What is Scientific Knowledge?

An Introduction to Contemporary Epistemology of Science

EDITED BY
KEVIN MCCAIN AND
KOSTAS KAMPOURAKIS

WHAT IS SCIENTIFIC KNOWLEDGE?

What Is Scientific Knowledge? is a much-needed collection of introductory-level chapters on the epistemology of science. Renowned historians, philosophers, science educators, and cognitive scientists have authored 19 original contributions specifically for this volume. The chapters, accessible for students in both philosophy and the sciences, serve as helpful introductions to the primary debates surrounding scientific knowledge. First-year undergraduates can readily understand the variety of discussions in the volume, and yet advanced students and scholars will encounter chapters rich enough to engage their many interests. The variety and coverage in this volume make it the perfect choice for the primary text in courses on scientific knowledge. It can also be used as a supplemental book in courses in epistemology, philosophy of science, and other related areas.

Key features:

- an accessible and comprehensive introduction to the epistemology of science for a wide variety of students (both undergraduate- and graduate-level) and researchers
- written by an international team of senior researchers and the most promising junior scholars
- addresses several questions that students and lay people interested in science may already have, including questions about how scientific knowledge is gained, its nature, and the challenges it faces.

Kevin McCain is Associate Professor of Philosophy at the University of Alabama at Birmingham, USA. His published research includes *Evidentialism and Epistemic Justification* (2014), *The Nature of Scientific Knowledge: An Explanatory Approach* (2016), and, with Kostas Kampourakis, *Uncertainty: How It Makes Science Advance* (2019).

Kostas Kampourakis is a researcher in science education and a lecturer at the University of Geneva, Switzerland. His most recent authored books are *Making Sense of Genes* (2017), *Turning Points: How Critical Events Have Driven Human Evolution, Life and Development* (2018), and, with Kevin McCain, *Uncertainty: How It Makes Science Advance* (2019). He has also co-edited, with Michael Reiss, *Teaching Biology in Schools: Global Research, Issues and Trends* (2018).

WHAT IS SCIENTIFIC KNOWLEDGE?

An Introduction to Contemporary Epistemology of Science

Edited by Kevin McCain and Kostas Kampourakis

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2020
by Routledge
52 Vanderbilt Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

Library of Congress Cataloging-in-Publication Data
A catalog record for this title has been requested

ISBN: 978-1-138-57016-0 (hbk)
ISBN: 978-1-138-57015-3 (pbk)
ISBN: 978-0-203-70380-9 (ebk)

Typeset in Bembo
by Newgen Publishing UK

CONTENTS

CONTRIBUTORS

Hanne Andersen is Professor of Philosophy of Science at the University of Copenhagen, Denmark. Her publications include *On Kuhn* (2001), *The Cognitive Structure of Scientific Revolutions* (with Peter Barker and Xiang Chen, 2006), and the entry on scientific method in the *Stanford Encyclopedia of Philosophy* (with Brian Hepburn).

Theodore Arabatzis is Professor of History and Philosophy of Science at the National and Kapodistrian University of Athens. He has published on the history of modern physical sciences and on historical philosophy of science. His publications include *Representing Electrons: A Biographical Approach to Theoretical Entities* (2006).

Christoph Baumberger is Senior Researcher at the Department of Environmental Systems Science, Swiss Federal Institute of Technology Zurich, Switzerland. He published (with Stephen R. Grimm and Sabine Ammon) *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science* (2017).

Matthew J. Brown is Associate Professor of Philosophy and the History of Ideas and Director of the Center for Values in Medicine, Science, and Technology at the University of Texas at Dallas. His published work in the history and philosophy of science has focused primarily on values in science, feminism and science, and science for policy. He has a book in progress, *Science and Moral Imagination: A New Ideal for Values in Science*.

Martin Curd is the co-author and co-editor (with Jan Cover and Christopher Pincock) of *Philosophy of Science: The Central Issues* (2013) and co-editor (with Stathis Psillos) of *The Routledge Companion to Philosophy of Science* (2014). He retired from Purdue University in 2018.

Olaf Dammann is Professor and Vice Chair of Public Health and Community Medicine at Tufts University School of Medicine, Boston. He is also a PhD candidate in Philosophy at the University of Johannesburg, South Africa. He has more than 200 publications in neuropediatrics, neonatology, epidemiology, philosophy of science and recently published (with Ben Smart) the book *Causation in Population Health Informatics and Data Science* (2019).

Henk W. de Regt is Professor of Philosophy of Natural Sciences at the Institute for Science in Society, Radboud University Nijmegen. He published *Understanding Scientific Understanding* (2017) and (with Sabina Leonelli and Kai Eigner) *Scientific Understanding: Philosophical Perspectives* (2009).

Jeroen de Ridder is Associate Professor of Philosophy at Vrije Universiteit Amsterdam and Professor by special appointment of Christian Philosophy at the University of Groningen. His work in social epistemology has appeared in journals such as *The Australasian Journal of Philosophy, Synthese, Erkenntnis, and Episteme*. He recently edited, together with Rik Peels and René van Woudenberg, *Scientism: Prospects and Problems* (2018).

Richard Fumerton is the F. Wendell Miller Professor of Philosophy at the University of Iowa. His research has focused mainly on epistemology, but he has also published books and articles in metaphysics, philosophy of mind, philosophy of science, value theory, and philosophy of law. He is the author of *Metaphysical and Epistemological Problems of Perception* (1985), *Reason and Morality: A Defense of the Egocentric Perspective* (1990), *Metaepistemology and Skepticism* (1996), *Realism and the Correspondence Theory of Truth* (2002), *Epistemology* (2006), *Mill* (co-authored with Wendy Donner) (2009), and *Knowledge, Thought and the Case for Dualism* (2013).

Susan A. Gelman is the Heinz Werner Distinguished University Professor of Psychology and Linguistics at the University of Michigan. She has published over 200 books and papers on concepts and language in young children, including co-editing (with Mahzarin Banaji) *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (2013) and authoring, *The Essential Child: Origins of Essentialism in Everyday Thought* (2003). Her honors include the Developmental

Psychology Mentor Award of the American Psychological Association and the G. Stanley Hall Award for Distinguished Contribution to Developmental Psychology (APA, Division 7).

Sara Gottlieb is a postdoctoral fellow at Yale University in the Cognition and Development Lab. She earned her PhD in cognitive psychology from the University of California Berkeley in 2018. Her main research area is folk epistemology, and she is particularly interested in how people think about science and scientific explanations.

Carrie S. I. Jenkins is Canada Research Chair and Professor of Philosophy at the University of British Columbia. Her publications include a monograph on the philosophy of arithmetic (*Grounding Concepts*, 2008) and a recent book on the nature of romantic love (*What Love Is and What It Could Be*, 2017).

Kostas Kampourakis is a researcher in science education and a lecturer at the University of Geneva, Switzerland. His most recent authored books are *Making Sense of Genes* (2017), *Turning Points: How Critical Events Have Driven Human Evolution, Life and Development* (2018), and, with Kevin McCain, *Uncertainty: How It Makes Science Advance* (2019), having also co-edited, with Michael Reiss, *Teaching Biology in Schools: Global Research, Issues and Trends* (2018).

Janet A. Kourany is Associate Professor of Philosophy and concurrent Associate Professor of Gender Studies at the University of Notre Dame, where she is also a Fellow of the Reilly Center for Science, Technology, and Values. Her books related to science include *Scientific Knowledge* (1987, 1998); *The Gender of Science* (2002); *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited* (with Martin Carrier and Don Howard; 2008); *Philosophy of Science after Feminism* (2010); and *Science and the Production of Ignorance: When the Quest for Knowledge Is Thwarted* (with Martin Carrier; 2019). She is currently working on a new book entitled *Bacon's Promise*.

Stephen Law is a freelance philosopher and author, and editor of the Royal Institute of Philosophy journal *THINK*. His books include *Humanism: A Very Short Introduction* (2011) and *The Philosophy Gym: 25 Short Adventues in Thinking* (2004).

Tania Lombrozo is a Professor of Psychology at Princeton University. Her research explores foundational topics in human cognition using the empirical tools of experimental psychology and the conceptual tools of analytic philosophy. She is the recipient of numerous early career awards from organizations including the National Science Foundation, the Association for Psychological Science, the American Psychological Association, and the Society for Philosophy and Psychology.

Kristan A. Marchak is an assistant professor at the University of Alberta Faculté Saint-Jean. She previously completed a postdoctoral fellowship at the University of Michigan's Weinberg Institute for Cognitive Science. Dr. Marchak's research focuses on children's and adults' concepts of individuals and kinds.

Kevin McCain is Associate Professor of Philosophy at the University of Alabama at Birmingham. His published research includes *Evidentialism and Epistemic Justification* (2014), *The Nature of Scientific Knowledge: An Explanatory Approach* (2016), and (with Kostas Kampourakis) *Uncertainty: How It Makes Science Advance* (2019).

Rik Peels is Assistant Professor of Philosophy at the Vrije Universiteit Amsterdam (the Netherlands). His most recent work includes *Responsible Belief: A Theory in Ethics and Epistemology* (2017) and the volume, edited with Jeroen de Ridder and René van Woudenberg, *Scientism: Prospects and Problems* (2018).

Erik L. Peterson is Associate Professor of the History of Science at the University of Alabama. His book, *The Life Organic: The Theoretical Biology Club and the Roots of Epigenetics,* was published in 2016. Two books are forthcoming: *Darwin vs. Chesterton: Eugenics, Race, Immigration, and the Battle of the Soul of British Science* and *One or Many?: A History of Race & Science from Columbus to Genomics* (with biological anthropologist Jim Bindon).

Ted Poston is the inaugural Director of the McCollough Institute for Pre-Medical Education and Professor of Philosophy at the University of Alabama. He has published two books *Reason & Explanation: A Defense of Explanatory Coherentism* (2014) and (with Adam C. Carter) *A Critical Introduction to Knowledge How* (2018). He has published over thirty articles on a variety of topics in epistemology, philosophy of science, and philosophy of religion.

Thomas A.C. Reydon is Professor of Philosophy of Biology in the Institute of Philosophy with cross-appointment in the Centre for Ethics and Law in the Life Sciences (CELLS), Leibniz University Hannover, Germany. He is Editor in Chief of the *Journal for General Philosophy of Science*, Editor in Chief of the book series *History, Philosophy and Theory of the Life Sciences*, and co-founder and Board Member of the German Society for the Philosophy of Science (GWP). For details about his academic work, see his website at www.reydon.info.

Andrew Shtulman is Professor of Psychology and Cognitive Science at Occidental College. He studies conceptual development and conceptual change and is the author of *Scienceblind: Why Our Intuitive Theories About the World are So Often Wrong* (2017).

Paul Thagard is Distinguished Professor Emeritus of Philosophy at the University of Waterloo. He is author of many books, including three in 2019: *Brain-Mind: From Neurons to Consciousness and Creativity; Mind-Society: From Brains to Social Sciences and Professions*; and *Natural Philosophy: From Social Brains to Knowledge, Reality, Morality, and Beauty*.

Dana Tulodziecki is Associate Professor of Philosophy at Purdue University. She received her degrees from the LSE and Columbia University and works mainly on general issues in Philosophy of Science and History and Philosophy of Science.

Andrew Young is Assistant Professor of Psychology at Northeastern Illinois University. He conducts research on learning and scientific reasoning in childhood.

PREFACE

Undeniably, scientific knowledge is extremely important to many aspects of our lives from our wellness to our technology and everyday conveniences. Given its importance, it is not surprising that scholars from many disciplines study scientific knowledge and how it is generated.

Here we have assembled a stellar group of philosophers, historians, and cognitive scientists to discuss aspects of scientific knowledge. The results of bringing this collection of scholars together are the 19 newly commissioned chapters in this volume which approach issues related to scientific knowledge from a variety of disciplinary perspectives. While each chapter is introductory, the aim is not to be perfectly neutral on the issues in question. Consequently, each chapter should be taken to be an opinionated introduction to its topic. In every case, the chapters cite the relevant literature so that readers can examine for themselves other viewpoints.

This volume is designed to introduce non-experts to key debates concerning scientific knowledge. In light of this, it could serve as a standalone text for a course on scientific knowledge. Alternatively, the chapters contained herein could be readily combined with other texts for courses in philosophy of science, epistemology, or other areas. In each instance the chapters in this volume do not presuppose familiarity with philosophy or science. Rather, they introduce readers to key issues surrounding scientific knowledge and its production.

The chapters in this volume are grouped around four major questions about scientific knowledge. All of the chapters in Part I explore questions related to how scientific knowledge is generated. Those in Part II take on issues concerning features of scientific knowledge. Part III consists of chapters examining the role that bias plays in science. Finally, the chapters in Part IV consider ways in which our scientific knowledge may be limited. Taken together, the chapters in the four parts of this volume cover a wide swath of issues related to scientific knowledge.

Consequently, someone who grapples with the debates described in these chapters will have a solid understanding of the key issues in the epistemology of science.

The present book thus offers an accessible and comprehensive introduction to the epistemology of science for a wide range of undergraduate students, researchers, and general audiences. Containing contributions from experts in epidemiology, epistemology, history of science, philosophy of science, cognitive science, and science education, it is rich in diversity. It covers many different aspects of scientific knowledge from a variety of perspectives and addresses several questions that students and lay people interested in science may have, including questions about how scientific knowledge is gained, its nature, and the challenges it faces.

Whereas the book is primarily intended for students in philosophy, and the arts and sciences more broadly, as well as researchers working on philosophy of science, epistemology, science education, and science studies, we believe that it is also appropriate for lay people interested in epistemological issues pertaining to science. The authors have tried to minimize philosophical jargon so that all chapters are accessible to people with limited background in philosophy. Most importantly, the topics they have written about are crucial ones that every well-informed citizen should know about. Last but not least, we do hope that scientists will find this book useful and thought provoking in that it will help them to think harder about issues they do not always have the opportunity to reflect upon.

We would like to thank Emma Starr for editorial assistance throughout the production process. We also would like to especially thank Andy Beck for his help in guiding this volume to its fruition. He was enthusiastic about this project from the beginning and great to work with from start to finish. Last, but certainly not least, we are indebted to the contributors for writing informative and accessible chapters for this volume.

Kevin McCain and Kostas Kampourakis

PART I

How Is Scientific Knowledge Generated?

1

HOW MANY SCIENTISTS DOES IT TAKE TO HAVE KNOWLEDGE?

Jeroen de Ridder

Introduction

Contemporary scientific research, especially in most of the STEM disciplines and the social sciences, is massively collaborative. This is something you can easily observe for yourself when visiting your local university's science labs. They will typically sport teams of researchers, technicians, and students involved in collaborative projects. These teams can be part of larger collaborations with people in other departments or at other universities. The collaborative nature of science also shows clearly in the official records of science: publications in peer-reviewed journals. A 2016 article in *The Economist* reports that the average number of authors on scientific papers rose from 3.2 to 4.4 between 1996 and 2015.[1] An earlier report from Thomson Reuters' *ScienceWatch* newsletter notes that fewer than one in five papers in the sciences, social sciences, and humanities had a single author (the percentage goes down to 12% for the natural sciences alone), and that 2011 saw over 600 papers with more than 100 authors and 146 with more than 1,000 authors.[2] As I'm writing this, a 2015 paper in particle physics, produced by scientists from all over the world collaborating with the teams operating the Large Hadron Collider at CERN in Switzerland, holds the record for highest author count on a single paper, with a total of 5,154 authors.[3] The list of authors takes up 24 of the paper's 33 pages! In the same year, biologists broke the 1,000 authors barrier for the first time with a genomics paper on fruit-flies authored by 1,014 people, more than 900 of whom were undergrads.[4]

These numbers demonstrate that science has become increasingly collaborative over the past decades, but it should be noted at the outset that they also reflect one of the less edifying aspects of contemporary science. The amount and (perceived) quality of publications can make or break academic careers. Scientists have a strong

incentive to rake in as many as they can. As a result, researchers frequently dish out authorships by very liberal criteria, engage in tit-for-tat exchanges of authorships on each other's papers, and add honorary authors to increase the chances of getting published in prestigious venues.[5] This leads to inflated author lists, which don't always reflect genuine collaboration and substantive research contributions. In response, many scientific journals have started to implement stricter guidelines for authorships.[6] These developments raise important questions, but for the purposes of this chapter, I will leave them aside.[7]

Instead, I want to explore some *epistemological* consequences of the fact that so much contemporary science is collaborative. That is, I want to investigate what collaboration means for the knowledge that is produced through it. What I will end up arguing is that there are a number of senses in which much contemporary scientific knowledge is *collective knowledge*. Often, *groups*, and not just individuals, have knowledge.

I start by surveying a number of salient features of scientific knowledge in collaborative settings. I then show how these features support the claim that scientific knowledge is collective knowledge in three different senses. I conclude with a short summary and some closing reflections.

Scientific Knowledge in Collaborative Contexts

Before we look at features of scientific knowledge specific to a collaborative context, we ought to get a better grip on scientific knowledge as such. Throughout this chapter, I rely on a characterization of knowledge that is widely endorsed in epistemology. According to it, knowledge is *warranted true belief*, where warrant is a general epistemically good-making property that makes the difference between a belief's being merely true and its constituting knowledge (cf. Plantinga 1993 and Burge 2003 for two different accounts of warrant). If a belief is warranted, that might mean, among other things, that it's not luckily true when it is true,[8] that it is reliably produced, or that it is based on good grounds.

Why is scientific knowledge held in such high regard? What sets it apart from non-scientific knowledge? For one thing, it is knowledge produced through scientific research. That is certainly true, but it tells us little about what is supposed to be good about scientific knowledge. A better starting point is the thought that scientific knowledge is *high-grade knowledge*, that is, knowledge that satisfies demanding epistemic standards and that, as a result, is highly reliable, robust, or well-established. There are different ways of understanding this general idea (De Ridder 2018). One tempting but mistaken reading is to think that scientific knowledge is the most certain knowledge that we have. That is false, however. Quantum mechanics or the standard model of particle physics are among the most firmly established results of science, but my knowledge that there is coffee in the cup in front of me or that $2 + 2 = 4$ is even more certain. That is because, unlike scientific knowledge, such humdrum beliefs are not based on complex inferences from

observations mediated by instruments and technology. Hence, they do not suffer from the inevitable uncertainties that attach to such inferences.

A better way to think about this issue is that scientific knowledge is *the most reliable knowledge we have about certain subject matters* – to wit, the underlying nature of reality and human beings. Science is our most reliable means for discovering non-obvious or non-superficial factual truths about the universe and ourselves.[9] Non-scientific methods for forming beliefs about such matters tend to be unreliable. Just consider the many false beliefs that people have had throughout the ages about the origin and age of the universe, about the ultimate constituents of reality, human nature, and so forth. Although not infallible, science is significantly more reliable in the long run when it comes to such matters. At the same time, we shouldn't overestimate how reliable science is. There is, by now, solid evidence that more than half of published scientific results in the biomedical sciences, psychology, neuroscience, and the social sciences are false (Ioannidis 2005; Harris 2017). Such unreliable published results are not scientific knowledge in the sense in which I am using that phrase here, because they lack appropriate warrant or are false.[10]

That scientific knowledge is high-grade knowledge also means that scientific knowers ought to be able to *justify* knowledge claims (Gerken 2015). Having scientific knowledge requires not only that scientific beliefs are produced by reliable methods, scientists should also understand these methods and be able to explain them, providing reasons for thinking that their knowledge claims are true. If a biologist claims to know what the key drivers of random genetic mutation are, she ought to be able to explain herself, provide evidence, or point to literature where evidence is presented. Having scientific knowledge thus requires having access to and grasping the reasons why your beliefs are likely to be true.[11] This is why consulting an oracle that reports true claims with perfect reliability would never produce scientific knowledge.

Let's shift our attention to features of *collaborative* scientific knowledge next. First, *most scientific knowledge claims nowadays are credited to groups* rather than individuals. Multi-authored publications have become the default. When you trace the original source of a scientific knowledge claim, it will often be a group. You could try to downplay the significance of this by pointing out that publications typically have a single designated corresponding author and that this individual is the real knowing subject, while the others are merely auxiliary. That's simply false, however. Corresponding authorships don't imply anything about credit or contribution; the choice for who fills that role can be purely pragmatic. Alternatively, you might suggest that the principal investigator (PI) for a research project or team bears ultimate responsibility for the knowledge produced and should thus be credited as the primary knower. But this, too, is wrong. While PIs bear certain sorts of ultimate responsibility – financial, managerial, intellectual ownership of a project's key ideas – it's not the case that the PIs have an exclusive responsibility to keep track of all the epistemic commitments and outputs in projects. Doing so

is often impossible for any single agent, especially in large or multi-disciplinary projects. Hence, the point that many scientific knowledge claims are credited to groups stands.

Second, *the reality behind multi-authored papers is collaborative work*. Scientists work together on research projects in smaller or larger teams; sometimes with teams in other departments or institutions. They divide up the work amongst each other, according to expertise, skills, and availability. To take a toy example, scientist A might be responsible for finding and reviewing relevant literature; B for designing the survey and making sure it uses validated instruments; C for recruiting a sufficiently large and representative sample of test subjects; D for the online survey system; E for processing the results when they come in and getting them in the right data format; and F for carrying out appropriate statistical analyses. One or more of these people might write a first draft of a paper, others might contribute specific sections and paragraphs, or revise the first draft. Progress will be discussed so that everyone stays abreast of the project as a whole, tasks might be reassigned if needed, and sometimes new people are added to the team or team members leave. All of these tasks are necessary and make real and significant contributions to the project as a whole. It's not as if some of these activities are easily dispensable or unimportant; they are all necessary and must be carried out well if the project is to produce scientific knowledge. Even though specific team members might be more involved, bear greater responsibility, or grasp the intellectual underpinnings of the project better than others, much research is a genuine team effort. The earlier example is relatively simple, but, to the extent that projects become larger and multi-disciplinary or interdisciplinary, mutual dependence will only become greater and the involvement of multiple researchers even more inevitable.

Third, *teamwork is not just an accidental feature of contemporary science*, which could easily be reversed.[12] Many questions that scientists are working on are so large and complex that answering them necessitates teamwork. There are two dimensions in which this is true: practical and cognitive.

Teamwork is *practically necessary* because the work needed to complete a research project is simply too much for any one person to complete on their own. Consider, for example, the Human Genome Project, the purpose of which was to identify and sequence the more than 3 billion (!) chemical units (nucleotides) in human DNA. This massively collaborative project took about 15 years from inception to completion (1985–2000) and involved research teams from twenty universities in the United States, Europe, and Asia. The sheer amount of work required to complete it necessitated the collaborative set-up. Even if one individual had possessed all the relevant expertise and skills, it would have been impossible for her to complete all this work within one lifetime. This project is an extreme example, but many projects are too big for individuals to carry out on their own.[13]

Next, teamwork is often *cognitively necessary*. Many research projects require expertise, skills, and background knowledge from different disciplines or

sub-disciplines. Individuals usually don't have formal training and experience in all the relevant disciplines or sub-disciplines and it is impossible for them to make up for this on the fly. Interdisciplinary projects are natural examples of this, but monodisciplinary projects can require more expertise and skills than one individual can muster too. Let's look at an example of each. In recent years, the interdisciplinary field of 'digital humanities' has taken off. As a result of large-scale efforts to digitize historical records, it has become much easier than before for historians, literary scholars, and philosophers to survey and compare large collections of historical materials and to ask questions about them that would have been unanswerable before. (A very simple example: the origin, spread, and prominence of key terms in thousands of texts spanning several centuries.) Doing so in a methodologically sound way, however, requires the skills of both computer scientists and humanities scholars. The former to get the historical materials in the right formats to allow them to be investigated; the latter to ask the right questions and to be sensitive to the interpretational difficulties of reading historical texts. Next, many research projects that may look like they're monodisciplinary still require the knowledge and skills of more than one researcher. Developing new drugs and testing their effectiveness, for example, requires not only medical expertise but also statistical acumen. In fact, the input of statisticians is crucial in quantitative research in many disciplines. Moreover, scientific disciplines have several sub-disciplines. Researchers usually specialize in just one sub-discipline. Hence, projects that straddle the boundaries of sub-disciplines require more than one researcher. Cosmologists studying distant galaxies, stars, or planets, for example, collaborate with physicists working in atomic, molecular and optical physics to figure out how to build telescopes and to make sense of the data collected by them. In all such cases, then, acquiring scientific knowledge involves cognitive resources beyond any individual's capacities. Teamwork is cognitively necessary.[14]

No Collective Knowledge?

Now that we have a better grip on scientific knowledge and scientific collaboration, we are ready to grapple with the central question: Is collaboratively produced scientific knowledge collective knowledge and, if so, in what sense(s)?

Before we do so, however, it is helpful to see what is at stake. Traditionally, knowledge has been conceived as involving *individual* mental states and this is why many philosophers reject the idea that there can be genuinely collective knowledge. The philosopher of science Philip Kitcher (1994, p. 118) refers to 'the traditional conception of knowledge as something that is located in (or possessed by) an individual subject'. When you know that it is sunny outside, your mind is in a certain state that is connected in the right way to the actual weather conditions. This traditional understanding of knowledge provides a clear reason to be skeptical of collective knowledge: Groups don't have minds of their own and, hence, no mental states. Since knowledge either is a mental state itself or at least involves

one (to wit, that of belief), it follows immediately that groups cannot literally have knowledge. Case closed.

