMA Psychiatry 70:638–645

Decety J, Skelly L, Yoder K et al (2014) Neural processing of dynamic emotional facial expressions in psychopaths. Soc Neurosci 9:36–49

Deigh J (1995) Empathy and universalizability. Ethics 105:743–763

Domes G, Hollerbach P, Vohs K, Mokros A, Habermeyer E (2013) Emotional empathy and psychopathy in offenders: an experimental study. J Personal Disord 27:67–84

Duff RA (1977) Psychopathy and moral understanding. Am Philos Q 14:189–200

Glannon W (1997) Psychopathy and responsibility. J Appl Philos 14:263–275

Glen A, Iyer R, Graham J et al (2009a) Are all type of morality compromised in psychopathy? J Personal Disord 23:384–398

Glen A, Raine A, Schug A, Young L, Hauser M (2009b) Increased DLPFC activity during moral decision making in psychopathy. Mol Psychiatry 14:909–911

Greene JD, Sommerville RB, Nystrom LE et al (2001) An fMRI investigation of emotional involvement in moral judgment. Science 293:2105–2108

Greenspan PS (1995) Practical guilt: moral dilemmas, emotions, and social norms. Oxford University Press, New York

Harenski CL, Harenski KL, Shane MS, Kiehl KA (2010) J Abnorm Psychol 119:864–874
Hare R (1993) Without conscience: the disturbing world of psychopaths among us. Pocket Books, New York
Hare R (2004) The Hare psychopathy checklist-revised. Multi-Health System, Toronto
Herpertz SC, Werth U, Lukas G et al (2001) Emotion in criminal offenders with psychopathy and borderline personality disorder. Arch Gen Psychiatry 58:737–745
Hiatt KD, Newman JP (2006) Understanding psychopathy: the cognitive side. In: Patrick CJ (ed) Handbook of psychopathy. Guilford Press, New York, pp 334–352
Hicks B, Markon K, Patrick C et al (2004) Identifying psychopathy subtypes on the basis of personality structure. Psychol Assess 16:276–288
Kant I (1785/1993) Grounding for the metaphysics of morals. (trans, Ellington J). Hackett Publishing, Indianapolis
Kennett J (2002) Autism, empathy and moral agency. Philos Q 52:340–357
Kiehl KA, Lushing J (2014) Psychopathy. Scholarpedia 9:30835
Koenigs M, Young L, Adolphs R et al (2007) Damage to the prefrontal cortex increases utilitarian moral judgments. Nature 446:908–911
Koenigs M, Kruepke M, Zeier J et al (2012) Utilitarian moral judgment in psychopathy. Soc Cogn Affect Neurosci 7:708–714
Levenston G, Patrick C, Bradley M et al (2000) The psychopath as observer: emotion and attention in picture processing. J Abnorm Psychol 109:373–385
Link NF, Scherer SE, Byrne PN (1977) Moral judgment and moral conduct in the psychopath. Can Psychiatr Assoc J 22:341–346
Lishner DA, Vitacco MJ, Hong PY, Mosley J, Miska K, Stocks EL (2012) Evaluating the relation between psychopathy and affective empathy: two preliminary studies. Int J Offender Therapy Comp Criminol 56:1161–1181
Lykken DT (1957) A study of anxiety in the psychopathic personality. J Abnorm Soc Psychol 55:6–10
Maibom HL (2005) Moral unreason: the case of psychopathy. Mind Lang 20:237–257
Maibom HL (2008) The mad, the bad, and the psychopath. Neuroethics 1:167–184
Marsh A, Cardinale E (2012) Psychopathy and fear: specific impairments in judging behaviors that frighten others. Emotion 12:892–898
Meffert H, Gazzola V, den Boer JA et al (2013) Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. Brain 136:2550–2562
Mullins-Nelson JL, Salekin RT, Leistico A-MR (2006) Psychopathy, empathy, and perspective taking ability in a community sample: implications for the successful psychopath concept. Int J Forensic Ment Health 5:133–149
Newman J, Kosson D (1986) Passive avoidance learning in psychopathic and non-psychopathic offenders. J Abnorm Psychol 96:257–263
Nichols S (2004) Sentimental rules: on the nature and foundation of moral judgment. Oxford University Press, New York
O'Kane A, Fawcett D, Blackburn R (1996) Psychopathy and moral reasoning: comparison of two classifications. Personal Individ Differ 20:505–514
Patrick C, Cuthbert B, Lang P (1994) Emotion in the criminal psychopath: fear image processing. J Abnorm Psychol 103:523–534
Porter S (1996) Without conscience or without active conscience: the etiology of psychopath revisited. Aggress Violent Behav 1:179–189
Prinz JJ (2007) The emotional construction of morals. Oxford University Press, New York
Reznek L (1997) Evil or ill: justifying the insanity defence. Routledge, London
Salekin RT (2006) Psychopathy in children and adolescents: key issues in conceptualization and assessment. In: Patrick C (ed) Handbook of psychopathy. Guilford Press, New York, pp 389–415
Shamay-Tsoory S, Harari H, Aharon-Perez J et al (2010) The role of orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. Cortex 46:668–677

Shoemaker D (2015) Responsibility from the margins. Oxford University Press, Oxford
Strawson PF (1962) Freedom and resentment. Proc Br Acad 48:1–25
Verona E, Bresin K, Patrick C (2013) Revisiting psychopathy in women: Cleckley/Hare conceptions and affective response. J Abnorm Psychol 122:1088–1093
von Borries AKL, Volman I, de Bruijn ERA et al (2012) Psychopaths lack the autonomic avoidance of social threat: relation to instrumental aggression. Psychiatry Res 200:761–766
Wallace RJ (1996) Responsibility and the moral sentiments. Harvard University Press, Cambridge, MA
Young L, Koenigs M, Kruepke M et al (2012) Psychopathy increases perceived moral permissibility of accidents. J Abnorm Psychol 121:659–667

Index

T. Schramme, S. Edwards (eds.), *Handbook of the Philosophy of Medicine*,
DOI 10.1007/978-94-017-8688-1

I