It is hard to object to this argument, because there is something undeniably right about it. Nonetheless, it flies in the face of the ease with which we think and speak about collective knowledge. Locutions such as 'Google knows everything about you'; 'The Republican leadership knew which candidate they wanted'; or 'The jury knew all the relevant facts' sound completely normal. Based on this argument, you might dismiss them all out of hand as purely metaphorical, but there is another option. Why not take their naturalness and ubiquitousness as evidence that there are other widely used and legitimate senses of knowledge, in addition to the traditional individualistic one (cf. Tollefsen 2002; Gilbert 2004)? Collective knowledge may differ from individual knowledge in not being exclusively tied to an individual mental state, but why should this be taken as a decisive reason to reject the existence of collective knowledge altogether?

That is the path I want to explore in the remainder of the chapter. I'll sketch three senses in which scientific knowledge can be collective knowledge. By the end of the chapter, we should have a clear view of how much contemporary scientific knowledge is intimately tied to collectives rather than individuals.

Collective Knowledge I: Production

As we saw earlier, much contemporary scientific knowledge is produced through collaboration – and inevitably so. This means that much scientific knowledge is collective in the straightforward sense that it has been produced by a group of people. It results from a group effort and it is officially credited to a group.

You might think this is a bit underwhelming as an account of collective knowledge. Shouldn't we simply distinguish between *producing* and *having* knowledge? Just because knowledge is produced by a collective doesn't mean that it is also had by the collective. To see that, suppose I'm trying to find out how many people are attending my party. I ask one friend to count the people outside, another those in the living room, and a third those in the kitchen. They report back to me and I add the numbers. Does this arrangement make my knowledge of the number of party people collective knowledge? I'm inclined to think not. It seems more accurate to say that I, individually, know how many people are at my party, even though my friends helped me obtain this piece of knowledge.

This worry prompts me to add some clarifications. First, a terminological point. Why precisely should the fact that multiple individuals were involved in producing a piece of knowledge *not* be enough to call the resulting knowledge collective? The term does not have a strictly delineated and clear meaning in everyday usage or academic philosophy, so we have some flexibility in fixing its exact meaning. Perhaps collective knowledge as collectively produced knowledge is not the most theoretically interesting sense of the term, but I don't think there

are strong reasons to resist this usage either, as long as we're clear about what we mean.

Second, two particularities of the party people example might change our unwillingness to ascribe collective knowledge: I'm the only one adding the numbers and I don't share the total with my friends. Suppose we modify the example so that I add the numbers and tell my friends, after which they check my calculation and correct me if I'm wrong. Now, it's more of a team effort. The mutual interaction and double checking also makes the example resemble scientific practice more. Arguably, this strengthens the motivation to ascribe collective knowledge.[15]

Third, an even more relevant difference between party people knowledge and scientific knowledge is that the collective production in the former case is entirely contingent. I could easily have gone around the house and done the counting myself. As we saw earlier, this is not so in contemporary science. Involvement of a collective is often practically or cognitively necessary. If you have qualms about using the term collective knowledge too liberally, this provides a principled restriction: Apply the term only if it is practically or cognitively necessary that knowledge is collectively produced. On this usage, party people knowledge is not collective (in either version of the case), but much scientific knowledge still is.

This last consideration – the practical and cognitive necessity of teamwork in science – provides what I take to be the strongest reason to see scientific knowledge as collective knowledge. That knowledge just happens to be produced by a collective might not be enough for 'real' collective knowledge, but if it (realistically) couldn't have been otherwise, we do get collective knowledge. Hence, the first sense in which a lot of scientific knowledge is collective is that it is knowledge that is collectively produced, where this could not have been otherwise for reasons of practical or cognitive necessity.[16]

Collective Knowledge II: Warrant

I have defined knowledge as warranted true belief. Because scientific knowledge is high-grade knowledge, scientific warrant must consist of explicit evidence and reasons (data, observations, analyses, inferences, etc.). Having scientific knowledge requires you to be able to justify your knowledge and this, in turn, requires that you can access the evidence and understand how it bears on the claim in question.

Now consider a case of interdisciplinary research; for instance, political scientists collaborating with computer scientists to investigate what happened in the right- and left-wing social media universes in the months leading up to the 2016 US elections. Warrant for their conclusions will depend, obviously, on whether the data gathered on social media and websites is representative and on whether it was reliably collected, processed, and analyzed. Securing these things requires expertise and skills from both computer scientists and political scientists. The former, for instance, for the technicalities of scraping social media data reliably and processing them into an analyzable format; the latter for identifying issues

on which to collect data, interpreting the data, identifying right- and left-wing individuals, organizations, or messages, and for putting the observed phenomena into a broader political science context. The political scientists will typically not be able to do computer scientists' work, nor will they be able to understand it all or verify its adequacy in any detail. They must simply rely on their colleagues' expertise. The same is true in the other direction. The computer scientists must also trust the political scientists.

This set-up implies that the full warrant for knowledge claims coming out of this investigation involves both evidence that can only be adequately grasped by computer scientists, and evidence that can only be adequately grasped by political scientists. Since having scientific knowledge requires access to the evidence and reasons that warrant a belief and understanding of how they support the belief, it follows that the *collective* of computer scientists and political scientists has scientific knowledge in the primary sense. No individual member of the interdisciplinary team has the expertise to appreciate everything that warrants the project's conclusions, even if some team members might individually have high-grade knowledge of partial or intermediate results.

There is nothing special about this particular example. The structure it exemplifies can be found in all seriously interdisciplinary projects, as well as in monodisciplinary projects that require expertise from different sub-disciplines. So a second sense in which much scientific knowledge is collective knowledge is that it is knowledge for which the warrant can only be possessed and understood by a collective.[17]

I should add two important clarifications. First, what I said shouldn't be taken to imply that when a team has collective scientific knowledge in the sense just specified, this knowledge is unavailable to individuals. Of course, they can acquire it by learning about it from a reliable oral or written source. But knowledge so acquired is not high-grade scientific knowledge in the sense specified earlier. When you acquire a piece of knowledge through testimony, you don't automatically also get all the original evidence supporting it, let alone a good grasp of how it does so. Rather, your warrant is testimonial and how good it is has to do with factors such as the sincerity and competence of the testifier, your ability to parse what is said or written, and your sensitivity to indications of unreliability (see Green (n.d.) for discussion).[18]

The general phenomenon here is that people can know the same things but know them with different kinds or degrees of warrant. This is familiar from everyday life too. If you saw a car accident and tell me about it, we know many of the same things, but you know them first-hand, through direct experience and memory, whereas I have second-hand testimonial knowledge of them (cf. Fricker 2006). Something similar holds in the earlier example. The team of computer scientists and political scientists collectively possesses high-grade scientific knowledge. When other scientists outside the team learn about their conclusions, they acquire testimonial knowledge of them. Collective knowledge in the present sense

can be shared with individuals, but they do not thereby automatically acquire high-grade scientific knowledge.

Second, the example was one in which collaboration was cognitively necessary, because of the different (sub-)disciplinary backgrounds required. But do we also get collective scientific knowledge in the same or a similar sense if collaboration is *practically* necessary? This is not obvious. In the interdisciplinary case, only the collective satisfied the conditions for having high-grade knowledge, because individual collaborators could not grasp all the relevant evidence. When collaboration is only practically necessary, however, individual collaborators could in principle access, understand, and evaluate all the evidence by themselves; they just don't have the time to do it.[19] This consideration could be a reason to deny that practically necessary collaboration gives rise to a similarly strong sense of collective knowledge. I believe, however, that we can do justice to this concern and also maintain that practically necessary collaboration does create collective knowledge. Since collaborating individuals mostly do not in fact access and assess all the evidence, but instead rely on their collaborators and assistants, a natural way to characterize their situation is that, even though they are *in a position* to acquire individual high-grade knowledge, they do not in fact possess it. Rather, the collective together has high-grade scientific knowledge.[20]

Hence, the second sense in which much scientific knowledge is collective knowledge is that the warrant required for it – that is, for high-grade knowledge – is possessed only by collectives and not by individuals.[21]

Collective Knowledge III: Function[22]

Knowledge plays various functional roles in our intellectual and practical lives. For instance, if you have knowledge, you can stop inquiry, unless you get new reasons for doubt. Knowledge is stored in memory, waiting to be retrieved and used. (If you can't retrieve a piece of knowledge, you forgot and you no longer know it.) Knowing something licenses you to use it in your practical or theoretical deliberations. If you want to get into law school and know that the best law schools select only applicants with GPAs well above 3.0, you had better study hard. Having knowledge is also connected to assertion: If you know something, you may assert it without qualification. If, in contrast, you merely believe or suspect something, you qualify your assertions accordingly. Knowledge is a basis for action: If you know, you can act on your knowledge.[23]

Scientific knowledge has many of these same functional roles. Scientific knowledge, especially if it is firmly established, stops further inquiry. Physicists now have conclusive empirical confirmation for the existence of the Higgs boson. Hence, they don't go on doing the same experiments, but move on to further questions. Scientific knowledge is stored in papers, books, proceedings, digital repositories, and the like. There, it can easily be accessed, either physically in libraries or electronically through the internet. Scientific knowledge is available for use in further

theoretical or practical deliberation. When mathematicians prove a novel theorem, it becomes part of the body of mathematical knowledge and mathematicians might then use it as a premise for further proofs. As for practical deliberation, once the causes of a disease have been identified, biomedical researchers might use that knowledge to devise new research projects aimed at finding a cure. Scientific knowledge also licenses assertions. Scientific papers obviously make assertions, but scientists also inform others about their field or offer expert advice to policy makers, private companies, courts of law, etc. Scientific knowledge also supports action. This happens within science when scientists build on each other's work to conceive new research projects, experiments, models, or theories, but scientific knowledge is also applied in technology, in policies, and in shaping people's worldview. Without knowledge of general relativity, our GPS systems wouldn't be at all accurate. Without systematic empirical study of hiring practices, we cannot know whether there is gender or racial discrimination in hiring and whether and how we should do something about it.[24]

Now note that, for many of these functional roles of scientific knowledge, research teams or larger collectives are the actors. As I emphasized repeatedly earlier, teams produce scientific knowledge and they possess the warrant for it. They also decide to stop inquiry. Once knowledge is stored in publications and digital repositories, it is available for retrieval and use by the scientific community and anyone else who has access. Research teams deliberate on the basis of scientific knowledge, they make assertions when appropriate, and they act on the basis of scientific knowledge, in conceiving new research plans, setting up novel experiments, or investing in equipment.

In short: Scientific knowledge plays many of the same roles for research teams and broader scientific communities as (individual) knowledge does for individuals. These strong functional parallels give rise to a third and final sense in which scientific knowledge is collective knowledge: Scientific knowledge functions for collectives as individual knowledge does for individuals.

Classifying or even identifying things and phenomena on the basis of functional similarities is a familiar style of reasoning. Technical artifacts, for instance, are grouped together on the basis of their functional roles. What chairs have in common is that they are for sitting on, even when they're very different qua physical structure and other functional roles. When early researchers in artificial intelligence used to say that the brain is – or is just like – a computer, they had in mind that it operates in the same ways, in spite of the differences in underlying hardware. Philosophers who defend the idea that groups can be agents with their own intentions and actions – or even their own minds – similarly rely on functional similarities with individual intentions, actions, and minds (cf. List and Pettit 2011; Chant, Hindriks, and Preyer 2014 for some recent proposals). For the case of knowledge specifically, we have been ascribing knowledge to animals for a long time and more recently people have started to talk about Facebook's algorithms or other online recommender systems possessing

intimate knowledge of us. Such uses are all based on observed functional similarities with human knowledge.[25]

A possible objection to this line of thought is that scientific knowledge doesn't bear *enough* functional similarity to individual knowledge. It may be similar in lots of ways, but not in all ways. Or not in the essential ways.[26] What I've said earlier already contains the ingredients for a reply, but let's make that explicit. As we saw, there aren't just one or two odd similarities between scientific knowledge and individual knowledge; there's a whole list. Moreover, several of the items on this list have been argued to be central roles for knowledge. Insisting that only 100% functional similarity would be enough to extend the concept of knowledge legitimately to collectives is overly strict. We extend concepts based on functional similarities in other domains more liberally than that, so why would knowledge be an exception?

Hence, much scientific knowledge is collective in the sense that it plays many of the central functional roles for scientific collectives that individual knowledge plays for individuals.

Conclusion

Large parts of contemporary science are collaborative. Intellectual and practical labor is carried out in groups. Of course, groups consist of individuals, so it's not as if individuals have dropped out of the picture altogether. Rather, the point is that much of what goes on in science is most naturally described and understood by taking groups, rather than individuals, as the primary agents. Contemporary science is teamwork and that is the only way it can be; projects are too cognitively demanding or too large for individuals to carry out on their own.

In consequence, a lot of scientific knowledge is collective knowledge in one or more of the three following senses. First, it is knowledge that is collectively produced and, for practical or cognitive reasons, could not have been produced otherwise. Second, it is knowledge for which the required high-grade warrant is possessed – and can only be possessed – by a collective. Third, it is knowledge that plays many of the central functional roles for collectives that individual knowledge plays for individuals.

These senses of collective knowledge are not mutually exclusive. A given piece of knowledge can be collective in two or even three of these senses at the same time. In fact, this might often be the case. If an instance of scientific knowledge is the result of practically or cognitively necessary collaboration, it will not only have been produced by a collective, but the warrant for it will often also be shared by the collective and the collective will be in a position to use this knowledge in a number of functional roles.

I believe these three senses capture central ways in which much contemporary scientific knowledge is collective knowledge, but I don't mean to claim that they exhaust the options. Some philosophers have argued that groups can

form collective beliefs by going through formal or informal decision procedures (Gilbert 1987; Wray 2007). Maybe this sometimes happens in science too, for instance in so-called consensus conferences where scientists try to reach agreement on the state of knowledge on pressing issues in their field. Others have argued that the conditions for collective knowledge are less demanding: It suffices if enough members of a group believe something and they jointly possess warrant for what is believed (Klausen 2015). There is room for further thinking here.[27]

To get back to the title of this chapter: How many scientists does it take to have knowledge? Occasionally, perhaps one is enough, but the earlier discussion shows that the popular myth of the lone scientific genius makes little sense in contemporary science.[28] Contemporary science is first and foremost a team game and scientific knowledge is had primarily by smaller or larger groups of scientists.

Acknowledgments

I'm grateful to the editors, Kevin McCain and Kostas Kampourakis, and to Valentin Arts, Wout Bisschop, Lieven Decock, Tamarinde Haven, Rik Peels, Chris Ranalli, and René van Woudenberg for helpful comments on an earlier version of this chapter. Research for this chapter was made possible through a Vidi grant (276-20-024) from the Netherlands Organization for Scientific Research (NWO).

Notes

1 See: www.economist.com/news/science-and-technology/21710792-scientific-publications-are-getting-more-and-more-names-attached-them-why.
2 www.nature.com/news/seven-days-3-9-august-2012-1.11139#/trend.
3 See this report in *Nature*: www.nature.com/news/physics-paper-sets-record-with-more-than-5-000-authors-1.17567.
4 www.nature.com/news/fruit-fly-paper-has-1-000-authors-1.17555.
5 Tilak, Prasad, and Jena (2015) investigate these issues in biomedical publications.
6 See Resnik et al. (2016) for a survey.
7 See Huebner, Kukla, and Winsberg (2017) for discussion.
8 That is, given (i) that the knower existed and (ii) was in a position to acquire the belief, it's not a matter of luck that her belief is true. See Pritchard (2005) for more on luck in epistemology.
9 For the purposes of the discussion, I'm disregarding the possibility of other sources of knowledge about the fundamental nature of reality, such as divine revelation.
10 Note that we sometimes use 'scientific knowledge' more loosely to cover everything that is published in scientific outlets, but that is not how I'm using the phrase here.
11 Epistemologists call this sort of justification for knowledge claims *internalist* justification (BonJour 2010).
12 See Hardwig (1985; 1991) for early observations to this effect.
13 I'm glossing over the fact that practical necessity may have vague boundaries. Suppose one scientist could complete a project on her own, but it would take her five years, during which she could not work on anything else. Realistically, this couldn't be done

in most academic settings nowadays. So is it practically necessary that a team is involved? It seems so. Alternatively, maybe one person could finish a project within reasonable time, but only if she puts in 70-hour weeks for six months. Is it practically necessary to get a team involved? Perhaps not quite, but it certainly seems highly preferable. Even with vague boundaries, however, the notion latches onto something real.

14 Cognitive necessity is bound to be vague in ways similar to practical necessity; cf. the previous note. Moreover, it could change as a result of technological advances such as cognitive enhancement.

15 To push back, you could suggest that, in the modified case, three individuals acquired a piece of knowledge together and now know the same thing, but there is no collective knowledge. My response would be the same as before: I appreciate this consideration, but I don't see why it should be taken as a decisive reason against calling the result collective knowledge.

16 See Goldberg (2010) for further discussion of the fact that we rely on others for much of what we know.

17 I've defended the idea that much scientific knowledge is collective knowledge in this sense in more detail elsewhere (De Ridder 2014).

18 Objection: if I learn about general relativity by reading a popular science book, don't I acquire scientific knowledge? Yes and no. Yes, in the sense that I acquire knowledge that is the result of scientific inquiry. But no in the sense that I do not acquire high-grade knowledge of general relativity. This would require deep familiarity with and understanding of the evidence, which I don't get from reading one popular science book.

19 Although note that, the larger the collaboration becomes, the more work the 'in principle' clause is doing. Looking at 3 billion chemical building blocks of human DNA is not something an individual could actually do, even with the right expertise.

20 Note that it is a consequence of this proposal that individuals can sometimes – when the collaboration isn't too big – acquire high-grade scientific knowledge of propositions that initially came to be known through a practically necessary collaboration. If a team member really takes the time and effort to sift through all the evidence, to familiarize herself with it, and to check the analyses and inferences, she acquires individual high-grade knowledge.

21 Miller (2015) defends a similar idea, which turns on the insight that scientific knowledge requires *strong* evidence. Because gathering sufficiently strong evidence often requires the efforts of many people and multiple teams, he, too, concludes that some scientific knowledge is primarily had by collectives rather than individuals.

22 Inspiration for the line of thought in this section comes from Bird (2010).

23 For discussion of some of these roles, see Benton (n.d.).

24 The connection between scientific knowledge and action is also behind evidence-based medicine, policy, and decision making. Whenever possible, we ought to base our medical and policy interventions on strong scientific evidence.

25 Note that, if one thinks that the concept of knowledge itself is a functional role concept, it becomes even easier to defend the line of thought espoused in this section. Craig (1990) can be read as defending a version of this idea.

26 See Lackey (2014a) for an objection along these lines to Bird's (2010) proposal.

27 Two recent collections of essays devoted to group epistemology are Lackey (2014b) and Brady and Fricker (2016).

28 Contributions by Kostas Kampourakis and Kathryn Olesko in Numbers and Kampourakis (2015) show that it is also historically inaccurate.

References

Benton, M. (n.d.) 'Knowledge Norms', *Internet Encyclopedia of Philosophy*, www.iep.utm.edu/kn-norms/. (Accessed: July 30, 2018)

Bird, A. (2010) 'Social Knowing: The Social Sense of "Scientific Knowledge"', *Philosophical Perspectives* 24, pp. 23–56.

BonJour, L. (2010) 'Externalism/Internalism', in: Dancy, J., Sosa, E., and Steup, M. (eds.) *A Companion to Epistemology*. Oxford: Wiley-Blackwell, pp. 364–368.

Brady, M., and Fricker, M. (eds.) (2016) *The Epistemic Life of Groups*. Oxford: Oxford University Press.

Burge, T. (2003) 'Perceptual Entitlement', *Philosophy and Phenomenological Research* 67(3), pp. 503–548.

Chant, S., Hindriks, F., and Preyer, G. (eds.) (2014) *From Individual to Collective Intentionality: New Essays*. New York: Oxford University Press.

Craig, E. (1990) *Knowledge and the State of Nature*. Oxford: Oxford University Press.

De Ridder, J. (2014) 'Epistemic Dependence and Collective Scientific Knowledge', *Synthese* 191(1), pp. 37–53.

De Ridder, J. (2018) 'Kinds of Knowledge, Limits of Science', in: de Ridder, J., Peels, R., and van Woudenberg, R. (eds.) *Scientism: Prospects and Problems*. New York: Oxford University Press, pp. 190–219.

Fricker, E. (2006) 'Second-Hand Knowledge', *Philosophy and Phenomenological Research* 73(3), pp. 592–618.

Gerken, M. (2015) 'The Epistemic Norms of Intra-Scientific Testimony', *Philosophy of the Social Sciences* 45(6), pp. 568–595.

Gilbert, M. (1987) 'Modelling Collective Belief', *Synthese* 73(1), pp. 185–204.

Gilbert, M. (2004) 'Collective Epistemology', *Episteme* 1(2), pp. 95–107.

Goldberg, S. (2010) *Relying on Others*. New York: Oxford University Press.

Green, C. (n.d.) 'The Epistemology of Testimony', *Internet Encyclopedia of Philosophy*, www.iep.utm.edu/ep-testi/. (Accessed: July 30, 2018)

Hardwig, J. (1985) 'Epistemic Dependence', *Journal of Philosophy* 82(7), pp. 335–349.

Hardwig, J. (1991) 'The Role of Trust in Knowledge', *Journal of Philosophy* 88(12), pp. 693–708.

Harris, R. (2017) *Rigor Mortis*. New York: Basic Books.

Huebner, B., Kukla, R., and Winsberg, E. (2017) 'Making an Author in Radically Collaborative Research', in: Boyer-Kassem, T., Mayo-Wilson, C., and Weisberg, M. (eds.) *Scientific Collaboration and Collective Knowledge: New Essays*. Oxford: Oxford University Press, pp. 95–116.

Ioannidis, J. (2005) 'Why Most Published Research Findings Are False', *PLoS Medicine* 2(8), p. e124, https://doi.org/10.1371/journal.pmed.0020124.

Kitcher, P. (1994) 'Contrasting Conceptions of Social Epistemology', in: Schmitt, F. (ed.) *Socializing Epistemology*. Lanham, MD: Rowman & Littlefield, pp. 111–134.

Klausen, S. (2015) 'Group Knowledge: A Real-World Approach', *Synthese* 192(3), pp. 813–839.

Lackey, J. (2014a) 'Socially Extended Knowledge', *Philosophical Issues* 24(1), pp. 282–298.

Lackey, J. (ed.) (2014b) *Essays in Collective Epistemology*. New York: Oxford University Press.

List, C. and Pettit, P. (2011) *Group Agency*. Oxford: Oxford University Press.

Miller, B. (2015) 'Why (Some) Knowledge Is the Property of a Community and Possibly None of Its Members', *Philosophical Quarterly* 65(260), pp. 417–441.

Numbers, R., and Kampourakis, K. (eds.) (2015) *Newton's Apple and Other Myths About Science*. Cambridge, MA: Harvard University Press.

Plantinga, A. (1993) *Warrant and Proper Function*. New York: Oxford University Press.

Pritchard, D. (2005) *Epistemic Luck*. Oxford: Oxford University Press.

Resnik, D., Tyler, A., Black, J., and Kissling, G. (2016) 'Authorship Policies of Scientific Journals', *Journal of Medical Ethics* 42, pp. 199–202.

Tilak, G., Prasad, V., and Jena, A. (2015) 'Authorship Inflation in Medical Publications', *INQUIRY* 52, pp. 1–4.

Tollefsen, D. (2002) 'Challenging Epistemic Individualism', *ProtoSociology* 16, pp. 86–117.

Wray, K. (2007) 'Who Has Scientific Knowledge?', *Social Epistemology* 21(3), pp. 337–347.

2

WHAT ATTITUDE SHOULD SCIENTISTS HAVE?

Good Academic Practice as a Precondition for the Production of Knowledge

Thomas A.C. Reydon

Setting the Stage

Good practice in science has long been a topic of discussion among scientists, professional scientific organizations, ethicists and philosophers of science, university administrators, governments, as well as the general public. To a large extent the discussions have been motivated by cases of deceit, fraud and other sorts of misbehavior by scientists that were widely given attention in the media – cases in which researchers fabricated results in their entirety, "cleaned up" data sets to better fit the view they wanted to advance, sabotaged the research projects of their competitors (Maher, 2010), and so on. Closer examinations of the history of science and the contemporary situation show that there is no shortage of such cases.[1]

Unsurprisingly, scientists and professional scientific organizations began to worry that more frequent occurrences of such cases (as well as increasing attention for them in the media) would lead to a decline in public trust in science, as well as in individual scientists.[2] After all, if scientists cannot be trusted to honestly report the results of their work, to not make up data and present them as outcomes of actual measurements, and to not distort the results of their work in order to make them look better, how can science be trusted to produce genuine knowledge? These worries show how epistemological issues and questions of good practice are intimately related. In response to such worries, in the past few decades most universities, research institutions, and funding organizations have implemented regulations and/or guidelines to safeguard good scientific practice, and have set up courses in research ethics for students and staff. Virtually all national academies of science and international scientific organizations have published memoranda on best practices in scientific research (e.g., Steneck, 2007; ESF, 2008; SAMS, 2008; UKRIO, 2009; ESF/ALLEA, 2011; DFG, 2013; VSNU, 2014). The overarching

aim of such measures is to restore and strengthen public trust in science by improving the ethical standards of conduct of scientists.

Most of the measures that have been implemented focus on preventing misbehavior in science, increasing the awareness among scientists of morally problematic situations that they might encounter in the various contexts of their work, and providing institutions as well as individuals with procedural guidelines for dealing with cases of (suspected) misbehavior. The relevant regulations, guidelines and policy documents are often conceived of as addressing *moral* issues, and accordingly emphasize traditional moral virtues and values such as honesty, fairness with respect to collaborators in research projects and co-authors of papers, and so on. The European Science Foundation's *European Code of Conduct for Research Integrity*, for example, mentions "honesty in communication; reliability in performing research; objectivity; impartiality and independence; openness and accessibility; duty of care; fairness in providing references and giving credit; and responsibility for the scientists and researchers of the future" (ESF/ALLEA, 2011, p. 5) as the main principles of research integrity. Similarly, the Office of Research Integrity's *Introduction to the Responsible Conduct of Research* mentions honesty, accuracy, efficiency, and objectivity as the main "shared values for the responsible conduct of research that bind all researchers together" (Steneck, 2007, pp. 2–3). Clearly, these are not *strictly* moral aspects of research, but following the written and unwritten rules of good scientific practice is still often seen as behaving in a morally acceptable way as a scientist.

No doubt, many aspects of everyday behavior in scientific settings indeed have to do with morality. After all, science is a human endeavor and as such takes place in social settings in which moral issues arise from the interactions among people. Moreover, many research projects involve and potentially affect sentient beings – think of biomedical projects involving experiments on animals, psychological research and clinical trials involving human test persons, and more generally research that can be expected to yield results that may affect specific groups of people or society at large (such as field trials with genetically modified plants). But thinking of the domains of good scientific practice or responsible conduct of research as first and foremost domains of applied ethics, I want to argue, is mistaken. In contrast, I want to propose that next to the moral issues that play a role in science, good practice and responsible conduct of research are essentially connected to the question how a person should perform *in the role of a scientist*, that is, in the role of someone tasked with the production of knowledge.

In this chapter, I discuss some of the principal elements of good scientific practice from an epistemological perspective and try to show that the normativity that these elements embody is not primarily a form of moral normativity, but of goal-directed normativity. That is, I argue that the rules and guidelines of good scientific practice are not moral imperatives per se that pertain to right and wrong conduct as such (even though some moral aspects of scientific work should be addressed), but rather specifications of what it is for a scientist or an academic more generally

to do a good job *as an academic*. Given that the aim of science and hence the job of scientists is to produce knowledge (which I take as an uncontroversial claim), doing a good job as a scientist implies doing a good job in producing knowledge.

From Good Scientific Practice to an Academic Attitude

I will begin by making a few clarifications regarding the scope of the ideas presented here. First, arguably producing knowledge is the job of people working in academia in general and not only in the natural sciences. There are ongoing debates in epistemology and in the philosophy of science on the question of under what conditions we can say that we have knowledge in a strict sense, and whether areas of investigation outside (natural) science produce knowledge in this sense at all or something else. Often, for example, the explanation of phenomena is seen as central to scientific knowledge (see for instance Kevin McCain's chapter in this volume). But according to a traditional dichotomy, the sciences explain phenomena by providing accurate descriptions of how phenomena occur that involve true generalizations (laws of nature), while academic areas of investigation that cannot be counted as science do not in fact explain phenomena, but rather contribute to our *understanding* of them.[3] Research in physics, for example, is aimed at providing explanations of physical phenomena, whereas research in history supposedly is aimed at a better understanding of what happened in particular episodes of human history, but not at actual scientific explanations of those occurrences. Because of its lack of explanatory content, such understanding is sometimes thought of as not being a kind of knowledge – De Regt (2015), for example, suggested that it should be seen as a kind of skill – and accordingly fields that do not provide explanations in a strict sense are often seen as not producing knowledge in any strict sense.[4]

Be that as it may, I think that when it comes to good practice we should be concerned with *academic* practice, that is, the practice of work in all areas of academia, as they all produce some sort of epistemic content. Talking about good *scientific* practice is too restrictive, I think, as it suggests that the discussion only applies to the natural (and possibly the social) sciences. In addition, talking about the responsible conduct of *research* seems too restrictive, too. If we restrict our thinking about what it means to be a good scientist (or rather, a good academic) to research contexts and the rules and regulations that are implemented at academic institutions to research only, we are missing important and large parts of everyday academic practice, such as teaching, public engagement, and service to the profession. For instance, editors of academic journals and reviewers who evaluate manuscripts submitted for publication make decisions about which results will be published and become available to the academic community, thus directly affecting the epistemic output of the community. The scope of discussions on good academic practice, then, should be much broader than only concerning research, and include all aspects of academic work.

Lastly, I think that the aim of guidelines on good academic practice, codes of conduct, regulatory frameworks implemented by universities, research institutions, and funding organizations, and introductory textbooks on the topic should be to foster what I call an *academic attitude*. Rather than aiming at establishing sets of fixed rules for good conduct, final solutions for normative problems, or sets of fundamental principles for how to be a good academic, efforts should be aimed at installing and strengthening a general attitude in future and already practicing academics towards what it means to do a good job as an academic. This attitude may encompass specific views of what academic integrity encompasses (cf. Matthew Brown's chapter in the present volume), what academics are responsible for (and what they cannot be held responsible for), what scientific misconduct encompasses, and so on. But as it is unlikely that any general agreement on all these aspects of the discussion will ever be reached, I think it is more fruitful to aim at cultivating an attitude of awareness of the various issues that may arise and of developing one's own views about how to deal with these issues in practice than to aim at generally valid rules and regulations, sets of principles, guidelines, and the like.

What I would like to achieve in what follows, is to take a few small steps towards the development of such an academic attitude by showing how the various core elements of good academic practice are connected to considerations of how knowledge production processes in academia work, and of how they could work better. Two areas of responsibility of scientists should be considered in this context: responsibilities that follow from *having knowledge* that others do not have, and responsibilities that follow from one's role as part of the *knowledge production* process. As will become clear in the following sections, though, the two areas are intimately connected.

Responsibilities Due to *Having* Knowledge

The dropping of the two atomic bombs on the Japanese cities of Hiroshima and Nagasaki in August 1945 marked an important point in the contemporary discussion on the responsibility of scientists and of the scientific community. The magnitude of the loss of lives and of the material destruction that had been caused in the two events kindled a feeling of deep unsettlement, which became widely shared by the scientific community and the general public. Had science finally gone too far by making the development of such powerful weapons possible? Had science now lost its innocence as a presumed morally neutral quest for the truth, or a quest for knowledge in the service for mankind? And had the events that occurred in August 1945 highlighted a new kind of responsibility of individual scientists or of the scientific community as a whole, namely the responsibility for keeping the potential negative consequences of their work in check?

Irrespective of how one answers these questions, the events at the end of World War II put "responsibility in science" on the map as a topic of consideration. It

is, however, far from clear what exactly "responsibility in science" encompasses. When it comes to the specific responsibilities of scientists, many authors "would argue that since scientists helped to create nuclear weapons, the scientific community today has a profound responsibility to help reduce and ultimately disarm them" (Rees & Browne, 2010). The idea behind a position like this is that the work that scientists do to *produce* particular items of scientific knowledge – theories, sets of solutions to sets of mathematical equations, conceptual advances that enable further advances in science, particular techniques, and so on – entails a responsibility for how the resulting knowledge is used. The thought is that doing research that yields results that make a particular technological application possible entails a responsibility for this application and its consequences on the part of the scientists who have achieved these results. But how involved must a scientist have been in the development of a particular application to bear a responsibility for its consequences? In the case of the atomic bombs, scientists who had not been directly involved in the development of the bombs but had provided crucial contributions to the development of the underlying theory, such as Werner Heisenberg, indeed felt a deep responsibility for what had occurred. Arguably, though, the responsibility of a scientist who worked in the context of a project that was specifically aimed at the development of an atomic bomb (such as the Manhattan Project) is very different from that of a scientist who only did theoretical work that enabled such later projects.

Two authors who have argued in this latter direction are physicist and Nobel Prize laureate Percy Bridgman, and developmental biologist Lewis Wolpert. Bridgman emphasized that practicing scientists neither possess a special moral competence nor the time to consider the possible consequences of their research, such that having to worry about moral responsibilities entailed by their work would hinder them in performing their research. According to Bridgman, "the justification for the favored position of the scientist is that the scientist cannot make his contribution unless he is free, and the value of his contribution is worth the price society pays for it." (Bridgman, 1947, p. 149). Wolpert distinguished between science and technology and argued that

> In contrast to technology, reliable scientific knowledge is value-free and has no moral or ethical value. Scientists are not responsible for the technological applications of science; the very nature of science is that it is not possible to predict what will be discovered or how these discoveries could be applied.
> (*Wolpert, 2005, p. 1253*)

Similarly to Bridgman, Wolpert felt that individual scientists should not be held responsible for consequences of the application of the knowledge that they contributed to, and that lacking specific ethical or political expertise scientists could not be expected to be involved in the making of decisions on such applications (Wolpert, 1989, p. 943; 2005, p. 1254).

Wolpert and Bridgman both rejected the view that having played a part in the production of knowledge entails any particular responsibility. But Wolpert did not agree with Bridgman that scientists should be free to do their work without having to consider the possible consequences of their work at all. On Wolpert's view, the fact that scientists have special access to knowledge (i.e., that they possess knowledge of their particular area of work that outsiders do not possess) entails the special obligation to examine the possible societal consequences of their work and to communicate these to the public (Wolpert, 1989, p. 942–943; 2005, p. 1253–1254). As he put it:

> The social obligations that scientists have […] comes from them having access to specialized knowledge of how the world works that is not easily accessible to others. Their obligation is to both make public any social implications of their work and its technological applications and to give some assessment of its reliability.
>
> (*Wolpert, 2005, p. 1254*)

What we see in Wolpert's view, in contrast to Bridgman's, is an element of the responsibility of academics that is derived from epistemological considerations. Bridgman focused on moral responsibilities and argued that scientists should not be thought of as having any moral responsibilities in connection to the knowledge that they helped produce. Wolpert's consideration, in contrast, is that *having* knowledge entails an obligation to think about what this knowledge might mean for society and the people that are part of it, for better and for worse. In the case of academics, this unpacks as the view that having privileged access to a particular area of knowledge that the general public does not have access to entails the obligation to examine its possible consequences for society and humanity, and communicate these to the general public. Clearly this does not hold only for *research* scientists. Those who teach but are not active in research, have a similar access to specialized knowledge and hence similar obligations. And it does not only hold for *scientists* in a strict sense either. Scholars in all other areas of academia, such as philosophers and historians, also have privileged access to particular areas of knowledge, and hence should be thought of as having similar obligations as scientists do.[5] Wolpert did not provide a conclusive argument for this view, and nor will I argue for it here – but I will take it as sufficiently plausible to accept as a central tenet in the account of the academic attitude that I want to develop here.

Wolpert recognized that the possession of knowledge entails a special responsibility, namely responsibility to explore the possible consequences of various usages of this knowledge and inform the general public accordingly. Having privileged access to an area of knowledge means having expertise that others lack, which makes scientists the ideal persons to think about the possible consequences of their research. This observation is not novel. As early as 1624 in his *New Atlantis* Francis Bacon sketched the scientific community as consisting of various groups, each

with a specific task in the knowledge production process. Among the tasks that Bacon distinguished were "looking into the experiments of their fellows, and cast about how to draw out of them things of use and practice for man's life and knowledge" (Bacon, 1906, p. 273) and to "have consultations, which of the inventions and experiences which we have discovered shall be published, and which not; and take all an oath of secrecy for the concealing of those which we think fit to keep secret" (Bacon, 1906, p. 274).[6] In addition, Bacon (1906, p. 275) wrote, "we have circuits or visits, of divers principal cities of the kingdom; where, as it cometh to pass, we do publish such new profitable inventions as we think good." The gist of Bacon's thinking here is, as is the case for Wolpert too, that having access to an area of knowledge that the general public does not have access to entails responsibilities with respect to thinking about possible uses and consequences, and communicating about them to the members of society at large. Today, these are widely recognized integral aspects of good academic practice.

Note that the normativity involved in relation to the possession of knowledge is not moral normativity but goal-directed normativity. My argument for this view rests on the points made by Bridgman and Wolpert: academics *as academics* (with the exception of philosophers) do not have special ethical knowledge or training that would allow them to make moral decisions as part of their professional role. The obligations entailed by having access to a body of knowledge should therefore not involve obligations to decide on potential ways in which this knowledge can be used but should be thought of as obligations to *enable* public discussion and decision making. These obligations are goal-directed in nature: academics bear a role responsibility to not only produce new knowledge, but also to bring the knowledge they have special access to as part of their professional role into the public domain, highlighting possible beneficial and adverse usages, so that the members of society can make well-informed decisions. Such outreach activities are part of the academic's professional role, and academics who publish results without contextualizing them with considerations of possible beneficial and harmful consequences can be said to have neglected an important aspect of their job. But there is no reason to think of them as having violated any moral norms.

Summarizing, the point is that if one has access to particular knowledge as part of one's public role, this entails that one has particular responsibilities as a *keeper* of this body knowledge, and being a keeper of a body of knowledge involves examining what this body of knowledge could mean for society and the people who are part of it, as well as communicating one's findings to the public. I think that this understanding of one's role as an academic should be seen as a part of the academic attitude that should be cultivated in universities and other academic institutions.

Responsibilities in the Context of *Producing* Knowledge

Academics are not only in positions with privileged access to particular areas of knowledge, however, they also (and foremost) are *producers* of knowledge. In this

section I want to explore some aspects of responsibility that surface in the context of knowledge production in academia.

Let me begin by pointing out that in a straightforward sense someone who produces something, or takes part in the production of something, bears a responsibility for the quality of the product (or at the very least for that part of the final product that they were involved in making). I take this to be uncontroversial. In the same way, one can say that academics who are part of knowledge production processes bear a responsibility for the quality of the knowledge that they help produce – or at the very least for that element in the production of which they were involved. What I have in mind here is a responsibility for the knowledge itself, that is, for how well it is supported, for how reliable it is, and so on, not for how it is used.

While some readers may not agree with my (perhaps overly business-like) characterization of science and more widely, academia, in terms of production, thinking of the relationship between academics and knowledge in terms of producers and products does open up an illuminating perspective on authorship and related issues. Consider the example of plagiarism.[7] Widely used definitions of scientific misconduct center around a core definition of scientific misconduct as "fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results" (Steneck, 2007, p. 20).[8] One might wonder why plagiarism is one of the core three categories of scientific misconduct, though. Plagiarism is not unique to academia, after all. At base, plagiarism is theft of the ideas or creative work of another person and presenting them as one's own. As such, plagiarism occurs in all areas in which ideas and creative work play a central role – literature, music, the visual and performing arts, the design, production and marketing of various kinds of goods (brand name shirts can be plagiarized, for example), and so on.

From a general ethical perspective, there are several reasons why plagiarism is wrong. For one, plagiarism is a kind of theft – theft of intellectual property. This is not different in academia and in other areas: theft of someone's research contribution and presenting it as one's own is not categorically different from theft of someone's literary work and presenting it as one's own. In addition, plagiarism can constitute a breach of copyright when parts of published works are plagiarized (or in the case of products, when the look and brand name of a product is imitated), and this can be the case in academia as well as elsewhere. More specifically for academic contexts, plagiarism can constitute fraud in cases in which plagiarized results are used in a work that is submitted to obtain an academic degree. In recent years, for instance, Germany has seen a number of publicly discussed cases in which politicians had been accused of having plagiarized substantial parts of their dissertations on the basis of which they had been awarded their doctorates. One important aspect of these cases was that the persons in question had allegedly committed fraud in their attempts at meeting the requirements for being awarded an academic degree. Fraud is not specific to academia, however, and as such is not an instance of specifically *scientific* or *academic*

misconduct – when it occurs in academic context, it is misconduct in academia, but not academic misconduct.[9]

What, then is the difference between misconduct in academia and academic misconduct? As I see it, the difference is that between misconduct in academic settings that does not directly (or much less directly) affect the content of the products of academic work (i.e., knowledge) and misconduct that does so directly (or much more directly). Some instances of misconduct, such as the fabrication or falsification of research data or representations of data, directly affect the body of knowledge that is the product of academic investigations. A paper that contains data that have been freely invented by its author, or that contains data that have been "cleaned up" to better fit the view the author favors, clearly contains unreliable information. Other researchers have to be able to trust that published work is solid in order to be able to use it in their work, and papers containing fabricated or falsified data can – if the distortions remain unnoticed – pollute the output of research and the basis on which further research is conducted. Because of their direct consequences for the product of academic research, fabrication and falsification are considered to be academic misconduct and not merely misconduct in academia.

A similar point holds for plagiarism. With authorship of a publication comes a responsibility for the knowledge contained in it. An author of an academic publication guarantees with their good name in the relevant academic community that the knowledge (in the broadest sense, including understanding, insight, etc.) contained in the publication has been produced with, to their abilities, the best possible methods, the best possible data analysis, the best possible argumentation, the best possible background literature study, the best possible theoretical interpretation, and so on. In brief, authors guarantee with their good name that the content of their publications are reliable and of good quality in much the same way as a baker's good name stands for the quality of their baked goods. This is why so-called "gift authorship," that is, listing someone (usually famous or important in the relevant community) as an author who did not contribute anything to the work that is being reported, is counted as scientific misconduct (Smith, 2000). A "gift author," after all, cannot even partly stand for the quality of the published results, as they did not contribute to producing it. A recent development in academic publishing makes this relation of responsibility between authors and their products more explicit. Scientific journals increasingly have begun to ask authors of articles to specify who contributed what to the publication. Thus, one increasingly finds statements like "A.B. did the experiment, C.D. and E.F. prepared the samples, G.H. performed the statistical analysis, I.J. wrote the manuscript, …" as disclaimers in journal articles. Rather than simply having all authors take equal responsibility for the entire content of the paper, in this model individual researchers take responsibility for parts of the production process, and do so explicitly.

The point regarding plagiarism is that it breaks the relation of responsibility between author and product. An author who copies parts of someone else's

publication without providing adequate citations or steals someone's intellectual work without attribution cannot stand for the quality of the work that has been copied, as they were not involved in its production. Providing citations is a way of deferring to the original author of the text or the ideas that one uses, such that one does not need to be able to guarantee the quality of the used texts or ideas oneself – the original author does. In addition, providing citations makes knowledge items traceable to the context in which they originated. Individual researchers as well as research groups and labs usually work within particular paradigms, theoretical frameworks, metaphysical worldviews, and so on, that affect the knowledge they produce, such that understanding knowledge items involves having some information about the context in which they originated. Thus, when the connection between authors and their products is severed, important epistemic features that help place knowledge items into context become at least partly lost. Thus, plagiarism in academic work constitutes both misconduct in academia (because of its general moral aspects) and academic misconduct (because of its epistemic consequences).[10]

Note that responsibilities in relation to the production of knowledge do not only arise in contexts of academic misconduct. Kendig (2016), for example, recently examined cases of what is called "proof of concept research" and argued that such research opens up new epistemic categories that in turn entail ethical categories. Such new categories, both the epistemic and the ethical ones, give rise to new questions that can be asked, new methods of investigation, new tools for research, and so on. According to Kendig, "emerging technologies each present fundamentally new sets of ethical issues" and the new ethical categories are both novel and proper to the new field of investigation, as they track the epistemic categories that the new field has introduced (2016, p. 741). Kendig considers Synthetic Biology as an example of such a newly emerging field that aims at the re-engineering of life forms or even creating them *de novo*. An example of the former are genomically re-engineered Cyanobacteria that have been prepared for the production of biofuel (Kendig, 2014; 2016, p. 739–740), while a widely discussed example of the latter is a synthetic *Mycoplasma* bacterium that was created at the J. Craig Venter Institute, named *Mycoplasma mycoides* subsp. *JCVI syn1.0* (Gibson et al., 2011; Kendig, 2014, 2016, p. 738).[11] In such contexts new epistemic categories of life forms are created, which are likely to raise new ethical issues: once we have obtained a new category of life form, we are faced with the question how to treat the members of the category. As Kendig pointed out

> Knowledge of *what it is* provides information to us about how we are to behave towards it or in our relationship with it. Is it a moral subject?, a moral object?, or something worthy of our moral consideration? Arguably, these kinds of questions are only answerable once we know what kind of thing we are talking about. Put another way, once we know *what* it is, we start thinking about how we should act towards it. (*2016, p. 744; original emphasis*)

What we have seen in this section are two different examples of how participation in knowledge production processes may give rise to normative questions. While in the case of scientific misconduct it is, I hope, clear how normative issues arise for individual researchers, in the case of emerging technologies or areas of research this might be less clear. For clarification, I refer to the previous section: researchers in emerging areas do not only participate in the production of knowledge that gives rise to new ethical issues, they also are the ones who should (at least help) address these ethical issues, as they have privileged access to the knowledge they helped produce. Arguably, then, cases of emerging areas of research are cases in which responsibilities follow for researchers from a combination of their being part of the knowledge production process and their having privileged access to a particular body of knowledge.

As was the case in the previous section, here, too, the normativity involved is not primarily moral normativity but goal-directed normativity. When it comes to scientific misconduct, the normative issues are clearly connected to the aim of science, namely the production of knowledge. The ethical issues that arise with respect to new epistemic categories in emerging areas of research will in part be moral issues (such as issues regarding how to treat the newly created beings, issues regarding risks they pose to humans and the environment, etc.), but in part will also involve goal-directed normative issues that follow from researchers having privileged access to particular knowledge.

Finally, I want to suggest that here, too, we could benefit from thinking of scientific misconduct (or rather, the prevention of it) and emerging ethical categories in terms of fostering an academic attitude. With respect to scientific misconduct, I believe that fostering an understanding with students and practicing researchers that as academics they bear responsibilities for the knowledge they produce is more effective than simply explaining that "FFP" and other categories constitute unacceptable behavior, or setting up codes of conduct or sets of rules. When it comes to emerging ethical categories, having the right attitude perhaps matters even more, because it will be largely up to the researchers themselves to identify potential normative issues relating to their work – as they are creating new epistemic categories, they are at the very forefront of research and thus in the best position to address normative questions.

Outlook

I have argued that good academic practice (or responsible conduct of research) is less a matter of morality than of goal-directed normativity. The goal is the goal of science, which I have taken as the production of knowledge, and the normativity stems from the role someone plays in achieving this goal. When someone assumes the role of an academic, be it as research scientist in an area of the natural sciences or the engineering disciplines, or as a researcher in the humanities or the social sciences, that person assumes particular responsibilities that come with this role.

I have argued that these responsibilities can be thought of as falling into two categories: responsibilities due to having privileged access to a particular domain of knowledge and responsibilities due to playing a part in the production of particular knowledge items. My aim was not to be exhaustive, but merely to illustrate how epistemology and normativity hang together in academic work.

I have suggested that this area of normativity is best understood in terms of researchers and teachers having the right *academic attitude*. This attitude principally encompasses an understanding of oneself as performing a public role in research and/or teaching, which is principally aimed at the production of knowledge. Accordingly, good academic practice to a large extent is an epistemological matter, rather than a matter of exhibiting morally correct behavior or simply following the rules of the scientific "game" that are presented in policy documents, guidelines, codes of conduct, and so on.

Notes

1 Historical overviews are given by Broad (1981) and Broad & Wade (1982, 1994). Cases that were widely discussed more recently include fraud in condensed matter physics (the case of Jan-Hendrik Schön; Reich, 2009), stem cell research (the case of Hwang-woo Suk; see Resnik, Shamoo & Krimsky, 2006) and social psychology (the case of Diederik Stapel; see Levelt Committee, Noort Committee & Drenth Committee, 2012). Among the empirical surveys that have attempted to map out the current situation are Martinson et al. (2005) and Fanelli (2009).

2 Interestingly, though, surveys showed that public trust in science has remained stable over the past four decades or so (Funk, 2017).

3 The dichotomy between explanation ("*erklären*" in German) and understanding ("*verstehen*") is often understood as a dichotomy between knowledge and something that does not count as knowledge. I do not have space here to engage with the dichotomy and the contemporary discussion on understanding as an aim of science next to explanation (see, for example, De Regt & Dieks, 2005; De Regt, 2015, 2017, and De Regt and Baumberger's chapter in the present volume), or with the extensive discussion on what exactly scientific explanations consist in (for overviews, see Skow, 2016; Woodward, 2017).

4 Note that it has been argued that history *does* provide explanations, albeit of a different kind than the sciences (Dray, 1957, 1964, 1968). More recently, a similar discussion has emerged on the question of whether the engineering disciplines produce explanations and knowledge (Vincenti, 1990; Pitt, 2000, p. 41–65).

5 I invite those readers who might think that knowledge in the humanities does not have societal applications that could beneficially or adversely affect societies and the people therein to consider how philosophical, political and economic systems of thought, such as Marxism and economic liberalism, have shaped and continue to shape our societies.

6 Traditionally, scientists are thought to have the obligation to make all their findings public and not attempting to publish results is often counted as scientific misconduct (Smith, 2000). Recently, however, Bacon's implicit point that some results might better be left unpublished re-emerged in the discussion on "dual use" research, that is, research with possible military uses or risks of abuse for criminal or terroristic purposes. For introductions to this issue, see Tucker (1994), Miller and Selgelid (2007), Selgelid (2007, 2010), or Rappert and Selgelid (2013).

7 I have discussed the case of plagiarism elsewhere (Reydon, 2015).
8 See also ESF/ALLEA (2011: 6) or DFG (2013: 3). Smith (2000) listed 15 categories, ranging from serious to minor research misconduct and with "FFP" at the serious end of the spectrum.
9 The distinction is not usually explicitly found in introductory texts on good scientific practice (but for a brief discussion, see Fuchs et al., 2010, p. 43).
10 An additional, more indirect epistemic effect of plagiarism is that the occurrence of cases of plagiarism will adversely affect the public's trust in science and its products.
11 It should be noted that the latter was not an instance of *de novo* creation of life. Researchers used an existing *Mycoplasma* bacterium and substituted its genome with a fully synthetically generated genome. Also, the synthetic genome was copied from the natural genome (Newman, 2012, p. 14).

References

Bacon, F. (1906) *Bacon's Advancement of Learning and the New Atlantis, With a Preface by Thomas Case*, London, New York & Toronto: Oxford University Press.
Bridgman, P.W. (1947) "Scientists and social responsibility," *Scientific Monthly* 65: 148–154.
Broad, W.J. (1981) "Fraud and the structure of science," *Science* 212: 137–141.
Broad, W.J. & Wade, N. (1982) *Betrayers of the Truth: Fraud and Deceit in the Halls of Science*, New York: Simon and Schuster.
Broad, W.J. & Wade, N. (1994) "Fraud and the structure of science," in: Erwin, E., Gendin, S. & Kleiman, L. (Eds), *Ethical Issues in Scientific Research: An Anthology*, New York & London: Garland Publishing, pp. 69–89.
De Regt, H.W. (2015) "Scientific understanding: Truth or dare?," *Synthese* 192: 3781–3797.
De Regt, H.W. (2017) *Understanding Scientific Understanding*, New York: Oxford University Press.
De Regt, H.W. & Dieks, D. (2005) "A contextual approach to scientific understanding," *Synthese* 144: 137–170.
DFG (2013) *Proposals for Safeguarding Good Scientific Practice (Memorandum): Recommendations of the Commission on Professional Self-Regulation in Science*, Weinheim: Wiley-VCH.
Dray, W.H. (1957) *Laws and Explanation in History*, London: Oxford University Press.
Dray, W.H. (1964) *Philosophy of History*, Englewood Cliffs, NJ: Prentice-Hall.
Dray, W.H. (1968) "Explaining what" in history," in: Brodbeck, M. (Hrsg), *Readings in the Philosophy of the Social Sciences*, London: Macmillan, pp. 343–348.
ESF (2008) *Stewards of Integrity: Institutional Approaches to Promote and Safeguard Good Research Practice in Europe*, Strasbourg: European Science Foundation.
ESF/ALLEA (2011) *The European Code of Conduct for Research Integrity*, Strasbourg: European Science Foundation & Amsterdam: ALLEA.
Fanelli, D. (2009) "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data," *PLoS ONE* 4: e5738.
Fuchs, M., Heinemann, T., Heinrichs, B. et al. (2010) *Forschungsethik: Eine Einführung*, Stuttgart & Weimar: Verlag J.B. Metzler.
Funk, C. (2017) "Mixed messages about public trust in science," *Issues in Science and Technology* 34, no. 1 (Fall 2017).
Gibson, D.G., Glass, J.I., Lartigue, C. et al. (2011) "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science* 329: 52–56.

Kendig, C.E. (2014) "Synthetic biology and biofuels," in: Thompson, P.B. & Kaplan, D.M. (Eds), *Encyclopedia of Food and Agricultural Ethics*, Dordrecht: Springer, pp. 1695–1703.

Kendig, C.E. (2016) "What is *proof of concept* research and how does it generate epistemic and ethical categories for future scientific practice?," *Science and Engineering Ethics* 22: 735–753.

Levelt Committee, Noort Committee & Drenth Committee (2012) *Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel*, Tilburg: Tilburg University.

Maher, B. (2010) "Sabotage!," *Nature* 467: 516–518.

Martinson, B.C., Anderson, M.S. & de Vries, R. (2005) "Scientists behaving badly," *Nature* 435: 737–738.

Miller, S. & Selgelid, M.J. (2007) "Ethical and philosophical consideration of the dual-use dilemma in the biological sciences," *Science and Engineering Ethics* 13: 523–580.

Newman, S.A. (2012) "Synthetic biology: Life as app store," *Capitalism Nature Socialism* 23: 6–18.

Pitt, J.C. (2000) *Thinking About Technology: Foundations of the Philosophy of Technology*, New York & London: Seven Bridges Press.

Rappert, B. & Selgelid, M.J. (Eds) (2013) *On the Dual Uses of Science and Ethics: Principles, Practices, and Prospects*, Canberra: Australian National University E Press.

Rees, M. & Browne, D. (2010) "Science's nuclear responsibility," *The Guardian*, 4 April 2010, www.theguardian.com/commentisfree/2010/apr/04/scientific-responsibility-nuclear-weapons.

Reich, E.S. (2009) *Plastic Fantastic: How the Biggest Fraud in Physics Shook the Scientific World*, New York: Palgrave Macmillan.

Resnik, D.B., Shamoo, A.E. & Krimsky, S. (2006) "Fraudulent human embryonic stem cell research in South Korea: Lessons learned," *Accountability in Research* 13: 101–109.

Reydon, T.A.C. (2015) "Plagiate als Professionalisierungsproblem," in: Lahusen, C. & Markschies, C. (Eds), *Zitat, Paraphrase, Plagiat: Wissenschaft zwischen guter Praxis und Fehlverhalten*, Frankfurt & New York: Campus, pp. 293–304.

SAMS (2008) *Integrity in Scientific Research: Principles and Procedures*, Bern: Swiss Academies of Arts and Sciences.

Selgelid, M.J. (2007) "A tale of two studies: Ethics, bioterrorism, and the censorship of science," *Hastings Center Report* 37, no. 3: 35–43.

Selgelid, M.J. (2010) "Ethics engagements of the dual-use dilemma: Progress and potential," in: Rappert, B. (Ed.), *Education and Ethics in the Life Sciences: Strengthening the Prohibition of Biological Weapons*, Canberra: Australian National University E Press pp. 23–34.

Skow, B. (2016) "Scientific explanation," in: Humphreys, P. (Ed.), *The Oxford Handbook of Philosophy of Science*, New York: Oxford University Press, pp. 524–543.

Smith, R. (2000) "What is research misconduct?," *Proceedings of the Royal College of Physicians of Edinburgh* 30 (Supplement 7), 3–8.

Steneck, N.H. (2007) *ORI Introduction to the Responsible Conduct of Research (Revised Edition)*, Washington, DC: U.S. Department of Health and Human Services.

Tucker, J.B. (1994) "Dilemmas of a dual-use technology: Toxins in medicine and warfare," *Politics and the Life Sciences* 13: 51–62.

UKRIO (2009) *Code of Practice for Research: Promoting Good Practice and Preventing Misconduct (September 2009)*, London: UK Research Integrity Office.

Vincenti, W.G. (1990) *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History*, Baltimore, MD: Johns Hopkins University Press.

VSNU (2014) *The Netherlands Code of Conduct for Academic Practice: Principles of Good Academic Teaching and Research (Revised 2014)*, The Hague: Association of Universities in the Netherlands.

Wolpert, L. (1989) "The social responsibility of scientists: Moonshine and morals," *British Medical Journal* 298: 941–943.

Wolpert, L. (2005) "Is science dangerous?," *Philosophical Transactions of the Royal Society B* 360: 1253–1258.

Woodward, J. (2017) "Scientific explanation," in: Zalta, E.N. (Ed), *The Stanford Encyclopedia of Philosophy (Fall 2017 Edition)*, https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/.

3

HOW DO MEDICAL RESEARCHERS MAKE CAUSAL INFERENCES?

Olaf Dammann, Ted Poston, and Paul Thagard

Introduction

Bradford Hill asked "In what circumstances can we pass from … [an] observed *association* to a verdict of *causation*? Upon what basis should we proceed to do so?" (Hill 1965, p. 295) Hill's expertise lay in the relationship between work conditions and illness. He often had information that revealed associations among many factors, and he had to determine which factors, if any, cause which others. He aimed to provide guidelines (what he called "viewpoints") for justifying a particular causal inference.

The Hill aspects are widely discussed and used in epidemiological inference, yet how they justify causal inference is poorly understood. Morabia (2013, p. 1526) remarked that "Hill's viewpoints may be philosophically novel, sui generis, still waiting to be validated and justified."

We advance Hill's contribution by interpreting his viewpoints as contributions to inference to the best explanation. We first introduce the Hill aspects, and then discuss explanatory coherentism based on the principles of explanatory coherence. We then apply these principles to three cases of epidemiological inference using the ECHO model of computing explanatory coherence: the recent case of inferring a causal relationship between the Zika virus and birth defects, the classic case of inferring that smoking causes cancer, and the historical case of Snow's inference to the cause of cholera. Each case illustrates the central coherentist theme that justified inferences require balancing various lines of evidence with various competing theoretical claims. Moreover, the cases illustrate the utility of the ECHO program for modeling epidemiological inference. Finally, we provide a general interpretation of Hill's aspects in terms of principles of explanatory coherence and reply to objections to our approach.

Hill's Viewpoints

Epidemiological inference is complex. It is rarely obvious what statistically correlated factors are causally responsible for others. It is typical for multiple possible causes to be viable explanations. Partial evidence is usually misleading. Some of the identified associations normally conflict with expected theory. In such a complex evidential situation, it is difficult to justifiedly infer any causal relationship. Even so, the need to improve public health makes it imperative to discern causal relationships.

Hill's viewpoints address this complex situation and provide particular questions that a medical researcher should attempt to answer. Reordering his list, we group his nine aspects as follows:

1. Temporality – does the putative cause precede the effect?
2. Strength of association – is the association strong?
3. Consistency of association – is the association consistent across a variety of conditions?
4. Specificity of association – how specific is the association?
5. Biological gradient – is there a strong dose-response curve (i.e., the curve of independent and dependent variables)?
6. Experiment – is the association supported by experimental study?
7. Plausibility – how plausible is the causal claim given existing biological knowledge?
8. Coherence – does the causal claim cohere with the existing history and biology of the disease?
9. Analogy – how similar is the potential causal claim with other accepted causal claims?

The first aspect, *temporality*, suggests that one should determine the beginning point of each factor and then formulate causal hypotheses guided by the rule that causes come before their effects. Hill observed that the onset of certain factors is not always evident. Illnesses often have a long incubation period, and an illness may cause a particular factor rather than vice versa. For example, Hill asked: "Does a particular diet lead to disease or do the early stages of the disease lead to those peculiar dietetic habits?" (Hill, 1965, p. 297).

The features of association are the *strength of association,* the *consistency of association,* the *specificity of association,* the *biological gradient,* and *experiment*. The strength of association between a possible causal condition C and an effect E should examine the ratios of (i) C&E to C&~E and of (ii) C&E to ~C&E. The first ratio compares the number of cases in which the putative cause and effect are present to the cases in which the putative cause but not the effect is present. The second ratio compares the number of cases in which the putative cause and effect are present to the number of cases in which the effect is present without the putative cause. Causal relations are consistent with a low ratio (i). For instance, smoking causes

lung cancer even though few smokers develop lung cancer. The key to detecting this causal relation is that lung cancer is rare in non-smokers so that there is a strong ratio (ii) of smoking and cancer to not-smoking and cancer.

The *consistency* of an association concerns whether it has been observed by different persons in different places at different times. This aspect is aimed against alternative explanations of an association such as chance and bias. Similarly, *experiment* looks for cases where removing a possible cause decreases an effect, also making less plausible alternative explanations such as chance and confounding factors. Consistency looks at existing studies in diverse circumstances, whereas experiment looks at interventional studies.

The *specificity* of association favors more precise causal paths over more general ones. Workers at several chemical plants may develop an illness, suggesting that working at chemical plants causes illness. But the suggestion is stronger if the association is limited to specific workers, sites, and diseases, and when there is no association between the work and other diseases.

The last aspect of association that Hill mentioned is *biological gradient*, which corresponds to John Stuart Mill's (1970; original 1843) method of concomitant variation. More of a cause is associated with more of an effect, and less of a cause produces less of an effect. Evidence for the causal connection between smoking and lung cancer is enhanced by the fact that the number of cigarettes smoked per day is proportional to the rate of lung cancer.

The guiding aspect of temporality together with the five aspects of association are alone inadequate to infer a causal relationship. Causal inference should also be guided by theory, captured by Hill's aspects of plausibility, coherence, and analogy. *Plausibility* assesses how the potential causal relationship fits with general biological knowledge. *Coherence* assesses how the potential causal relationship fits with the history and biology of the disease. Finally, *analogy* assesses whether the potential causal relationship is similar to other established causal relationships.

The importance of background theory in causal inference is illustrated by the history of the practice of bloodletting. Based on the theory that disease involved an imbalance of the four humors (blood, black bile, yellow bile, and phlegm), bloodletting evacuated 'bad blood' from the body to restore the proper balance of the humors. This practice was supported both by the association between bleeding patients and fever reduction, and by the theory of disease as humoral imbalance. The germ theory of disease introduced by Pasteur dramatically changed the biological background and led to the abandonment of bloodletting. A strength of Hill's perspective is his sensitivity to the theoretical dynamics in causal inference.

Explanatory Coherence

The complexity of epidemiological inference suggests a coherentist interpretation. Evidence for a claim *emerges* from a body of information in which the relations of support between claims are bi-directional and may involve rejecting some of

the originally accepted claims. On a coherentist picture of inference each claim in a body of information may contribute to the justification of any other (see, for example, Poston, 2014, ch. 3).

Medical researchers highlight the emergence of conclusions from evidence. Rasmussen et al. note the emergence of a causal relation in the case of the Zika virus. They write,

> As is typically the case in epidemiology and medicine, no 'smoking gun' (a single definitive piece of evidence that confirms Zika virus as a cause of congenital defects) should have been anticipated. Instead, the determination of a causal relationship would be expected to emerge from various lines of evidence, each of which suggests, but does not on its own prove, that prenatal Zika virus infection can cause adverse outcomes. *(2016, p. 1982)*

Dammann (2018) proposed that epidemiological inferences concerning the causes of disease can be understood in terms of explanatory coherence through Poston's (2014) development of explanatory coherentism. Furthermore, Dammann conjectured that Thagard's (1989) ECHO model of coherence computation could provide a rigorous account of such inferences. We now develop Dammann's proposal both specifically and generally. We show how Thagard's principles of explanatory coherence apply to three important cases of epidemiological reasoning, all of which can be simulated using ECHO. We then describe more generally how these principles connect with Hill's viewpoints and similar attempts to characterize inference in epidemiology. Our results confirm and deepen the remark of Broadbent (2017, p. 104) that Hill's reasoning is a good example of inference to the best explanation.

Philosophers such as Wilfred Sellars (1973), Gilbert Harman (1973), and Ted Poston (2014) have argued that knowledge is justified by explanatory coherence: you are justified in believing that P if P is part of the best explanation of the evidence as determined by coherence with everything that you know. Thagard (1989) proposed a precise theory of explanatory coherence accompanied by a computational model, ECHO, which has been used to simulate numerous examples of scientific, medical, legal, and everyday inference (Thagard, 1992, 1999, 2000, 2004, 2012; Eliasmith and Thagard, 1997; Nowak and Thagard, 1992a, 1992b).

Box 3.1 presents principles of explanatory coherence. In the Zika case, the data (Principle E4) are the results of observations, for instance the Brazilian finding of a strong association between Zika virus infection and cases of microcephaly. The hypotheses are conjectures about what might be causing the data, for example that Zika virus causes microcephaly and other birth defects. Principle E2 says that hypotheses cohere with what they explain, so the hypothesis that Zika virus causes birth defects coheres with the evidence concerning increased microcephaly

in Brazil. Hypotheses can be stacked up in complex causal networks, for example Zika virus causes birth defects because of biological mechanisms of infection disrupting cell growth. In accord with Principle E1, the coherence relation is symmetrical: hypothesis and data cohere with each other. In contrast, the probability of a hypothesis given data is usually very different from the probability of data given evidence.

BOX 3.1 PRINCIPLES OF EXPLANATORY COHERENCE, FROM THAGARD (2006).

Principle E1. Symmetry. Explanatory coherence is a symmetric relation, unlike, say, conditional probability. That is, two propositions *p* and *q* cohere with each other equally.

Principle E2. Explanation. (a) A hypothesis coheres with what it explains, which can either be evidence or another hypothesis; (b) hypotheses that together explain some other proposition cohere with each other; and (c) the more hypotheses it takes to explain something, the lower the degree of coherence.

Principle E3. Analogy. Similar hypotheses that explain similar pieces of evidence cohere.

Principle E4. Data priority. Propositions that describe the results of observations have a degree of acceptability on their own.

Principle E5. Contradiction. Contradictory propositions are incoherent with each other.

Principle E6. Competition. If *P* and *Q* both explain a proposition, and if *P* and *Q* are not explanatorily connected, then *P* and *Q* are incoherent with each other. (*P* and *Q* are explanatorily connected if one explains the other or if together they explain something.)

Principle E7. Acceptance. The acceptability of a proposition in a system of propositions depends on its coherence with them.

Principle E3 recognizes that analogy can contribute to coherence, for example when Darwin (1859) argued that one of the supports for his theory of evolution by natural selection was the analogy with artificial selection carried out by breeders (Thagard 1978). In the Zika case, epidemiologists note analogous explanations such as the causation of birth defects by the rubella virus.

Principles E5 and E6 establish incoherence relations between hypotheses that are flat-out contradictory or merely competing to explain the same data. The alternatives to the hypothesis that the Zika virus causes birth defects are that something else causes birth defects, or that the defects occur randomly.

Principles E1–E6 establish complex networks of data, explanations, and competing hypotheses at different levels. Principle E7 directs how to determine what to believe and what not to believe, based on how well a proposition (hypothesis or piece of evidence) fits with everything else. For example, the hypothesis that Zika virus causes birth defects should fit with all the data and outcompete alternative hypotheses. In a complex evidential situation, it is difficult to determine the best fit of all the explanatory constraints. The computer program ECHO shows how to best satisfy these constraints.

ECHO Simulations

To determine overall coherence, the computer program ECHO uses a neural network algorithm for approximately maximizing coherence. ECHO represents each proposition by a unit, a simplified artificial neuron that is connected to other units by excitatory and inhibitory links. As in real neurons, an excitatory link is one that enables one neuron to increase the firing of another, whereas an inhibitory link decreases firing. After cycles of excitation and inhibition, the firing rates (activations) of the units settle into stable patterns.

Zika Simulation

In the Zika example, we can represent the hypothesis that the virus causes defects by the unit ZIKA-CAUSES-DEFECTS, and the Brazilian association between virus and defects by a unit BRAZIL-ASSOCIATION. Then whenever principles E2 and E3 establish relations of coherence between two propositions, the units that represent the propositions get excitatory links between them. So ZIKA-CAUSES-DEFECTS and BRAZIL-ASSOCIATION have an excitatory link between them that is symmetric in accord with principle E1. Principle E4 is implemented by making an excitatory link between a special unit EVIDENCE and any unit such as BRAZIL-ASSOCIATION that represents a proposition based on observation. Principles E5 and E6, which establish incoherence between competing hypotheses, are implemented by inhibitory links between units: when two hypotheses are incoherent, for example, ZIKA-CAUSES-DEFECTS versus OTHER-CAUSE, then the units that represent them get an inhibitory link between them.

The acceptability of a unit is represented by its activation, corresponding roughly to the firing rate of a real neuron. Just as firing rates of neurons are determined by their excitatory and inhibitory neurons, the activation of units in ECHO are determined by their excitation and inhibition and the activation of the units to which they are connected. When the network settles (i.e., activations stabilize), the resulting activations (positive or negative) indicate whether the hypotheses and data represented by the units are accepted or rejected. A test of the theory of explanatory coherence is whether examples such as the Zika virus can be plausibly modeled using the program ECHO.

The neural networks used by ECHO are not biologically plausible because single neuron-like units represent complex propositions such as that Zika virus causes birth defects, and because the excitatory and inhibitory links between units are symmetric. Thagard and Aubie (2008) showed how to translate ECHO networks into more biologically realistic networks with one-directional links between neurons in groups that collectively represent propositions. Techniques are now available for translating complex symbolic propositions into neural networks (Eliasmith and Thagard, 2001).

The input to ECHO for the Zika virus simulation consists of statements of what explains what, analogies, and evidence, shown in Box 3.2. Fleshed out, the main evidence and hypotheses are:

E1. Infection is present during prenatal development.
E2. Rare microcephaly is associated with Zika. Reports of fetuses and infants with microcephaly who are born to women with brief periods of travel to countries with active Zika virus transmission are consistent with Zika virus being a rare exposure. The defect, congenital microcephaly, is rare, with a birth prevalence of approximately 6 cases per 10,000 liveborn infants, according to data from birth-defects surveillance systems in the United States.
E3. Zika virus is in brain tissue.
E4. A study during the outbreak in Brazil found a significant association between Zika virus infection and microcephaly. Eighty-eight pregnant women who had had an onset of rash in the previous 5 days were tested for Zika virus RNA. Among the 72 women who had positive tests, 42 underwent prenatal ultrasonography, and fetal abnormalities were observed in 12 (29%); none of the 16 women with negative tests had fetal abnormalities. The abnormalities that were observed on ultrasonography varied widely, and some findings lacked postnatal confirmation because the pregnancies were ongoing.
E5. A study on subjects in French Polynesia found a significant association between Zika virus infection and microcephaly.
E6. No results of an animal model with Zika virus infection during pregnancy and fetal effects have yet been published.
E7. Birth defects are associated with rubella virus.
E8. Animal models have shown that Zika virus is neurotropic, supporting biologic plausibility.
E9. Zika virus infects neural progenitor cells and produces cell death and abnormal growth.
H1. Zika virus causes microcephaly.
H2. There is some other cause of microcephaly.
H3. Rubella causes birth defects.
H4. Zika virus is neurotropic.

BOX 3.2 INPUT TO ECHO FOR ZIKA SIMULATION. THE PARENTHESES AND QUOTATION MARKS ARE ARTIFACTS OF THE IMPLEMENTATION OF ECHO IN THE PROGRAMMING LANGUAGE LISP.

```
; EVIDENCE
(proposition 'E1 "infection-during-prenatal-development")
(proposition 'E2 "rare-microencephaly-with-zika")
(proposition 'E3 "zika-virus-in-brain-tissue")
(proposition 'E4 "Brazil-more-microencaphaly-after-infection")
(proposition 'E5 "Polynesia-more-microencaphaly-after-infection")
(proposition 'E6 "no-animal-models")
(proposition 'E7 "birth-defects-rubella")
(proposition 'E8 "animal-models-neurotropic")
(proposition 'E9 "Zika-produces-abnormal-growth")
(data '(E1 E2 E3 E4 E5 E6 E7 E8 E9))
; HYPOTHESES
(proposition 'H1 "zika-virus-causes-microencephaly")
(proposition 'H2 "other-cause")
(proposition 'H3 "rubella-causes-defects")
(proposition 'H4 "Zika-virus-is-neurotropic")
; EXPLANATIONS
(explain '(H1) 'E2)
(explain '(H1 E1 E3) 'E4)
(explain '(H1 E1 E3) 'E5)
(explain '(H2) 'E6)
(explain '(H3) 'E7)
(explain '(H4 E9) 'H1)
; ANALOGY
(analogous '(H1 H3) '(E4 E7))
; CONTRADICTION
(contradict 'H1 'H2)
```

Figure 3.1 provides a simplified picture of the causal network that ECHO turns into a neural network. When ECHO is run, all units begin with activation 0, and after 118 cycles of activation adjustment activations stabilize. ECHO accepts the hypothesis that the Zika virus causes birth defects while rejecting the alternative hypothesis of some other cause. The specific numbers for activation are not significant: what matters is whether the final activation is above 0, indicating acceptance, or below 0, indicating rejection. Hence explanatory coherence and the ECHO model explain how the conclusion that the Zika virus causes brain defects arises by inference to the best explanation.

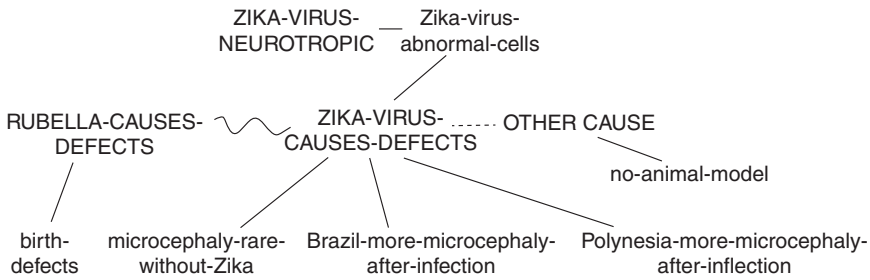

FIGURE 3.1 Causal and neural network for the Zika simulation, simplified. Solid lines indicate excitatory links based on coherence from explanation. The wavy line indicates an excitatory link based on coherence from analogy. The dotted line indicates an inhibitory link based on incoherence from contradiction. Hypotheses are shown in capital letters, and evidence in lower case.

Smoking/Cancer Simulations

One of the great public health accomplishments of epidemiology is the demonstration that tobacco smoking causes cancer and other diseases. Hill was one of the earlier researchers to find a statistical association between smoking and cancer (Doll and Hill, 1950), but the overall case that smoking causes cancer was made by the American Surgeon General (1964). This study used five "criteria" for establishing causal relationships based on statistical associations: consistency, strength, specificity, temporal relationship, and coherence. All of these are included in Hill's viewpoints, and the four viewpoints not included in the report (biological gradient, experiment, plausibility, analogy) might be absorbed into coherence. Subsequent reports to the Surgeon General (e.g., 2010, 2014) made an even stronger case that smoking causes many diseases, including various forms of cancer, cardiovascular and pulmonary problems, and reproductive effects.

Proctor (2012) reviewed the history of the discovery of the connection between cigarettes and lung cancer. He says that four lines of evidence converged to establish cigarette smoking as the leading cause of lung cancer: population studies, animal experimentation, cellular pathology, and cancer-causing chemicals in cigarette smoke. Population studies repeatedly found that smokers of cigarettes were far more likely to contract lung cancer than non-smokers. Animal experiments found that applying tobacco products to rabbits and mice led to cancer. Cellular pathology research showed that smokers experienced damage to lung cells. Finally, chemical research determined that cigarette smoke contains many carcinogens.

How these lines of evidence converged to back the conclusion that smoking causes cancer is a matter of explanatory coherence, as shown in Figure 3.2. The hypothesis that smoking causes cancer explains why smokers and tobacco-applied animals are more likely to get cancer. The studies about cellular

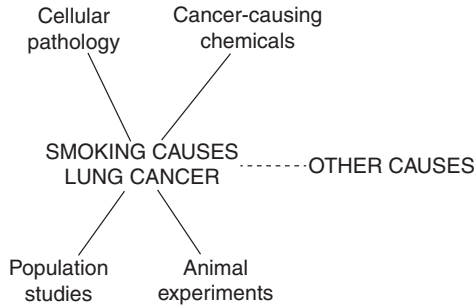

FIGURE 3.2 Explanatory coherence of the conclusion that smoking causes cancer, following Proctor (2013). Conventions are the same as in Figure 3.1.

pathology and cancer-causing chemicals sketch the mechanisms which explain how smoking causes cancer, through the effects of carcinogenic chemicals on lung cells.

The input for the ECHO simulation of this case is shown in Box 3.3. After 95 cycles of activation adjustment, the neural network produced by this input settles, with positive activation of the unit smoking-causes-cancer indicating acceptance of this hypothesis, rejecting other-causes.

BOX 3.3 INPUT TO ECHO FOR PROCTOR SIMULATION.

```
; EVIDENCE
(proposition 'population "population studies associate smoking and lung
    cancer")
(proposition 'animal "animal experimentation associate tobacco and cancer")
(proposition 'cellular "cellular pathology finds that smoking damages cells")
(proposition 'chemicals "there are cancer causing chemicals in smoke")
(data '(population animal cellular chemicals))
; HYPOTHESES
(proposition 'smoking-causes-cancer "tobacco smoking causes cancer")
(proposition 'other-cause "cancer has other causes")
; EXPLANATIONS
(explain '(smoking-causes-cancer) 'population)
(explain '(smoking-causes-cancer) 'animal)
(explain '(chemicals cellular) 'smoking-causes-cancer)
; CONTRADICTION
(contradict 'smoking-causes-cancer 'other-cause)
```

John Snow's Communication Theory of Cholera

John Snow (1855) is considered one of the originators of epidemiology because of his arguments in the 1840s and 1850s that cholera is caused by communication via excremental evacuations. Tulodziecki (2011) has provided a thorough analysis of Snow's arguments showing the explanatory power of his theory compared to the prevalent view that cholera results from miasma – bad air due to decaying matter.

This analysis translates into explanatory coherence as shown in Figure 3.3. The superior explanatory power of the communication theory comes primarily from its ability to explain numerous phenomena that the miasma theory cannot. For example, communication of "cholera poison" via evacuation explains why cholera usually starts with digestive problems and why people with bad hygiene got cholera more often than people with good hygiene. In addition, the communication theory explains why physicians (who were careful about washing hands and not eating while visiting the sick) were less likely to get cholera than ordinary people. In contrast, on the miasma theory physicians would be more likely to get the disease via miasmic effluvia from the sick people they visited (Tulodziecki, 2011, p. 312). Figure 3.3 displays the superior explanatory power of the communication theory.

The relations between propositions shown in Figure 3.3 generate the input to ECHO shown in Box 3.4. In less than a second, with 142 cycles of activation updating, ECHO settles with the acceptance of H2 (communication causes cholera) and the rejection of H1 (miasma causes cholera).

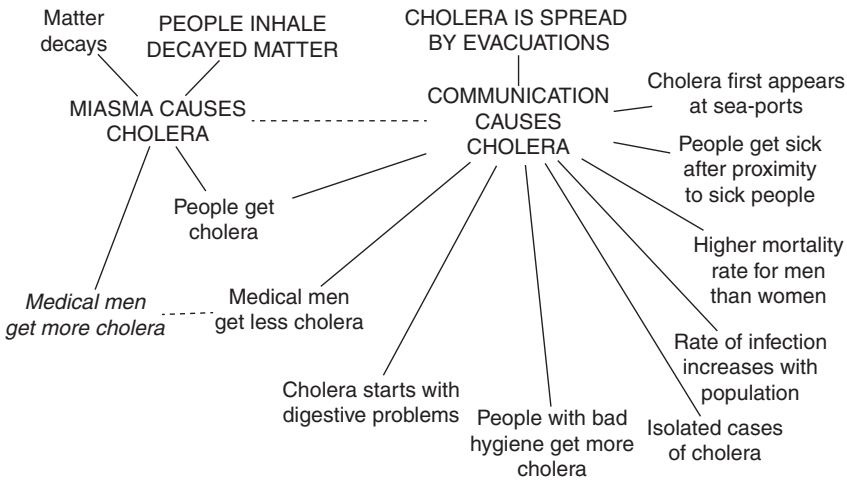

FIGURE 3.3 Explanatory coherence of the conclusion that smoking causes cancer, following Tulodziecki (2011). Conventions are the same as in Figure 3.1.

BOX 3.4 INPUT TO ECHO FOR SNOW SIMULATION

```
; EVIDENCE
(proposition 'E1 "people get cholera")
(proposition 'E2 "medical men get less cholera")
(proposition 'NE2 "medical men get more cholera"); negative evidence
(proposition 'E3 "cholera starts with digestive problems")
(proposition 'E4 "people with bad hygiene get more cholera")
(proposition 'E5 "there are isolated cases of cholera")
(proposition 'E6 "the rate of infection increases with population")
(proposition 'E7 "higher mortality rater for men than women")
(proposition 'E8 "people get sick after proximity to sick people")
(proposition 'E9 "cholera first appears at sea-ports")
(proposition 'E10 "matter decays")
(data '(E1 E2 E3 E4 E5 E6 E7 E8 E9 E10))
; HYPOTHESES
(proposition 'H1 "miasma causes cholera")
(proposition 'H2 "communication causes cholera")
(proposition 'H3 "people inhale decayed matter")
(proposition 'H4 "cholera is spread by evacuations");
CONTRADICTIONS
(contradict 'E2 'NE2)
; EXPLANATIONS
(explain '(H1) 'E1)
(explain '(H1) 'NE2)
(explain '(H3 E10) 'H1)
(explain '(H2) 'E1)
(explain '(H2) 'E2)
(explain '(H2) 'E3)
(explain '(H2) 'E4)
(explain '(H2) 'E5)
(explain '(H2) 'E6)
(explain '(H2) 'E7)
(explain '(H2) 'E8)
(explain '(H2) 'E9)
(explain '(H4) 'H2)
```

Connections with Epidemiological Standards for Causality

Thagard (1998, 1999) described how an explanatory coherence account of reasoning concerning the causation of stomach ulcers by *H. pylori* bacteria fit well

with the criteria of causality advocated by Evans (1993). More generally, Table 3.1 maps the relation between principles of explanatory coherence and additional ways that epidemiologists have characterized determination of causality, due to Hill (1965) and Shepard et al. (1994). Rasmussen et al. (2016) connected their inference that Zika virus causes birth defects with Shepard's criteria, which map onto Hill's viewpoints and Evans's criteria as shown by Table 3.1. These viewpoints and criteria are not necessary and sufficient conditions for causality but serve as standards of evaluation.

According to principle E2, the hypothesis that an environmental condition causes a disease coheres with the evidence that it explains. But if the disease

TABLE 3.1 Mapping of standards for causation onto explanatory coherence principles

Hill's viewpoints	*Evans's criteria*	*Shepard's criteria*	*Principles of explanatory coherence*
1. Temporality	4. Temporally, disease follows exposure	1. Exposure to the agent	E2. Explanation
2. Strength of association	1. Prevalence 2. Exposure 3. Incidence	2. Epidemiology findings	E2. Explanation E4. Data priority
3. Consistency of association		2. Consistent findings 3. Delineation of cases	E4. Data priority E5. Contradiction E7. Acceptance
4. Specificity of association		4. Rarity of exposure and defect	E2. Explanation E5. Contradiction
5. Biological gradient	5. Spectrum of host responses 6. Measurable host response		E2. Explanation
6. Experiment	7. Experimental reproduction 8. Elimination -> reduction 9. Prevention of host response	5. Experimental animals 7. Experimental system for agent	E2. Explanation
7. Plausibility	10. Biological sense	6. Biological sense	E2. Explanation (higher order) E3. Analogy
8. Coherence	10. Biological sense		E5. Contradiction E7. Acceptance
9. Analogy		6. Biological sense	E3. Analogy

precedes the condition, then there is no causation, hence no explanation, and hence no coherence. The rule that causes come before effects is not true *a priori*, because it is conceivable that time travel could enable a future event such as getting into a time machine in the year 3000 to cause an earlier event such as arriving in a place in the year 1000. But there has never been an observed case of the future causing the past, so the temporality rule is a reasonable way of dismissing cases of backward causation. Thus, failure of temporality blocks some applications of E2 where an explanatory cause happens after the event explained.

Principle E2 is also key to understanding Hill's aspects 2–6. The hypothesis that an environmental condition causes a disease can explain why there are associations between the condition and the disease that are strong, consistent, and specific. These explanations therefore enhance the coherence and acceptability of the causal hypothesis. Alternative hypotheses such as chance and the occurrence of some unknown factor cannot furnish comparable explanations.

Similarly, the hypothesis that a condition causes a disease explains why there is a biological gradient such that more of the condition causes more disease. When experimental evidence of a successful intervention is available, it also increases explanatory coherence because the hypothesis that a condition causes a disease explains why changing the condition changes the disease. Alternative hypotheses concerning chance and unknown factors cannot mount similar explanations, and hence gain no support from E2.

Principles E4, E5, E6, and E7 are also relevant to understanding why aspects 2–6 help to indicate causality. E4 (data priority) ensures that evidence collected by observations and experiments gets a degree of priority over hypotheses. E4 does not imply that data are always taken at face value, because observations and experimental results can be mistaken; but it does help to ensure that evidence will have a greater contribution to coherence than hypotheses that may be fanciful. Aspects E5 and E6 set up a battle between the hypothesis that a condition causes a disease and its alternatives, either because the alternative is flat-out contradictory (e.g., cause vs. chance) or merely competitive in cases where multiple causes might be operating.

One weakness of Hill's method is that he gave no indication of how all the aspects can be combined into an overall inference that a condition causes a disease. Principle E7 asserts that maximizing coherence is the key to evaluating a causal hypothesis and other beliefs. E7 does not say how to maximize coherence, but the construal of coherence as constraint satisfaction and the availability of various algorithms for approximately maximizing coherence (including the neural network algorithm used by ECHO) take care of this problem. From the explanatory coherence perspective, medical researchers should use Hill's aspects to establish non-rigid constraints that need to be coherently satisfied. The ECHO program then determines best overall fit.

Hill's aspects 7–8 of plausibility and coherence also fall under the theory of explanatory coherence. They urge that a causal hypothesis should fit with general

biological and medical knowledge. Principle E5 (contradiction) handles the most extreme case where a new hypothesis contradicts what is generally believed. As Hill noted, contradicting orthodoxy does not always provide the grounds for rejecting a hypothesis because the orthodoxy may be wrong.

Another source of fit with biological and medical knowledge comes from the availability of higher-order explanations. The hypothesis that smoking causes cancer became more plausible once mechanisms were understood for how the ingredients in smoke irritate tissues and encourage the development of mutations that lead to growth of tumors. Also relevant is E7 (coherence) which encourages an overall fit with all knowledge, not just the narrow domain in which the causal hypothesis operates.

Hill's ninth aspect, analogy, encourages that a causal hypothesis be analogous to other kinds of explanations used in biology and medicine. Analogy is taken care of by Principle E3 of explanatory coherence, and analogical reasoning can also be understood as a kind of parallel constraint satisfaction (Holyoak and Thagard, 1995).

Objections and Replies

Our explanatory coherence account of epidemiological reasoning generates worries.

1. Explanation. The theory of explanatory coherence is empty without an account of the nature of explanation.

Reply. Explanation follows different patterns in different fields (see Poston 2014, pp. 70–80). For example, the hypothesis that the Zika virus causes birth defects explains the evidence that in Brazil the virus is associated with microcephaly because of a partially understood mechanism where the parts are viruses and neurons, the main interaction is infection, and the regular changes are defective neurons and brains. In contrast, Snow had an explanation even though he lacked a detailed understanding of the mechanism of cholera infection, which needed the germ theory developed by Pasteur in the 1860s.

2. Causality. The hypothesis that a condition causes a disease is meaningless without an understanding of causality.

Reply: Hill acknowledged the difficulty of analyzing causality, and no definition has ever survived for long. But causality can be characterized using the method of 3-analysis, which is based on a new theory of concepts that describes how they combine exemplars (typical examples), typical features, and explanations (Blouw, Solodkin, Thagard, and Eliasmith, 2016). There are many familiar exemplars of causes, such as pushes, pulls, motions, collisions, actions, and diseases whose effects are symptoms (Thagard, 2019).

The typical features of causality include:

1. Causes happen before effects.
2. Causes operate in sensory-motor-sensory patterns, for example, when you see and feel a bike not moving, move the pedals, then see and feel the bike move.
3. Cause and effects sometimes yield regularities, for example that hitting your finger with a hammer always hurts.
4. Statistical dependencies occur, with causes increasing the probabilities of effects.
5. Manipulations and interventions lead from causes to effects.

None of these typical features is a necessary or sufficient condition of causality, but matching a lot of them suggests that cause/effect relations have been identified. There are obvious relations between these five typical features of causality and the epidemiological criteria in Table 3.1.

Such relations provide explanations of why things happen and how they can be changed. Causality in particular cases is explained by the presence of underlying mechanisms connecting cause and effect. Before concluding that C causes E, you need to consider alternative explanations such as that E has a different cause, or that C and E are both caused by something else, or that E occurs randomly. We cannot directly observe causal relations but can infer that they exist as part of the best explanation of systematic observations, in accord with explanatory coherence.

3. Inferences against causality. Epidemiology sometimes leads to the rejection of causal hypotheses, not just their acceptance.

Reply. Explanatory coherence understands the rejection of causal hypotheses as resulting from the acceptance of alternatives concerning other causes, chance, bias, or confounding. For example, the popular hypothesis that stomach ulcers are caused by excess acidity was rejected because of the explanatory coherence of the new hypothesis that bacteria cause ulcers (Thagard, 1999). More recently, the hypothesis that multiple sclerosis is caused by compromised flow of blood in veins to the head has been largely rejected for many reasons. Factors include the shoddiness of initial studies used to support the hypothesis, the conflicts of interest of the investigators who proposed it, the failure of more careful studies to find that balloon angioplasty reduces the symptoms of multiple sclerosis, and the finding that the correlation between venous insufficiency and multiple sclerosis is dubious (Traboulsee et al., 2014; Kruger, Patel, and Lee, 2015). All of these factors could be incorporated into an explanatory coherence analysis and ECHO model of rejection of the hypothesis that venous insufficiency causes multiple sclerosis.

Conclusion

Our chapter addresses causal reasoning in epidemiology, but explanatory coherence extends to other kinds of medical inference. Thagard and Larocque (2018) model mental health assessment as inference to the best explanation performed by ECHO. Other forms of diagnosis can also be construed as abductive inference, that is, inference to the best explanation (Josephson and Josephson, 1994; Peng and Reggia, 1990), in ways that naturally translate into explanatory coherence. For example, physicians who diagnose lung cancer in patients can take into account (1) evidence explained by the diagnosis such as coughing and test results, (2) history of heavy smoking which explains why the patient got sick, and (3) alternative explanations such as emphysema. Finally, reasoning in evidence-based medicine concerning the effectiveness of medical treatments can be understood as inference to the best explanation (Thagard, 2010), but detailed analysis in terms of explanatory coherence remains to be developed.

More narrowly, we have provided an epistemological interpretation and justification for Bradford Hill's influential recommendations about how to infer causality in epidemiology. Our interpretation is based on the epistemology of explanatory coherentism, fleshed out using a detailed theory of explanatory coherence. We have shown the applicability of this approach by applying the ECHO computational model for calculating explanatory coherence to three important cases of epidemiological reasoning, concerning the Zika virus, smoking, and cholera. The result is a deeper understanding of the nature of medical inference concerning the causes of disease.

Acknowledgments

Thanks to Mike Bishop, Kostas Kampourakis, Kevin McCain, and Chase Wrenn for comments on an earlier draft.

References

Blouw, P., Solodkin, E., Thagard, P., & Eliasmith, C. (2016). Concepts as semantic pointers: A framework and computational model. *Cognitive Science, 40*, 1128–1162.

Broadbent, A. (2017). Philosophy of epidemiology. In J. A. Marcum (Ed.), *The Bloomsbury companion to contemporary philosophy of medicine* (pp. 93–112). London: Bloomsbury.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. Chicago: University of Chicago Press.

Dammann, O. (2018). Hill's heuristics and explanatory coherentism in epidemiology. *American Journal of Epidemiology, 187*(1), 1–6.

Darwin, C. (1859). *The origin of species*. London: Murray.

Doll, R., & Hill, A. B. (1950). Smoking and carcinoma of the lung. *British Medical Journal, 2*(4682), 739–748.

Eliasmith, C., & Thagard, P. (1997). Waves, particles, and explanatory coherence. *British Journal for the Philosophy of Science, 48*, 1–19.

Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science, 25*, 245–286.

Evans, A. S. (1993). *Causation and disease: A chronological journey*. New York: Plenum.

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*, 295–300.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press/Bradford Books.

Josephson, J. R., & Josephson, S. G. (Eds.). (1994). *Abductive inference: Computation, philosophy, technology*. Cambridge: Cambridge University Press.

Kugler, N., Patel, P. J., & Lee, C. J. (2015). Chronic cerebrospinal venous insufficiency in multiple sclerosis: A failed concept. *Vascular Specialist International, 31*(1), 11–14.

Mill, J. S. (1970). *A system of logic* (8th ed.). London: Longman.

Morabia, A. (2013). Hume, Mill, Hill, and the sui generis epidemiologic approach to causal inference. *American Journal of Epidemiology, 178*(10), 1526–1532.

Nowak, G., & Thagard, P. (1992a). Copernicus, Ptolemy, and explanatory coherence. In R. Giere (Ed.), *Cognitive models of science* (Vol. 15, pp. 274–309). Minneapolis: University of Minnesota Press.

Nowak, G., & Thagard, P. (1992b). Newton, Descartes, and explanatory coherence. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, sognitive psychology and educational theory and practice.* (pp. 69–115). Albany, NY: SUNY Press.

Peng, Y., & Reggia, J. (1990). *Abductive inference models for diagnostic problem solving*. New York: Springer.

Poston, T. (2014). *Reason and explanation: A defense of explanatory coherentism*. London: Palgrave Macmillan.

Proctor, R. N. (2012). The history of the discovery of the cigarette–lung cancer link: Evidentiary traditions, corporate denial, global toll. *Tobacco Control, 21*(2), 87–91.

Rasmussen, S. A., Jamieson, D. J., Honein, M. A., & Petersen, L. R. (2016). Zika virus and birth defects – reviewing the evidence for causality. *New England Journal of Medicine, 2016*(374), 1981–1987.

Sellars, W. (1973). Givenness and explanatory coherence. *Journal of Philosophy, 70*, 612–624.

Shepard, T. H. (1994). "Proof" of human teratogenicity. *Teratology, 50*(2), 97–98.

Surgeon General. (1964). *Smoking and health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC.

Surgeon General. (2010). *How tobacco smoke causes disease: The biology and behavioral basis for smoking-attributable disease*. Washington, DC.

Surgeon General. (2014). *The health consequences of smoking – 50 years of progress*. Washington, DC.

Snow, J. (1855). *On the mode of communication of cholera*. London: John Churchill.

Thagard, P. (1978). The best explanation: Criteria for theory choice. *Journal of Philosophy, 75*, 76–92.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435–467.

Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.

Thagard, P. (1998). Ulcers and bacteria I: Discovery and acceptance. *Studies in history and philosophy of science. Part C. studies in history and philosophy of biological and biomedical sciences, 29*, 107–136.

Thagard, P. (1999). *How scientists explain disease*. Princeton: Princeton University Press.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.
Thagard, P. (2004). Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence, 18*, 231–249.
Thagard, P. (2006). Evaluating explanations in science, law, and everyday life. *Current Directions in Psychological Science, 15*, 141–145.
Thagard, P. (2010). *The brain and the meaning of life*. Princeton: Princeton University Press.
Thagard, P. (2012). *The cognitive science of science: Explanation, discovery, and conceptual change*. Cambridge, MA: MIT Press.
Thagard, P. (2019). *Natural philosophy: From social brains to knowledge, reality, morality, and beauty*. Oxford: Oxford University Press.
Thagard, P., & Aubie, B. (2008). Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and Cognition, 17*, 811–834.
Thagard, P., & Larocque, L. (2018). Mental health assessment: Inference, explanation, and coherence. *Journal of Evaluation in Clinical Practice*, *24,* 649–654.
Traboulsee, A. L., Knox, K. B., Machan, L., Zhao, Y., Yee, I., Rauscher, A., … Sadovnick, D. (2014). Prevalence of extracranial venous narrowing on catheter venography in people with multiple sclerosis, their siblings, and unrelated healthy controls: A blinded, case–control study. *The Lancet, 383*(9912), 138–145.
Tulodziecki, D. (2011). A case study in explanatory power: John Snow's conclusions about the pathology and transmission of cholera. *Studies in History and Philosophy of Biological and Biomedical Sciences, 42*(3), 306–316.

4

HOW DO EXPLANATIONS LEAD TO SCIENTIFIC KNOWLEDGE?*

Kevin McCain

Introduction

The famous physicist Erwin Schrödinger (1954, p. 90) – perhaps you've heard of his cat? – once said that all of science presupposes the truth of the "hypothesis that *the display of Nature can be understood*." This seems absolutely right. After all, if Nature/the world around us could not be understood (at least to a significant degree), there would be no science and no reason to engage in scientific activities. Why conduct scientific experiments if they aren't going to help you understand what you're studying? Why theorize about the nature of the world when your theories are doomed from the start? Fortunately, there's good reason to think that science's presupposition is correct – the world around us can be understood, and science is great at increasing our understanding. How does science do this though?

Science is in the business of explaining things. This is because our scientific understanding of the world comes to us by way of explanations. Science explains why sugar dissolves in water, why unsuspended objects fall, why the planets move like they do, and so on. Explanation is one of the primary aims of science (along with the prediction and control of natural phenomena). In fact, it's not unreasonable to think that explanation is *the* primary aim of science.[1] Once we understand how and why something happens we can make quite accurate predictions about when it will happen and sometimes even control whether it happens. For instance, understanding the process of combustion can help us predict when fires might occur as well as sometimes control whether they will happen at all. If, as Michael Strevens (2013) claims, understanding in science only comes by way of having an explanation, it's plausible that explanation is the primary goal of science because it is required to achieve the other two goals (prediction and control). Either way, whether explanation is the primary goal of science or simply one important goal, explanation plays a big role in science because it leads to scientific understanding.

Okay, so explanations are featured prominently in science, but what about the question of this chapter? How do explanations lead to scientific knowledge? The short answer is that they do so by way of inferences. Specifically, we can come to have knowledge by way of inferring that the best explanation of some natural phenomenon is true. Inferring to the best explanation (a method of inference sometimes referred to as "abduction") roughly has this form: we come to possess some data, consider various potential explanations of the data, and then infer that the best potential explanation is the actual explanation of the data, that is, we infer that the best explanation is true. Of course, much more needs to be *explained* about this process before we can rest assured that we understand how explanations lead to scientific knowledge. Let's turn our attention toward these details now.

What Is an Explanation?

We've presupposed that explanations can lead to scientific knowledge (after all, our question isn't *whether* explanations can lead to scientific knowledge, but *how* they do so). Up to this point we've talked a bit about what explanations can provide, and we've seen that they somehow lead to scientific knowledge by way of inference to the best explanation (IBE). However, we haven't really said what an explanation is. So, first things first, let's briefly pause to look into the nature of explanation.

To begin, we should distinguish between an *explanation* and *explaining* something. Explaining is an action that we sometimes perform. In successfully completing that action we provide an explanation. Think of when a student explains to his professor why he didn't complete his assignment on time. The student talks to the professor (or perhaps emails her) about his grandparent's illness or his unexpected computer troubles or whatever happens to be the reason why he failed to complete the assignment. When he is engaging in the activity of telling this information to his professor, the student is *explaining* his failure to turn in the assignment. He is providing an *explanation*, which is constituted by the reasons he gives for failing to complete the assignment. The explanation is a set of claims: my computer crashed right before the assignment was due, I didn't have a backup file saved, and so on. Now, whether or not the student is conveying the *actual* (true) explanation when explaining all this to the professor depends on whether he is telling her the real reason for his not completing the assignment. This is important because it helps to make clear that explanations are independent of our acts of explaining. One might engage in the act of explaining without providing the correct explanation (though we might question whether this counts as really explaining or simply offering excuses/trying to explain). Or, one might have an explanation without engaging in the act of explaining at all such as when the student simply chooses not to tell his professor why he didn't turn in the assignment. In this case, the student has an explanation (he knows why he didn't complete the assignment), but he doesn't explain his failure to complete the assignment. In this

chapter we're concerned with explanation rather than acts of explaining. It is by inferring the truth of best explanations that we come to have scientific knowledge, not necessarily by engaging in acts of explaining (though this is often how we share information and scientific knowledge with others).

In light of its central role in science it should come as no surprise that philosophers of science have been keenly interested in providing accounts of the nature of explanation. There are way too many such accounts for us to explore them all in detail, or even give much of an overview of the prominent theories here.[2] Luckily, we don't have to settle the debate concerning the proper account of explanation to answer our question. If one is worried about this, consider this fact: scientists have been using explanations to come to have scientific knowledge ever since there's been scientific knowledge. This is so despite the fact that the issue of the correct theory of explanation remains unsettled. So, for our purposes we can adopt a working model of explanation just to have something specific in mind when we discuss IBE; that will be sufficient for our current needs.

A plausible working model of explanation for our present discussion is philosopher Jaegwon Kim's (1994, p. 68) idea that "explanations track dependence relations." The idea is that an explanation consists of information about how the thing being explained (the *explanandum*) depends upon other things (the *explanans*). For example, an explanation of a window's breaking (the *explanandum*) consists of information about how the event of the window breaking is causally dependent upon other events, such as a baseball flying into it (the *explanans*). One thing that makes this "dependence" view of explanation so helpful is that it covers all sorts of relations: causal relations, constitution relations (when some things make up something else), mereological relations (relations that exist between the parts of an object), and so on. This view of explanation is consistent with all of the major views of the nature of explanation. So, it will work well for us.

We now have a solid working model of explanation in hand. An explanation is a set of claims/information about dependence relations that exist between the explanandum (what is being explained) and the explanans (what is doing the explaining). We're now ready to look at exactly how explanations lead to scientific knowledge; we're ready to dive into IBE.

How Do Explanations Generate Scientific Knowledge?

There is an obvious sense in which explanations generate scientific knowledge. When you know that a particular scientific explanation is correct you can use it to come to know what happened in a particular situation. For example, when you know the physical explanation for a solid dissolving in a liquid, you can use that explanation to generate knowledge about what happened when this particular bit of sugar was placed in that particular cup of water. You understand how a solid's holding together depends upon attractive forces between its molecules, and how the binding of those molecules with the liquid's molecules depends upon

their different polarizations, and so on. Your grasping this correct explanation of dissolving allows you to determine why the particular bit of sugar dissolved. Relatedly, you can use knowledge of an explanation to know what is likely to happen – this is simply using the explanation to make predictions. Given your knowledge of scientific explanations you can predict what will happen if someone were to drop some sugar into water. While both of these (knowing what happened in a particular circumstance and predicting what will or would happen in other circumstances) are important ways that explanations lead to scientific knowledge, they aren't our primary concern here. In other words, we are not concerned with applying explanations to particular instances or with drawing on explanations to make predictions. Instead, we are concerned with how we come to know the relevant scientific explanations are correct in the first place. Such explanations are themselves the heart of scientific theories, and thus the heart of what we typically think of as scientific knowledge.

What Is Inference to the Best Explanation?

We come to knowledge of correct explanations via a specific kind of inference, inference to the best explanation (IBE). IBE is something that you are familiar with even if you haven't thought of it in a formal manner. We use IBE all the time in our everyday lives – so much so that it isn't far-fetched to say that we employ it in such an automatic way that we often don't even notice we are doing it. Here's just a few examples.

(a) You come home ready to eat the noodles that you saw in the refrigerator this morning, but they're gone. You only have one roommate, and he has been home all day. You infer that he ate the noodles.
(b) You put air into your bicycle tire yesterday, but this morning it's flat again. The valve stem cap is still on it. Your bicycle has been locked up all night. You infer that you have a leaky tire.
(c) You and a friend are taking the same class. The last exam was pretty easy, if you read the assigned readings. You find out that your friend, who typically does as well as you on assignments and exams, did significantly worse than you on the last exam. You infer that he didn't read the assigned readings.

In each of (a)–(c) you're inferring that the best explanation of the data you have is true; inferring as you do is completely reasonable. Are there other potential explanations for what happened in these situations? Sure. It could be that a thief broke into your house, ate the noodles that were in your refrigerator but stole nothing else. It could be that someone has been breaking in and letting the air out of your tire at night. It could be that the professor has it in for your friend and gave him a low score on the exam even though he did the assigned readings. Each of these is a *possible* explanation for what happened in (a)–(c). But, it seems

that in the cases as described it wouldn't be reasonable to think any of these are what actually happened. The initial explanations that you inferred to be true are better explanations than the other possible explanations just provided. And so, although these alternative explanations are *possible*, the reasonable thing to do in each case is to infer that the best explanation (the initial explanation discussed) is true – exactly what you did. Could you be wrong? Sure, but that doesn't mean your inference wasn't reasonable or that you don't know what happened (this idea is explored later in this chapter). In each of these cases you are using IBE.

There are many other instances where we use IBE. When your physician diagnoses your illness, she's using IBE. When your mechanic figures out what's wrong with your car, she's using IBE. It's plausible that we use IBE in order to gain knowledge from testimony (what other people say), whether this comes by way of what we are told or by way of reading books, articles, and so on (Fricker, 1994; Lipton, 1998). Some, such as Jerry R. Hobbs (2004), go so far as to claim that we use IBE to even understand language, that is, we rely on IBE in order to figure out what people mean when they say things. We use IBE all the time!

To help get a better handle on what each of the earlier situations have in common it will be helpful to spell out IBE more schematically. In these cases, we are reasoning in the following manner:

IBE

1. There are some data in need of explanation.
2. A particular explanation, X, explains the data very well.
3. No competing explanation explains the data as well as X.
4. Therefore, X is true.

Recall (a)–(c) from earlier. In each case you are employing IBE to arrive at your conclusion. Your roommate eating the noodles best explains the fact that they're gone; your tire having a leak best explains why it went flat; your friend failing to do the assigned readings best explains why he did poorly on the exam. In each case, you are inferring that the best explanation of the data is true.

Before moving on there are two points about IBE that we should clarify. The first is that IBE involves comparing various explanations and then inferring that the best one is true. Since the explanations that are compared when making an IBE are competing, at most one of them can be true. But, you might worry, "If an explanation isn't true, is it even an explanation at all?" This is a reasonable worry. After all, if your explanation of why sugar dissolves in water is that 1,000 invisible fairies break the sugar apart, you don't have an *actual* explanation of why the sugar dissolves. You're simply confused. In light of this sort of concern it is helpful to keep in mind that when we refer to explanations in IBE we mean "potential explanations" – things that, if true, would explain the facts. As a result, another way to put IBE is as inferring that the best *potential* explanation is the *actual*

explanation. So, in (a) for example, there are various potential explanations of the missing noodles: your roommate ate them, a thief broke in just to eat your noodles, the noodles simply vanished, and so on. Of these potential explanations one stands out as clearly the best: your roommate ate the noodles. Consequently, you infer that the best potential explanation (my roommate ate the noodles) is the actual explanation of why the noodles are gone.

The second point to clarify concerns what it means for an explanation to be the "best." When making an inference to the best explanation one assesses the virtues of the various potential explanations. A large number of explanatory virtues have been proposed in philosophy of science and appealed to by scientists. Some of the most common are: *simplicity* (as Sir Isaac Newton [1687/1999, p. 794] put it "No more causes of natural things should be admitted than are both true and sufficient to explain their phenomena. … For nature is simple and does not indulge in the luxury of superfluous causes"), *explanatory power* (the range of data explained), *conservatism* (consistency with what we already know/currently accepted theories), and *predictive power* (making accurate predictions). There are several other explanatory virtues that have been proposed, but for our purposes we don't have to explore them all. A grasp of some of the most common explanatory virtues is sufficient for getting a good handle on IBE.[3]

Take the example of the missing noodles again. What makes the claim that your roommate ate the noodles a better explanation than its rivals? It's simpler than either of the other explanations – it doesn't have to posit mysterious thieves who only steal noodles or weird unknown mechanisms that lead to noodles simply vanishing. Although all three potential explanations account for the missing noodles, the explanation that your roommate ate them has more explanatory power than the others. This explanation accounts for why your roommate didn't notice a thief even though he's been at home all day, and it accounts for why your roommate wasn't completely freaked out by the vanishing noodles. It's more conservative than the others too – it fits much better with what we know about the typical behavior of thieves and with what we know about objects not simply vanishing into thin air. It also has more predictive power – it will allow accurate predictions whereas the others won't. For example, you can accurately predict that your roommate likes noodles, that he will answer "yes" if you ask him if he ate the noodles, and so on. The other potential explanations don't provide you with any accurate predictions. In light of all of this, your roommate eating the noodles is clearly the best explanation in this case.

Does IBE Occur in Science?

We've explored what IBE is, and we've seen that IBE is quite common in our everyday lives. But, you might wonder: is it really used in science? After all, science is different than everyday life – it involves labs, experiments, precise measurements, and so on. Since our concern in this chapter is how explanations lead to *scientific*

knowledge – not how they lead to everyday knowledge, it's very important to consider whether IBE actually occurs in scientific practice.

The short answer to this concern is a resounding "Yes!" IBE is very common in science. IBE is how we gain scientific knowledge of theories. When a particular theory generates hypotheses that offer really good explanations of large sets of data, and those hypotheses better explain the data than the hypotheses generated by rival theories, we infer that the theory is true. Of course, this scientific knowledge, like all other knowledge in science, is tentative (we might make revisions as we come to have new data or new theories), but it is knowledge nonetheless. To help see the use of IBE in science, let's consider some important cases where major scientific breakthroughs came as the result of gaining scientific knowledge by way of IBE.

The Discovery of Neptune

> At the beginning of the nineteenth century, it was discovered that the orbit of Uranus, one of the seven planets known at the time, departed from the orbit as predicted on the basis of Isaac Newton's theory of universal gravitation and the auxiliary assumption that there were no further planets in the solar system. One possible explanation was, of course, that Newton's theory is false. Given its great empirical successes for (then) more than two centuries, that did not appear to be a very good explanation. Two astronomers, John Couch Adams and Urbain Leverrier, instead suggested (independently of each other but almost simultaneously) that there was an eighth, as yet undiscovered planet in the solar system; that, they thought, provided the best explanation of Uranus' deviating orbit. Not much later, this planet, which is now known as "Neptune," was discovered (Douven, 2017).

The Theory of Natural Selection

Charles Darwin himself gave this line of reasoning in support of the theory of natural selection in *The Origin of Species:*

> It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection, the several large classes of facts above specified. It has recently been objected that this is an unsafe method of arguing; but it is a method used in judging of the common events of life and has often been used by the greatest natural philosophers. *(1859/1962, p. 476)*

The Oxygen Theory of Combustion

Antoine Lavoisier supported his theory of combustion by appealing to IBE when he inferred the truth of his theory because of its simplicity and explanatory power:

> I have deduced all the explanations from a simple principle, that pure or vital air is composed of a principle particular to it, which forms its base, and which I have named *the oxygen principle*, combined with the matter of fire and heat. Once this principle was admitted, the main difficulties of chemistry appeared to dissipate and vanish, and all the phenomena were explained with an astonishing simplicity. *(1862/1978, p. 623)*[4]

There are many other examples of IBE in science. Copernicus's theory that the sun is the center of the solar system replaced the older Ptolemaic model that claimed the earth was the center of the solar system because Copernicus's heliocentric theory better explained things like the observed retrograde motion of various planets. Joseph John Thomson's discovery of the electron was the result of IBE. Thomson inferred that electrons exist because such a particle best explained the observed behavior of cathode rays. We could go on and on. IBE is widespread in science.[5]

Now we have a good answer to our question for this chapter. Explanations lead to scientific knowledge via IBE. So, we're done, right? Not quite. Although many of history's greatest scientists explicitly relied on IBE, and we use it in everyday life, some object that IBE is not a legitimate form of reasoning. If IBE is an illegitimate way to reason, then it doesn't provide us with scientific knowledge, even if it has gotten us to true theories in science. If these criticisms of IBE are correct, then the times that IBE has yielded true scientific theories have just been a matter of luck. And, lucky guesses aren't knowledge. Fortunately, those who doubt IBE's legitimacy are mistaken. However, it's important to not just know that they are mistaken, but to understand why. Let's look at some of these criticisms and see why IBE is unfazed by them.

Is IBE a Bad Way to Reason?

There have been many criticisms of IBE as a method of reasoning. Unfortunately (or, fortunately, depending on how quickly you want to finish reading this chapter!), we don't have the space to examine all of these criticisms or to delve too deeply into the ones we will consider. Nevertheless, we will take a look at three of the more prominent objections to IBE and briefly consider what can be said in response to them.

What if Our Best Explanation Is Only the Best of a Bad Lot?

One of the most well-known objections to IBE is Bas van Fraassen's "Best of a Bad Lot". At the heart of this objection is the idea that when we choose the best available explanation from a set of competing explanations "our selection may well be the best of a bad lot" (van Fraassen, 1989, p. 143). As van Fraassen explained:

> To believe is at least to consider more likely to be true, than not. So to believe the best explanation requires more than an evaluation of the given hypothesis. It requires a step beyond the comparative judgment that this hypothesis is better than its actual rivals. … For me to take it that the best of set X will be more likely to be true than not, requires a prior belief that the truth is already more likely to be found in X, than not. *(1989, p. 143)*

There are two ways of understanding van Fraassen's Best of a Bad Lot objection. One way is to understand it as a version of "No Sign of Truth" objection, which we will discuss in the next section. Another way of understanding van Fraassen's objection is as the challenge that we might have considered only bad explanations. Essentially, the thought is that we can't infer that the best explanation is true because even though it's better than the other explanations we've thought about, it might still be a crummy explanation overall. Although this way of construing the objection has a straightforward response, it is worth considering for at least two reasons. First, one might think of the Best of a Bad Lot in this way. Second, thinking about the response to this way of construing the Best of a Bad Lot helps make clear an important qualification of IBE.

How do we defend the legitimacy of IBE as a method of reasoning from the Best of a Bad Lot? We remember premise (2) from earlier – in order to reason via IBE the explanation whose truth we are inferring must "explain the facts very well." In other words, the best explanation must be "good enough."[6] So, van Fraassen's objection doesn't defeat IBE, it simply illuminates the fact that IBE should be understood as inference to the best available explanation that it is of sufficiently high quality.

Two questions are worth briefly pausing over here: first, how could van Fraassen, or anyone else, ever think IBE had this problem since premise (2) is clearly a component of it? And, second, what does it take for an explanation to be "good enough"? The first question has a pretty simple answer – IBE often isn't presented as carefully as it should be. So, in many formulations of IBE, premise (2) is left out. Without the restriction to sufficiently good explanations, van Fraassen's objection is a major problem for IBE. Hence, we should be careful to always understand this qualification to be in place when discussing IBE.

The second question isn't quite as easy to answer. Of course, the explanatory virtues that we mentioned above will play a key role. They must be used to determine how good an explanation is. However, there's still a difficulty here. There must be a cutoff so that some explanations are good enough to be inferred (when they are the best) and others aren't (even if they are the best). Exactly where to draw this line is difficult to determine. It may be that the exact cutoff will depend upon the circumstances. In some cases, an explanation might be good enough to be inferred, but in other circumstances that same explanation may not be good enough to be inferred. It could very well be that the stakes of accepting a particular explanation as true can affect when it counts as "good enough" to infer. It

is also likely that there will be tough cases – situations where the best explanation is close to the cut-off, and so it's hard to tell whether it is good enough or not. Unfortunately, sometimes science, like life in general, is hard. There will be cases where it's hard to determine whether an explanation is good enough to infer its truth even if we can determine that it is the best available explanation. This shouldn't trouble us though. After all, we're talking about science – we can always run more experiments and gather more data; doing that will allow us to better determine the quality of an explanation. Not to mention that there will be many cases where the answer is clear – there are many cases where the best explanation is clearly good enough, and many where it's clearly not.

Why Should We Think that Explanatory Virtues Are Signs of Truth?

Another objection to IBE concerns whether we should think that explanatory virtues are signs of truth. What we might call the "No Sign of Truth" objection is the idea that we don't have good reason to think that explanatory virtues, for example, simplicity, explanatory power, and so on, are actually related to the truth. As van Fraassen (1980, p. 90) puts it, "some writings on the subject of induction suggest that simpler theories are more likely to be true. But it is surely absurd to think that the world is more likely to be simple than complicated."[7] More broadly, those, like van Fraassen, who press the No Sign of Truth objection, ask: why think that the world is such that best explanations are true? As Peter Lipton (2004, p. 144) explains, those who make this objection to IBE think "It would be a miracle if using explanatory considerations as a guide to inference were reliably to take us to the truth." The worry is that if we don't have good reason to think that explanatory virtues are indicators of the truth of an explanation, IBE is merely guesswork. And, guesses aren't scientific knowledge.

Fortunately, we have good reason to think that explanatory virtues are signs of truth. First of all, there are several instances in the history of science where a theory was initially accepted because of IBE (inferring that the most explanatorily virtuous explanation is true) and then later confirmed via observational methods. For example, as we mentioned above, the existence of Neptune was inferred via IBE before we could actually observe it. Later, we were able to observe this eighth planet and find that our IBE was correct. Again, as we saw above, the electron was thought to exist because of an IBE. We now have things like electron microscopes whose existence gives confirmation that our IBE was correct. There are many more such examples.

In addition to these cases there are numerous instances in ordinary life where we make IBEs and then confirm the conclusion directly via some other method. Recall (a)–(c) from earlier. You infer that your roommate ate the noodles because that is the best explanation of the data. You confirm this later when he tells you that he ate them. You reason via IBE that your tire has a leak in it. You later confirm this when you get it fixed. You come to think your friend didn't do the

assigned readings because this best explains his doing poorly on the exam. You confirm this when he tells you that he didn't do the readings. We use IBE all the time in our daily lives, and the conclusions it leads us to tend to be correct. This gives us plenty of confirmation that explanatory virtues are providing us with good indications of what's true.

What about Explanations We Haven't Thought of Yet?

The final objection we'll consider here is another from van Fraassen (1989, p. 146):

> I believe, and so do you, that there are many theories, perhaps never yet formulated but in accordance with all evidence so far, which explain at least as well as the best we have now. Since these theories can disagree in so many ways about statements that go beyond our evidence to date, it is clear that most of them by far must be false. I know nothing about our best explanation, relevant to its truth-value, except that it belongs to this class. So I must treat it as a random member of this class, most of which is false. Hence it must seem very improbable to me that it is true.

Essentially, van Fraassen is worried about potential explanations that we haven't thought of yet. He claims that not only are there explanations we haven't thought of, but also that many of these undiscovered explanations are as good as, or better than, our current best explanation. Since there are all these really good undiscovered potential explanations, how can we infer that our current best explanation really is the best, let alone that it is true?

Admittedly, throughout the history of science we've replaced what were our best theories with better ones. So, it is possible that we may do the same when it comes to some of our current best theories. Nonetheless, this shouldn't cause us to doubt IBE. For one thing, it is questionable whether our current best theories really are just one among many equally good potential explanations as van Fraassen thinks. It is plausible that the way in which we come to form theories today, at least in the empirical sciences, precludes our best explanations from being merely members of a large set of mostly false theories like van Fraassen claims.[8] We have a large amount of background knowledge that goes into our theory formation practices, and this background knowledge increases the likelihood that our best explanations are correct. Additionally, even if there are other equally good explanations that we haven't thought of yet, that doesn't mean that we can't trust IBE. At most this means that IBE is fallible – it's not a perfect way to reason; sometimes it can lead us astray. Sometimes the current best explanation isn't the best that we could have thought of; sometimes the current best explanation isn't true. This shouldn't be troubling or at all surprising; all human reasoning is fallible. What matters is that the current best explanation is often true. The fact that IBE isn't perfect, that is, it will sometimes yield a false conclusion, doesn't mean that

it's not a good way to reason any more than the fact that airplanes sometimes malfunction means that flying isn't a good way to travel. As J.D. Trout (2016, p. 204) says, "IBE works, so it is unclear the ultimate purpose of philosophers' criticisms of IBE." At the end of the day, IBE is a good way to reason and its fallibility simply means that we should be cautious when it comes to drawing inferences, and we should regard our scientific knowledge as tentative (we should be ready to change what scientific theories we accept if future discoveries call for a change).[9]

What's the Upshot?

So, what does our discussion reveal? First, when IBE is properly understood (when it has been qualified so as to avoid the objections mentioned in the section **Is IBE a Bad Way to Reason?**) it is how explanations lead us to scientific knowledge. Second, the scientific knowledge that we gain from employing IBE is tentative and revisable, that is, IBE is fallible, so we may have to revise what we take to be knowledge in light of new evidence. Does this mean that scientific knowledge can be false? In other words, does the tentative nature of scientific knowledge mean that we might know now that a scientific theory is true and then later come to know that it is false? No. What this means is that we might think that we know a particular theory is true, and then later we might come to realize that we didn't know that theory after all. This happens in science, and it happens in our daily lives. Does this mean that we shouldn't trust IBE? No, it means that we should approach our use of IBE, like all of our scientific practices, with caution and humility. Ultimately, the upshot here is that we should recognize that explanations lead us to scientific knowledge by way of IBE, but there's no easy route to the truth and even our best methods (which include IBE) aren't perfect.

Acknowledgments

Thanks to Kostas Kampourakis, Kevin Lee, Molly McCain, Ted Poston, and Siddhu Srikakolapu for helpful comments on earlier drafts.

Notes

⋆ The points made in this chapter are discussed in greater detail in McCain (2015) and (2016).

1 Admittedly, not everyone will agree with this. In particular, anti-realists about science argue that the primary aim of science is to construct theories that simply fit what we can observe. They will insist that explanations that go beyond what is observable by claiming to explain underlying phenomena are suspect. For more on realism versus anti-realism, see Chapter 16 of this volume.

2 For helpful overviews of the various accounts of explanation that have been defended, see Salmon (1990) and Woodward (2014).

3 For a nice overview of even more explanatory virtues, see Beebe (2009).
4 This is Paul Thagard's (1978) translation.
5 See Kampourakis (2014) and Trout (2016) for more examples of IBE in science. In fact, Trout goes so far as to argue that the success of modern science in general is due to IBE.
6 For more on the importance of restricting IBE to explanations that are good enough, see Lipton (2004).
7 Recall, that van Fraassen's (1989) "Best of a Bad Lot" objection might also be understood as pressing this objection as well.
8 For more on this, see Lipton (2004) and Psillos (1999).
9 Importantly, when it comes to our well supported scientific theories changes often occur at the periphery rather than the core. In other words, we don't often completely abandon a previously successful theory; rather, when we encounter enough conflicting evidence, we revise parts of the theory while retaining the core components that yielded its previous successes.

References

Beebe, J. (2009). The Abductivist Reply to Skepticism. *Philosophy and Phenomenological Research*, 79, pp. 605–636.

Darwin, C. (1859/1962). *The Origin of Species*. New York: Collier.

Douven, I. (2017). Abduction. In E.N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, Summer 2017 Edition. URL = https://plato.stanford.edu/archives/sum2017/entries/abduction/.

Fricker, E. (1994). Against Gullibility. In B.K. Matilal and A. Chakrabarti, eds, *Knowing from Words*. Netherlands: Kluwer, pp. 125–161.

Hobbs, J.R. (2004). Abduction in Natural Language Understanding. In L. Horn and G. Ward, eds., *The Handbook of Pragmatics*. Oxford: Blackwell, pp. 724–741.

Kampourakis, K. (2014). *Understanding Evolution*. Cambridge: Cambridge University Press.

Kim, J. (1994). Explanatory Knowledge and Metaphysical Dependence. *Philosophical Issues*, 5, pp. 51–69.

Lipton, P. (1998). The Epistemology of Testimony. *Studies in History and Philosophy of Science*, 29, pp. 1–31.

Lipton, P. (2004) *Inference to the Best Explanation*. 2nd ed. New York: Routledge.

McCain, K. (2015). Explanation and the Nature of Scientific Knowledge. *Science and Education*, 24, pp. 827–854.

McCain, K. (2016). *The Nature of Scientific Knowledge: An Explanatory Approach*. Dordrecht: Springer.

Newton, I. (1687/1999). *The Principia: Mathematical Principles of Natural Philosophy*. I.B. Cohen and A. Whitman, trans. Berkeley: University of California Press.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. New York: Routledge.

Salmon, W. (1990). *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

Schrödinger, E. (1954). *Nature and the Greeks*. Cambridge: Cambridge University Press.

Strevens, M. (2013). No Understanding Without Explanation. *Studies in History and Philosophy of Science*, 44, pp. 510–515.

Thagard, P.R. (1978). The Best Explanation: Criteria for Theory Choice. *Journal of Philosophy*, 75, pp. 76–92.

Trout, J.D. (2016). *Wondrous Truths: The Improbable Triumph of Modern Science*. Oxford: Oxford University Press.

van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.

van Fraassen, B. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.

Woodward, J. (2014). Scientific Explanation. In E.N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, Winter 2014 Edition. URL = http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/.

5

WHAT IS SCIENTIFIC UNDERSTANDING AND HOW CAN IT BE ACHIEVED?

Henk W. de Regt and Christoph Baumberger

Introduction

Science has not only produced a vast amount of knowledge about a wide range of phenomena, from the nature of elementary particles to the structure of the universe, but it has also enhanced our *understanding* of these phenomena. Indeed, understanding can be regarded as one of the central aims of science. Moreover, scientific understanding is not only important for its own sake, but also highly relevant to society. Climate scientists, for example, want to understand the process of global warming and other climate changes, because such understanding is a first, necessary step towards solving today's environmental problems. Accordingly, the main task of the Intergovernmental Panel on Climate Change (IPCC) is to assess progress in scientific understanding of climate change, as explicitly stated in their latest report *Climate Change 2013: The Physical Science Basis* (IPCC 2013, p.4).

But what exactly *is* scientific understanding, and how can it be achieved? These questions are hotly debated in contemporary epistemology and philosophy of science. While philosophers have long regarded understanding as a merely subjective and psychological notion that is irrelevant from an epistemological perspective, nowadays many of them acknowledge that a philosophical account of science and its aims should include an analysis of the nature of understanding. This chapter reviews the current debate on scientific understanding. We first present the main philosophical accounts of scientific understanding, and we then discuss topical issues such as the relation between understanding and knowledge, the phenomenology of understanding, and the role of understanding in scientific progress.

The Contextual Theory of Scientific Understanding

In the current debate on the nature of scientific understanding, there appear to be two types of approaches. On the one hand, some philosophers emphasize that scientific understanding should ultimately be grounded in objective scientific explanations or knowledge of, for example, causal relations, where understanding consists in a 'grasp' of those explanations or causal relations (Grimm 2017; Khalifa 2017; Strevens 2013). Although they acknowledge that understanding involves a 'grasp', which is a cognitive achievement that relates to the psychology of the subject, they do not see grasping as a sufficient condition for genuine scientific understanding. On the other hand, there are philosophers who put the pragmatics of understanding center stage in their analysis. This typically leads to approaches that invoke the results of empirical study by, for example, psychologists, historians, or sociologists of science. Thus, Faye (2014) bases his pragmatic-rhetorical theory of (scientific) understanding in part on results from cognitive science and evolutionary theory, while De Regt (2017) has developed his contextual theory of scientific understanding on the basis of historical case studies of scientific development. In this section, we will outline De Regt's theory, which was one of the first full-fledged theories of scientific understanding that appeared on the scene, in more detail. In the next section, we will come back to the more objectivist approaches mentioned earlier.

De Regt's theory is based on the analysis of examples from the history of science (esp. physics) and on recent insights that philosophers of science have derived from studying scientific practice. Its central idea is the thesis that scientists achieve understanding of a phenomenon *P* if they construct an appropriate model of *P* on the basis of a theory *T*, following the model-based account of explanation defended by Cartwright (1983, pp.143–162). More specifically, the contextual theory is built upon a Criterion for Understanding Phenomena (De Regt 2017, p.92):

> **CUP:** A phenomenon *P* is understood scientifically if and only if there is an explanation of *P* that is based on an intelligible theory *T* and conforms to the basic epistemic values of empirical adequacy and internal consistency.

The key term in this criterion is 'intelligible': understanding of phenomena requires an intelligible theory, where intelligibility is defined as (De Regt 2017, p.40):

> **Intelligibility:** the value that scientists attribute to the cluster of qualities of a theory *T* (in one or more of its representations) that facilitate the use of *T*.

This definition entails that intelligibility is not an intrinsic property of theories, but a context-dependent value: whether or not a theory is intelligible to scientists

depends on, for example, their skills and their background knowledge. Why do scientific theories need to be intelligible to the scientists who use them? De Regt's argument for this claim draws on the work of philosophers Nancy Cartwright (1983) and Margaret Morrison (1999, see esp. pp.60–64), who highlighted the pivotal role of modelling in scientific practice, and in explanatory practices in particular. On the model-based account of scientific explanation, scientists acquire understanding of the phenomena by constructing models, which 'mediate' between relevant theories and the phenomenon-to-be-explained. Constructing such mediating models involves pragmatic judgments and decisions, since models do not follow straightforwardly from theories (and neither do they follow from the empirical data). For example, suitable idealizations and approximations need to be made. De Regt submits that the construction of such models – which provide explanatory understanding of phenomena – requires theories that are intelligible in the sense defined above. Only if scientists' ability to work with the theory allows them to make suitable pragmatic judgments, will they succeed in constructing explanatory models. In sum, understanding a phenomenon on the basis of *T* depends on an appropriate combination of skills of *S* and qualities of *T*.

An example of such a quality is visualizability. This is a theoretical quality that is widely valued, because for many scientists visualizable theories are more tractable and easier to work with (De Regt 2014). But the contextual theory does not imply that visualizability is a necessary condition for the intelligibility of scientific theories. Depending on the context, there may be alternative ways to render theories intelligible.[1] A concrete example of model construction in which visualization plays a role is the explanation of gas phenomena on the basis of the kinetic theory of gases, as it was developed by James Clerk Maxwell and Ludwig Boltzmann in the nineteenth century. The kinetic theory represents gases as aggregates of particles (molecules) in motion obeying the laws of Newtonian mechanics. Specific models of the molecules and their structure have to be constructed in order to explain particular gas phenomena on the basis of the theory. For example, the kinetic explanation of Boyle's law involves the construction of a model (the ideal gas model) that represents gas molecules as point masses, such that application of Newton's laws leads to a theoretical prediction of the relationship between pressure and volume. The ideal gas model does not follow deductively from the kinetic theory, as its construction involves idealizations and approximations. Specific features of the theory – its visualizability, but also its causal aspects – guaranteed its intelligibility to Maxwell and Boltzmann (and to physicists ever since).[2] Subsequently, the construction of more specific molecular models, such as the van der Waals model and the dumbbell model for diatomic gases, has yielded additional or more detailed understanding of gas phenomena.[3]

Alternative Accounts of Scientific Understanding

According to the contextual theory introduced in the previous section, scientific understanding of a phenomenon requires the ability to use a theory to construct

(a model through which one can derive) an explanation of the phenomenon that conforms to the epistemic values of empirical adequacy and logical consistency. Alternative accounts of understanding are typically more demanding in what they require from the explanation and less demanding with respect to the abilities they require from the scientist.[4] Many of them identify understanding with 'grasping' a correct and thus an at least approximately true explanation. Grasping an explanation is distinguished from merely believing or even from knowing the explanation, but it does not require being capable of using a theory to construct the explanation from scratch. The accounts differ in how they conceive of the grasping that they take to be characteristic of scientific understanding.

Stephen Grimm (2006, 2010) suggests that the distinction between grasping and believing an explanation lies in one's ability to answer counterfactual 'what-if-things-had-been-different' questions. This, in turn, is the ability to anticipate the sort of changes that would result if the factors cited as explanatory were different in various ways. Understanding why a certain plane can fly, for example, requires not only the ability to see how Bernoulli's principle applies to the relevant details about the plane and thus to recognize that the shape of its wings (curved on the top and flat on the bottom) creates a difference in the velocity of air and thus in the pressure exerted along the top and the bottom of the wings. One needs also to be able to anticipate how changes, for instance in the velocity, would lead to a change in the pressure of the air, and, as result, see that if the top of the wings were flattened out, the plane would not be able to fly anymore (Grimm 2010, pp.340–341). Since one can be able to reason counterfactually about an explanation in this way without being capable of actually deriving the explanandum, Grimm's condition of grasping is less demanding than De Regt's requirement of intelligibility.

Mark Newman (2012, 2017) agrees that grasping a scientific explanation involves abilities that are not required for believing and for knowing the explanation. He deems central the ability to draw correct inferences about why the explanans explains the explanandum; for example, why the staying aloft of a plane is a consequence of Bernoulli's principle being applied to the relevant details about the plane. Newman (2012) has developed this idea in his Inferential Account of Scientific Understanding. However, he argues that the grasping necessary for scientific understanding requires neither the ability to construct an explanation, nor the ability to apply it to counterfactual cases and solve new problems. Newman (2017) suggests that if someone has such problem-solving abilities as well, we should rather say that she understands a theory, which she can use to comprehend a whole range of phenomena.

Strevens (2013, 2017) also identifies explanatory understanding with grasping a correct scientific explanation, but ascribes abilities a less prominent role than Grimm or Newman. Strevens takes grasping to be 'the fundamental relation between mind and world' (2013, p.511). He does not give an account of this relation but suggests that it involves a more intimate epistemic acquaintance than knowledge. By way of testimony, you might know Bernoulli's principle, the relevant details about a plane, and that the possibility of flight can be deduced from

them, even if you comprehend only dimly the content of your beliefs. Grasping the explanation requires a firmer grip on the explanatory connections that enable you to see why the principle together with the details about the plane explain why the plane can fly. This idea is familiar from Newman. But whereas Newman identifies the grasping required for scientific understanding with having inferential abilities, Strevens (2017, p.41) suspects that this gets the order of dependence wrong and suggests that the abilities are grounded in the psychological state of grasping.

Unlike the previous accounts, Kareem Khalifa (2017) explicates explanatory understanding with reference to propositional knowledge and acknowledges that understanding admits of degrees. He combines an account of minimal understanding with two comparative principles of understanding. Minimal understanding of why a phenomenon occurs is identified with believing an approximately true explanation of the phenomenon. The first comparative principle says that one person understands better why a phenomenon occurs than another if she has a more complete grasp of the correct explanations of the phenomenon and the relations between them. How complete her grasp is, depends on the number, the quality, and the level of detail of the explanations. The second comparative principle says that one person understands better why a phenomenon occurs than another if her grasp of the phenomenon's explanations and their interrelations bears greater resemblance to scientific knowledge than the second person's. Scientific knowledge of why a phenomenon occurs is based on a scientific evaluation of the explanations. Such an evaluation considers many plausible potential explanations for the phenomenon, compares those using scientific methods, and assigns an appropriate degree of belief for the explanations based on the comparison. Grasping explanations is on this view simply a cognitive state that resembles scientific knowledge of explanations. How close the resemblance is, depends on the number of plausible potential explanations that one has considered and compared, the scientific status of the methods used in this comparison, and the safety and accuracy of the resulting beliefs.

The accounts introduced so far are accounts of explanatory understanding: the understanding of why a phenomenon occurs that results from a scientific explanation of that phenomenon. Explanatory understanding can be distinguished from the more holistic objectual understanding. Here, scientists seek more than a single explanation of a phenomenon that is defined in terms of a small set of salient features and use a theory or model to understand a bunch of phenomena or a system. Examples are understanding climate change through climate models, or the origin of species in terms of evolutionary theory. In his Understanding as Representation Manipulability account, Daniel Wilkenfeld (2013) suggests that to understand a system is to have a mental representation of it that one can modify in such a way that enables one to manipulate or make relevant inferences about the system. The ability to modify a representation consists in being able to correct minor errors in one's representation and apply it to similar cases to cast predictions and give explanations. Wilkenfeld (2017) conceives of the ability to make relevant inferences about a system as an evaluative criterion rather than a necessary condition for scientific understanding,

and suggests representational accuracy as a further criterion for assessing how good someone's understanding is. Building on this proposal, Baumberger (forthcoming) adds justification as a third evaluative criterion and commitment as a necessary condition for objectual understanding (cf. Elgin 2017, p.44). Taking the ability to make relevant inferences, representational accuracy and justification as evaluative criteria rather than necessary conditions accommodates the insight that these dimensions play an important role in the ascription and assessment of scientific understanding, without denying that there may be contexts in which we rightly ascribe understanding while some of the conditions are hardly met or not met at all.

Key Issues in the Current Debate about Scientific Understanding

The accounts that we discussed earlier raise a number of systematic issues. For example, is understanding a form of knowledge, as some accounts assume? How are understanding and the grasping that many accounts take to be essential for understanding related to the 'feeling' or 'sense' of understanding? What is the role of understanding in scientific progress, and how should we account for this role? These questions are addressed in the following paragraphs.

Understanding, Truth, and Knowledge

You can obviously know *that* a phenomenon occurs without understanding why it occurs, but is it also possible to have knowledge-*why* without understanding-why? Not for knowledge-based accounts that identify understanding with explanatory knowledge. An example is Peter Lipton (2004, p.30) who declared: 'Understanding is not some sort of super knowledge, but simply more knowledge: knowledge of causes.' Philosophers who take explanatory understanding to be more demanding than explanatory knowledge typically claim that the former involves abilities that are not necessary for the latter, for example, the ability to use a theory to construct an explanation (De Regt), the ability to draw inferences about why the explanans explains the explanandum (Newman), or the ability to engage in counterfactual reasoning (Grimm).

Defenders of knowledge-based accounts object that some of these abilities are not necessary for (minimal) understanding, and those that are necessary are already required for explanatory knowledge (Khalifa 2017, pp.54–60). It seems, for example, possible to have some understanding of a phenomenon even if one is neither able to construct an explanation, nor to draw inferences about why the explanans explains the phenomenon. The ability to engage in counterfactual reasoning, on the other hand, seems already necessary for explanatory knowledge. Even believing an explanation requires that one is able to answer some what-if questions (Grimm 2014, p.388). For example, believing that the lift of a plane depends (according to Bernoulli's principle) on a difference in the velocity of air traveling over the wings, requires being able to see that had there been no

difference in velocity, there would not have been a difference in pressure and hence no lift. Thus, (minimal) understanding does not seem to involve abilities that are categorically distinct from those required for believing an explanation. Good or deep understanding may involve additional abilities, but they might also be required for more demanding instances of scientific knowledge-why (Khalifa 2017, pp.61–63).

Whether some form of knowledge is *sufficient* for understanding is thus mainly a question of whether the latter requires abilities that are not required for the former. A further issue is whether knowledge is *necessary* for understanding. Is it for instance possible to understand why a phenomenon occurs without knowing why it occurs? This depends on our notion of knowledge. Epistemologists typically conceive of knowledge as justified true belief, the truth of which is not due to epistemic luck. For each necessary condition for knowledge, it has been argued that it is not necessary for understanding. Some have argued that understanding is compatible with certain types of epistemic luck that undermine knowledge.[5] Others have suggested that, in contrast to knowledge, understanding requires neither belief nor justification.[6] However, the liveliest debate about a possible divergence between knowledge and understanding concerns the question of whether understanding implies truth.

Knowledge is factive: if one knows that p, then p is true. If understanding is a form of knowledge, understanding must be factive too. Since scientific explanations are often complex and involve for instance initial conditions and generalizations, a factivity condition for explanatory understanding requires that all propositions constituting the explanation be true. Adapted to objectual understanding such a condition requires that all propositions constituting one's representation of the target system be true. At least for objectual understanding, such a strong factivity condition seems too demanding. A few peripheral falsehoods may degrade one's understanding, but do not undermine it completely. Thus, moderate factivists only require that all central propositions be true (Kvanvig 2003, pp.201–202).

Non-factivists argue that scientific understanding is not even moderately factive. They point out that we gain understanding through idealized models (e.g., the ideal gas model) and superseded theories (e.g., Newton's theory of gravitation), even though both contain non-peripheral falsehoods (e.g., the assumption that a gas consists of perfectly elastic point masses that do not interact with each other in the case of the ideal gas model, and the assumption that there are gravitational forces in the case of Newton's theory of gravitation). Moreover, we can understand phenomena in terms of non-propositional representations such as diagrams and material models, which are not even truth-apt (De Regt 2015; Elgin 2017).

Moderate factivists can pursue two strategies to save factivity (Baumberger, Beisbart and Brun 2017, pp.8–10). First, they can argue that the alleged counterexamples do not display genuine understanding. The idea is that if we ascribe understanding in such cases, we use the term honorifically, as when we speak of 'the current state of scientific knowledge' while conceding that part of it

may be false (Greco 2014, pp.297–298). However, since at least idealized models do hardly preclude genuine understanding, the second strategy seems more promising. It claims that in those examples that display genuine understanding, moderate factivity is not really violated. Understanding a phenomenon with an idealized model, for example, requires that a scientist knows what idealizations the model involves, which aspects of the phenomenon it is intended to describe, and under which conditions the phenomenon approximately behaves as the model. If a scientist knows all this, her central beliefs about the phenomenon are true (Greco 2014, pp.296–297). In the ideal gas case, we need to distinguish between the ideal gas law, the conditions for its application, and the idealizing assumptions that are needed to derive the ideal gas law. These assumptions (e.g., that the particles do not interact) are indeed false, but they are at the periphery of the model since they do not belong to the description of the behavior of real gases. This behavior is rather characterized by the ideal gas law. The law and its conditions of applicability constitute the central propositions of the model. Since in successful applications of the model, the conditions are satisfied and the law is approximately true, moderate factivity seems to hold (Mizrahi 2012; cf. Khalifa 2017, pp.173–175).

Non-factivists can respond to this defense by pointing out that in some cases, we credit scientists with an understanding of a phenomenon even though they do not exactly know how their models diverge from the phenomenon or under which conditions the models provide an approximately true description of the phenomenon. Moreover, De Regt (2015) suggests examples from economics and ecology, in which scientists acquire understanding by applying models whose central proposition are not even approximately true.

Factivists and non-factivists face different challenges. Factivists need to explain how idealized models and flawed theories can contribute to understanding. To meet this challenge, Strevens (2017) argues that idealizations enhance understanding by highlighting that certain factors make no difference to the explanandum, and that we can learn why other factors are difference-makers by manipulating idealized models. Non-factivists, on the other hand, need to explain why we cannot gain understanding by any kind of just-so story or false theory. To meet this challenge, De Regt and Gijsbers (2017) suggest that representations provide understanding only if they reliably lead to scientific success, i.e. to true predictions, successful practical applications, and fruitful ideas for further research. Newton's theory of gravitation enables us to understand certain phenomena because it is very successful in a broad range of applications. Understanding and truth are thus connected for De Regt and Gijsbers, but what needs to be true are predictions rather than the theories or models themselves.

The Phenomenology of Understanding

Understanding as insight, or the grasping of an explanation, is often associated with a so-called Aha-feeling, a *Eureka*-moment of the kind experienced by Archimedes

when he took a bath and suddenly understood how he could find out whether or not the crown of the king was made of pure gold, thereby discovering the law of buoyancy that was later named after him. The emotion he experienced was so strong that he allegedly jumped out of his bath and ran naked through the streets of Syracuse, shouting '*Eureka!*': 'I have found it!' The story shows that understanding can come with a particular phenomenology. While interesting from a psychological point of view, one may wonder what the philosophical import of the phenomenology of understanding is. Does the feeling or 'sense' of understanding carry any epistemic weight, and should it therefore be included in a philosophical theory of understanding? Opinions diverge among contemporary philosophers of science and epistemologists.

The debate started with a provocative paper by J.D. Trout (2002), who argued that the sense of understanding is a highly unreliable feeling that is prone to cognitive biases such as the hindsight bias and the overconfidence bias.[7] Accordingly, scientists should never trust feelings of understanding such as Archimedes' *Eureka*-experience. And philosophers wanting to develop a theory of scientific explanation should stay away from the notion of understanding. The reason is, Trout stated, that the phenomenology of understanding does not give a clue as to whether the explanation that gives rise to it is actually correct. Trout suggested that the history of science is replete with examples of scientists experiencing a feeling of understanding and yet being completely wrong, and he cited Ptolemy as a case in point.

Trout's claim is confirmed by empirical studies in cognitive psychology, carried out by Leonid Rozenblit and Frank Keil, which reveal the existence of a so-called illusion of explanatory depth: 'People feel they understand complex phenomena with far greater precision, coherence, and depth than they really do' (Rozenblit and Keil 2002, p.521; cf. Keil 2006). It can be argued that there is no reason to assume that scientists are less prone to this illusion than people in general, and that scientific understanding may therefore be equally biased (Ylikoski 2009).

Most philosophers agree that these empirical results show that the feeling or sense of understanding should be distinguished from the understanding itself, or 'understanding proper', as Kuorikoski (2012) has called it. Still, the results do not entail an unambiguous conclusion about the relation between the two, since that obviously also depends on which conception of 'proper understanding' is adopted.[8] If one assumes that proper understanding can be reduced to explanatory knowledge (as, e.g., Trout and Khalifa have suggested), then the feeling of understanding appears to be neither necessary, nor sufficient for proper understanding and thereby irrelevant for philosophical theories of understanding. But if proper understanding involves some kind of grasping, it is not *a priori* clear how it relates to the feeling of understanding since grasping is a cognitive, psychological state that may come with a particular phenomenology. As we have seen, many authors associate grasping with some kind of ability, for example to construct models of the objects of understanding (De Regt), to make counterfactual inferences about them (Newman, Kuorikoski), or to manipulate representations

of them (Wilkenfeld 2013). As Grimm (2009, p.85) notes, 'at a more basic level the "sense of understanding" seems to refer to the exercise of an *ability* or *faculty* that undergirds the phenomenology', and it is an open question whether our sense of understanding is reliable in the sense that the distinctive phenomenology of understanding typically leads us to a correct grasp of the objects of understanding.

Grimm (2009) answers this question in the affirmative, arguing that the feeling of understanding can serve an epistemic function. He admits that it is not reliable per se, but argued that it is 'conditionally reliable', that is, it is 'reliable so long as our background beliefs are more or less sound, our intellectual practices are more or less virtuous, and so on' (Grimm 2009, p.93). Scientists are most likely to feel that they understand something when that bit of information coheres with the rest of their beliefs about the world. While this does not guarantee the reliability of the feeling of understanding, it does imply that it can be checked by reviewing one's background knowledge.

Also Lipton (2009, pp.54–60) argued that although the subjective feeling of understanding should be sharply distinguished from objective proper understanding, the former is not irrelevant to the latter. To begin with, feelings such as the Eureka-experience work as an incentive: the prospect of such pleasant feelings can motivate the search for proper understanding (cf. Gopnik 2000). Moreover, the feeling of understanding may guide our practice of inference to the best explanation, as Lipton held that the best explanation is typically the 'loveliest' explanation: the explanation which, if correct, would provide the most understanding.[9] Of course, this does not by itself prove that the feeling of understanding is a *reliable* guide, but Lipton (2009, pp.59–60) suggested that just as our perceptual system is mediated by subjective experience and yet is a reliable source of knowledge, our subjective experiences of understanding may be 'well calibrated and a reliable guide to theory choice'.

In conclusion, while there is general agreement that proper (scientific) understanding should be distinguished from its phenomenology, there is an ongoing debate over the question of whether the feeling of understanding is at least to some extent reliable as a cue to proper understanding and has accordingly epistemic value. This is both an empirical and a conceptual matter and merits further investigation on both fronts.

Understanding, the History of Science, and Scientific Progress

Since it is generally accepted that science, in the course of its historical development, has progressed enormously, and since progress in science can be defined as increasing success in achieving its aims, it seems obvious that scientific progress involves increase in understanding. Surprisingly, however, it is hard to find statements to this effect in the work of philosophers of science. The reason may be that although few philosophers will deny that science has progressed, there is no agreement about the nature of scientific progress, and philosophers of science have

debated the question of what it exactly consists in for decades. In this section, we will focus on the relation between scientific progress and scientific understanding. Does scientific progress involve, or even consist in, an increase of understanding? Or does progress comprise something else, for example, increase of knowledge or of problem-solving power, or an ever closer approximation to the truth? In the latter case, it is not *a priori* true that progress in science implies an improvement of (increase in) understanding.

As mentioned, it is remarkable that the relation between progress and understanding has not received much explicit attention in the philosophical literature until very recently. Most earlier accounts of scientific progress were framed in terms of the concepts listed above: knowledge, problem-solving power, or truth-approximation.[10] A recent exception is Angela Potochnik, who in her 2017 book *Idealization and the Aims of Science* defends the claim that understanding rather than truth is the key aim of science and suggests that this allows for a more convincing account of progress in science. She argues that accounts of scientific progress in terms of truth-approximation face at least two problems that suitable understanding-based accounts do not (Potochnik 2017, p.121). The first is the so-called pessimistic meta-induction, which implies that our current scientific theories may be radically false, such that no convergence to the truth has been achieved so far. The second is the crucial role of idealizations in science, which suggests that progress can be made by means of departing from the truth. Potochnik claims that an account of scientific progress in terms of increase in understanding would avoid these problems.

Whether that is the case depends of course on the conception of scientific understanding that one invokes. Earlier in this chapter we encountered many competing accounts of scientific understanding, so which one should we choose? Since the issue of scientific progress concerns the historical development of science, we suggest looking at the contextual theory first, since this theory is based on historical evidence and sensitive to changes in the historical context. However, this immediately leads to a problem: while the contextual theory acknowledges that understanding is, and has always been, an aim of science, it asserts that standards for the intelligibility of scientific theories – which is a necessary condition for understanding phenomena – may change over the course of history, and sometimes do so in a radical way. Examples include the development of theories of gravitation from Descartes and Huygens, via Newton, to Einstein (see De Regt 2017, ch. 5), and the Chemical Revolution, in which phlogiston theory was replaced by Lavoisier's oxygen theory of combustion (see De Regt and Gijsbers 2017). These historical developments embody radical changes in our understanding of the phenomena, which involved equally radical changes in intelligibility standards and the associated set of skills required for constructing scientific explanations. While nobody will deny that the shift from Newton's to Einstein's theory of gravitation constitutes scientific progress, it is less clear how this progress can be accounted for on the contextual theory of scientific understanding. It may seem that we

confront a problem similar to Kuhnian incommensurability (which haunted the problem-solving account of scientific progress): How can we measure an increase of understanding if intelligibility standards change with the historical context?

An answer is provided by Finnur Dellsén (2016), who defends an account of scientific progress in terms of increasing understanding, which he defines as 'grasping how to correctly explain and/or predict aspects of the target phenomenon', where grasping in turn involves the ability to anticipate the behavior of the phenomenon in a variety of circumstances (pp.74–75). Dellsén dubs his view the 'noetic' account of scientific progress and compares it with the traditional epistemic account, which explains progress in terms of an increase of knowledge (Bird 2016). He argues that the noetic account is preferable because it accommodates cases such as Einstein's explanation of Brownian motion on the basis of the kinetic theory, which feature an increase of understanding but no increase of knowledge, as well as cases in which knowledge increases but no understanding is gained. One advantage of his noetic account, Dellsén (2016, p.81) claims, is that it also explains the relevance of pragmatic virtues to scientific progress: while pragmatic virtues such as simplicity do not make a difference to the truth or propositional content of a theory, they do affect the ability to grasp explanations and predictions, and hence the understanding that can be achieved. In the contextual theory of scientific understanding the role of these pragmatic virtues is encapsulated in the definition of intelligibility, the value that promotes the epistemic aim of science: understanding.

Conclusion

In this chapter, we have reviewed the main trends in the philosophical debate about scientific understanding, a debate that started relatively recently and will surely continue and expand in the years to come. Understanding scientific understanding is important for philosophers who want to understand the nature of science, but it is also of interest to a wider audience. Both (aspiring) scientists and the general public may profit from deeper insight in the way science provides us with understanding of the world around us. Today, the dominant view of science in public debates is still based on the idea that scientific research can and should uncover the truth about reality by producing knowledge of incontrovertible facts. When it turns out that real science does not achieve this, that scientists cannot deliver certain knowledge and disagree about the 'facts', the result can be a relativist or even outright anti-scientific attitude among the general public and in politics. This has happened in recent years, and according to many, we are now living in a 'post-truth society', where facts are less relevant than emotions and 'alternative facts' may shape public opinion. A more realistic view of science will help to turn the tide, and a focus on understanding as a central aim of science may play a crucial part here. Acknowledging that science does not and cannot produce certain knowledge but that its value lies rather in the understanding that it delivers, is a first step towards a philosophical account of science that has immediate societal relevance.[11]

Understanding why things are as they are, and understanding how the world works, is more valuable than mere knowledge of facts (if attainable at all). While the pragmatic nature of understanding perhaps detracts from its objectivity, understanding allows for prediction and control in a way that ordinary knowledge cannot, and is thereby of vital importance for the solution of societal problems. Returning to the example of climate science, mentioned in the introduction, it is clear that objective facts regarding climate change are hard to obtain: determining how exactly the climate will change on global and regional scales is an intricate matter, and even though some consensus about it may be achieved within the scientific community, as has been shown by the IPCC, it will always be possible for skeptics to throw doubt on such knowledge-claims. By contrast, scientific understanding of the process of climate change is less prone to such skeptical challenges and at the same time more useful for coping with present and future environmental problems – at least if one adopts a pragmatic conception of understanding, for example along the lines of the contextual theory. For it is the ability to *use* the relevant physical, chemical and biological theories to build climate models, rather than the absolute *truth* of these theories, that allows for prediction, manipulation, and control of the environmental system. As such, scientific understanding of the world is a prerequisite for making it a better place.

Notes

1 See De Haro and De Regt (2018) for examples from current theoretical physics.
2 See De Regt (2017, pp.31–35 and pp.103–106) for an extensive discussion of this example.
3 See De Regt (2017, pp.205–216).
4 A reason is that De Regt focusses on the 'primary' understanding that is gained when scientists discover a new explanation for a phenomenon, while most alternative accounts focus on the 'secondary' understanding that is achieved when someone comprehends an already existing explanation for a phenomenon (De Regt 2017, p.100).
5 See Kvanvig (2003), Pritchard (2010), Hills (2016); for a criticism of this view Grimm (2006), Greco (2014) Khalifa (2017, Chapter 7).
6 See Dellsén (2017); Hills (2016); Wilkenfeld (2017).
7 The hindsight bias is the effect that people systematically overestimate their predictive power in after-the-fact contexts: they remember their earlier predictions as having been more accurate than they actually were. The overconfidence bias is the reason why people are 'systematically prone to believing that they are right when they are not' (Trout 2002, p.226).
8 One question is whether philosophers are free to adopt any notion of "proper understanding" they want, independently of empirical psychology. Waskan et al. (2014) argue against anti-psychologism, on the basis of experimental evidence regarding classifications of explanations by laypeople and scientists. They suggested that philosophers of science should take seriously psychological conceptions of explanation (e.g., those that invoke considerations of intelligibility), or else run the risk of being completely out of step with everyday and scientific explanatory practice.

9 See Lipton (2004) for an analysis of inference to the best explanation; see esp. pp. 59–62 for the distinction and relation between lovely and likely explanations.
10 See Bird (2016) for a recent overview of theories of scientific progress, in which the idea that science progresses in achieving the aim of understanding is conspicuously absent.
11 Cf. Kampourakis and McCain (2019), who show that science does not, and cannot, provide certainty but can still contribute to solving societal problems.

References

Baumberger, Christoph (forthcoming) Explicating objectual understanding: Taking degrees seriously. *Journal for General Philosophy of Science*.

Baumberger, Christoph, Claus Beisbart, and Georg Brun (2017) What is understanding? An overview of recent debates in epistemology and philosophy of science. In: Stephen R. Grimm, Christoph Baumberger and Sabine Ammon (eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science.* New York: Routledge, pp.1–34.

Bird, Alexander (2016) Scientific progress. In: Paul Humphreys (ed.), *The Oxford handbook of philosophy of science*. New York: Oxford University Press, pp.544–563.

Cartwright, Nancy (1983) *How the laws of physics lie*. Oxford: Clarendon Press.

Dellsén, Finnur (2016) Scientific progress: Knowledge versus understanding. *Studies in History and Philosophy of Science* 56, pp.72–83.

Dellsén, Finnur (2017) Understanding without justification or belief. *Ratio* 30, pp.239–254.

De Haro, Sebastian, and Henk W. de Regt (2018) Interpreting theories without a spacetime. *European Journal for Philosophy of Science* 8, pp.631–670.

De Regt, Henk W. (2014) Visualization as a tool for understanding. *Perspectives on Science* 22, pp.377–396.

De Regt, Henk W. (2015) Scientific understanding: Truth or dare? *Synthese* 192, pp.3781–3797.

De Regt, Henk W. (2017) *Understanding scientific understanding*. New York: Oxford University Press.

De Regt, Henk W., and Victor Gijsbers (2017) How false theories can yield genuine understanding. In: Stephen R. Grimm, Christoph Baumberger and Sabine Ammon (eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science.* New York: Routledge , pp.50–75.

Elgin, Catherine Z. (2017) *True enough*. Cambridge, MA: MIT Press.

Faye, Jan (2014) *The nature of scientific thinking: On interpretation, explanation, and understanding*. London: Palgrave Macmillan.

Gopnik, Alison (2000) Explanation as orgasm. *Mind and Machines* 8, pp.101–118.

Greco, John (2014) Episteme: Knowledge and understanding. In: Kevin Timpe and Craig A. Boyd (eds.), *Virtues and their vices*. Oxford: Oxford University Press, pp.285–302.

Grimm, Stephen R. (2006) Is understanding a species of knowledge? *British Journal of Science* 57, pp.515–535.

Grimm, Stephen R. (2009) Reliability and the sense of understanding. In: H. W. de Regt, S. Leonelli and K. Eigner (eds.), *Scientific understanding: Philosophical perspectives*. Pittsburgh: University of Pittsburgh Press, pp.83–99.

Grimm, Stephen R. (2010) The goal of explanation. *Studies in History and Philosophy of Science* 41, pp.337–344.

Grimm, Stephen R. (2014) Understanding as knowledge of causes. In: A. Fairweather (ed.), *Virtue epistemology naturalized: Bridges between virtue epistemology and philosophy of science*. Cham: Springer, pp.329–345.

Grimm, Stephen R. (2017) Understanding and transparency. In: Stephen R. Grimm, Christoph Baumberger, and Sabine Ammon (eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science.* New York: Routledge, pp.212–229.

Hills, Alison (2016) Understanding why. *Noûs* 50, pp.661–668.

IPCC (2013) *Climate change 2013 – The physical science basis, contribution of working group I to the fifth assessment report of the IPCC*. Cambridge: Cambridge University Press.

Kampourakis, Kostas and Kevin McCain (2019) *Uncertainty: How it makes science advance*. New York: Oxford University Press.

Keil, Frank (2006) Explanation and understanding. *Annual Review of Psychology* 57, pp.227–254.

Khalifa, Kareem (2017) *Understanding, explanation, and scientific knowledge*. Cambridge: Cambridge University Press.

Kuorikoski, Jaakko (2012) Simulation and the sense of understanding. In: C. Imbert, and P. Humphreys (eds.), *Models, simulations, and representations*. London: Routledge, pp.168–187.

Kvanvig, Jonathan (2003) *The value of knowledge and the pursuit of understanding*. New York: Cambridge University Press.

Lipton, Peter (2004) *Inference to the best explanation*. 2nd ed. London: Routledge.

Lipton, Peter (2009) Understanding without explanation. In: H.W. de Regt, S. Leonelli and K. Eigner (eds.), *Scientific u*nderstanding*: Philosophical p*erspectives. Pittsburgh: University of Pittsburgh Press, pp.43–63.

Mizrahi, Moti (2012) Idealizations and scientific understanding. *Philosophical Studies* 160, pp.237–252.

Morrison, Margaret (1999) Models as autonomous agents. In: M.S. Morgan and M. Morrison (eds.), *Models as mediators*. Cambridge: Cambridge University Press, pp.38–65.

Newman, Mark (2012) An inferential model of scientific understanding. *International Studies in the Philosophy of Science* 26, pp.1–26.

Newman, Mark (2017) Theoretical understanding in science. *British Journal for the Philosophy of Science* 68, pp.571–595.

Potochnik, Angela (2017) *Idealization and the aims of science*. Chicago: University of Chicago Press.

Pritchard, Duncan (2010) Knowledge and understanding. In: Duncan Pritchard, Alan Millar and Adrian Haddock, *The nature and value of knowledge: Three investigations*. Oxford: Oxford University Press, pp.1–88.

Rozenblit, Leonid, and Frank Keil (2002) The misunderstood limits of folk science: An illusion of explanatory depth, *Cognitive Science* 26, pp.521–562.

Strevens, Michael (2013) No understanding without explanation. *Studies in History and Philosophy of Science* 44, pp.510–515.

Strevens, Michael (2017) How idealizations provide understanding. In: Stephen Grimm, Christoph Baumberger and Sabine Ammon (eds.), *Explaining understanding: New essays in epistemology and the philosophy of science.* New York: Routledge, pp.37–49.

Trout, J.D (2002) Scientific explanation and the sense of understanding. *Philosophy of Science* 69, pp.212–233.

Waskan, Jonathan, Ian Harmon, Zachary Horne, Joseph Spino, and John Clevenger (2014) Explanatory anti-psychologism overturned by lay and scientific case classifications. *Synthese* 191, pp.1013–1035.

Wilkenfeld, Daniel (2013) Understanding as representation manipulability. *Synthese* 190, pp.997–1016.

Wilkenfeld, Daniel (2017) MUDdy understanding. *Synthese* 194, pp.1273–1293.

Ylikoski, Petri (2009) The illusion of depth of understanding in science. In: H. W. de Regt, S. Leonelli and K. Eigner (eds.), *Scientific understanding: philosophical perspectives*. Pittsburgh: University of Pittsburgh Press, pp.100–119.

PART II

What Is the Nature of Scientific Knowledge?

6

WHAT ARE SCIENTIFIC CONCEPTS?

Theodore Arabatzis

Introduction

Scientific concepts play representational and heuristic roles in the acquisition of scientific knowledge. On the one hand, they represent entities, properties, and processes in nature. On the other hand, they facilitate, or even make possible, the investigation of those entities, properties, and processes. Not surprisingly, the formation, development, and identity of scientific concepts have been key issues in the epistemology of science. Their evolution across time, especially, has given rise to intractable problems about the rationality of scientific change and the ability of science to approach truth.

In this chapter I discuss the nature and function of scientific concepts, what it takes to possess them, how they can be represented, and how they can be studied by examining the uses of the scientific terms associated with them. I then examine the epistemological issues that arise when considering conceptual change. Furthermore, I draw a distinction between concepts referring to manifest entities (accessible to observation) and concepts referring to hidden entities (temporarily or permanently unobservable). I argue that the function of scientific concepts is different in the two cases. In the former case, their function is primarily classificatory; whereas in the latter case, their function is primarily explanatory. Finally, I suggest that the epistemological problems generated by the evolution of scientific concepts are more severe in the latter case than in the former.

The Nature and Function of Scientific Concepts

Concepts are supposed to be things in the head: mental representations of objects, properties, processes, and so on (cf. Margolis & Lawrence 2014). In that sense, they

are theoretical constructs of cognitive psychology. They are posited to account for various abilities that humans have, such as the ability to unify and to discriminate. Furthermore, they are taken to be involved in central cognitive processes: "perception … memory … classification and inferences" (Johnston and Leslie 2012, p. 133).

As a historian and philosopher of science who studies scientific practices and their products, I do not pretend to know what goes on in scientists' heads. So, here I adopt an approach inspired by Wittgenstein, who has taught us "that concepts cannot be divorced from the practices of their employment" (Davidson 2001, p. 181; cf. Kindi 2012). Rather than taking concepts to be hidden psychological entities, I consider them to be associated with public, not mental, representations of objects (e.g., atoms, cells, genes, planets), properties (e.g., electric charge, mass), processes (e.g., electric currents, electromagnetic waves), and phenomena in particular domains (e.g., planetary motion). Thus, one can study scientific concepts by looking at the associated representations (e.g., linguistic, mathematical, diagrammatic) in textbooks, research articles, monographs, observational and experimental reports, and so on.

Scientific concepts play several epistemic roles in scientific practice. **First**, they structure scientific observation and the reporting of experimental results. Physicists, for example, report their observations of cloud chamber tracks in terms of 'electrons', 'positrons', 'muons', and so on. **Second**, scientific concepts enable the detection of regularities and empirical laws. For instance, the formation of the concepts of positive and negative electricity in the early 18th century made possible the discovery of various regularities in the domain of electrical phenomena (Steinle 2012). **Third**, they guide the design of experiments. For instance, in the late 19th century experiments for detecting the influence of magnetism on radiation were guided by the concept of the electron (Arabatzis 1992). **Fourth**, they go hand in hand with the classification of objects. For example, the formation of the concepts of 'planet' and 'fixed star' in Ptolemaic astronomy were associated with a corresponding classification of the heavenly realm. **Fifth**, they enable inferences about the objects they refer to (Johnston & Leslie 2012, p. 118). **Sixth**, they facilitate the solution of theoretical problems. For example, the concept of the electron made possible the understanding of electrical conduction, which had been a vexing problem in Maxwell's electromagnetic theory (see Arabatzis 2006). And, **seventh**, they enable the explanation of phenomena via hidden entities and mechanisms. For instance, the concept of the atom in 19th-century chemistry enabled the explanation of phenomena such as the laws of definite and multiple proportions (see Chalmers 2009). In all these respects, concepts function as tools for generating and validating knowledge (see, e.g., Brigandt 2010; Feest 2010; MacLeod 2012, Steinle 2012).

Possessing Scientific Concepts

What does it take to possess a scientific concept? The ability to use (correctly) the corresponding scientific term, or in other words, to be able to apply it

for descriptive, explanatory, and problem-solving purposes. Thus, the ability to use correctly a scientific term 'X' amounts to possessing the concept of X. Explaining this ability though is a non-trivial matter.[1] Various possibilities have been debated in the psychological and philosophical literature (see Margolis and Lawrence 2014; and Cheon and Machery 2016). According to the 'classical' view, possessing a concept amounts to knowing its definition: the set of necessary and sufficient conditions for membership in the extension of the corresponding term. However, there is considerable psychological evidence against this view. People do not learn how to use a word by assimilating a definition specifying necessary and sufficient conditions for its application.[2] In response to this difficulty, three other conceptions of 'concept' have been suggested: the 'prototype' view, the 'exemplar' view, and the 'theory-theory' view. According to the prototype view, possessing a concept amounts to knowing "the typical or diagnostic properties" of its referent (Cheon and Machery 2016, p. 514). According to the exemplar view, possessing a concept amounts to having a representation of exemplary instances of its referent. Finally, according to the theory-theory view, to possess a concept is to command a full theory about its referent. There seems to be a consensus that the classical view is deficient,[3] but as regards the other three possibilities the jury is still out.[4]

My own inclination is to adopt a pluralist account of scientific concepts. In certain cases (e.g., in mathematics) the classical view might be pertinent, whereas in others (e.g., in natural history or in particle physics) versions of the other three views might be more appropriate. Be that as it may, for the purposes of this chapter I refrain from entering into this explanatory issue, not the least because it is impossible to do justice to the extensive psychological literature dealing with it.[5] Suffice it to say that someone who possesses a scientific concept is usually able to state salient properties of the concept's referent. This ability is somehow associated with the mastery of a concept, which, moreover, comes in degrees: one can use a concept correctly in certain circumstances but not in others.

In any case, what is crucial for my purposes is that concepts, whatever their underlying structure turns out to be, are associated with public representations. The public character of scientific concepts, such as 'electron', 'field', 'gene', and so on, as opposed to their private mental counterparts, makes it possible for historians and philosophers of science to study them by examining the evolving representations associated with them; their uses, that is, the objects, properties, and processes to which they are applied; and their relations to other concepts.[6]

Representing Scientific Concepts

The representations associated with concepts can be best characterized by multidimensional schemata, such as those developed by Hilary Putnam and Nancy Nersessian. In the 1970s Putnam defended a view of concept possession that is along the lines suggested earlier:

> an organism possesses a *minimal concept* of a chair if it can recognize a chair when it sees one, and … it possesses a *full-blown concept* of a chair if it can employ the usual sentences containing the word *chair* in some natural language.
>
> *(Putnam 1975b, p. 3)*

Thus, Putnam identified the possession of a concept with the ability to use the corresponding word. This ability, according to Putnam, derives from knowing the "stereotype" associated with the word's referent. In the case of words referring to natural kinds, the stereotype consists of:

> a standardized description of features of the kind that are typical, or 'normal', or at any rate stereotypical. The central features of the stereotype generally are criteria – features which in normal situations constitute ways of recognizing if a thing belongs to the kind …
>
> *(Putnam 1975d, p. 230)*

These features, though, do not function as necessary and sufficient conditions. It may turn out that they have been mistakenly associated with a concept without, however, threatening its identity. As I indicate below, this is crucial for coming to terms with some of the philosophical implications of conceptual change.

In another paper from that period, Putnam articulated further his theory of concepts, by suggesting a representation of the concept associated with a word, such as 'tiger', in terms of a four-dimensional "vector" consisting of:

> (1) the syntactic markers that apply to the word, e.g. 'noun'; (2) the semantic markers that apply to the word, e.g. 'animal'…; (3) a description of the additional features of the stereotype, if any;[7] (4) a description of the extension.
>
> *(Putnam 1975d, p. 269)*

Knowledge of the components of this vector enables one to use correctly the corresponding word. Particularly significant, in this respect, are the semantic markers, which "attach with enormous centrality to the [corresponding] words … form part of a widely used and important *system of classification*", and are "*qualitatively* harder to revise" (Putnam 1975d, p. 267).

Putnam illustrated his proposal with the example of the word 'water'. The four-dimensional vector in this case includes the syntactic markers "mass noun" and "concrete"; the semantic markers "natural kind" and "liquid"; the stereotype "colorless", "transparent", and "tasteless"; and the description of the extension "H_2O" (Putnam 1975d, p. 269). It should be emphasized that the final component of a term's meaning, the description of its extension, leads to a distinction between concepts and meanings. The former are internal psychological entities whereas the latter include an external real-world component. Thus, two speakers may share the same concept (i.e., may be in the same psychological

state) without, however, referring to the same entity. This possibility is illustrated by Putnam's famous "Twin Earth" thought experiment. "Twin Earth is *exactly* like Earth", apart from a few "peculiarities": for instance, on Twin Earth "the liquid called 'water' is not H_2O but a different liquid whose chemical formula is very long and complicated … [say] *XYZ*". Moreover, "*XYZ* is indistinguishable from water at normal temperatures and pressures" (Putnam 1975d, p. 223). Thus, a speaker on Earth and a speaker on Twin Earth would associate the same concept with the term 'water' (i.e., they would be in the same psychological state when using that term), while referring to different substances. The speaker on Earth would refer to H_2O, whereas his or her counterpart on Twin Earth would refer to *XYZ*.

Putnam's schema is very useful for tracking the development of concepts and coming to terms with their evolving identity. One may follow the evolution of the stereotype associated with a concept, while at the same time attending to its extension. The instability of the former need not undermine the stability of the latter and, thus, an evolving concept may retain its identity.

Another significant proposal that captures salient aspects of the structure and identity of concepts has been made by Nancy Nersessian. She argued that the classical view of concepts is at odds with their historical character, their evolution over time in response to empirical and theoretical problems. If a concept were captured by necessary and sufficient conditions, then even the slightest change in those conditions would imply that the concept, as previously used, was vacuous. Nothing would fall under that concept, because nothing would satisfy the necessary and sufficient conditions associated with it. Rather, the concept in question would now be replaced by an altogether different one, associated with altered necessary and sufficient conditions that pick out different things in the world. Thus, conceptual change, which presupposes that in some sense concepts persist through change, would be impossible and would have to be reconceptualized as conceptual replacement.

To allow for the possibility of conceptual change, Nersessian proposed, instead, a more complex representation of concepts:

> The meaning of a scientific concept is a two-dimensional array which is constructed on the basis of its descriptive/explanatory function as it develops over time. I will call this array a "meaning schema". A "meaning schema" for a particular concept, would contain, width-wise, a summary of the features of each instance and, length-wise, a summary of the changes over time.
>
> *(Nersessian 1984, p. 156)*

The features associated with a concept are organized along four different lines:

> "stuff", "function", "structure", and "causal power." … Here, "stuff" includes what it is (with ontological status and reference); "function" includes what

> it does; "structure" includes mathematical structure; and "causal power" includes its effects.
>
> *(Nersessian 1984, p. 157)*

This multi-dimensional schema, besides capturing the complexity of concepts, enables us to chart both synchronic and diachronic variations in their features. Thus, it is well-suited for tracking the continuities and discontinuities of conceptual change (cf. Nersessian 1992, p. 36).

In the sciences, concepts are often embedded in conceptual frameworks, systems of interconnected concepts that order wide domains of objects and phenomena. In Aristotelian cosmology, for instance, the concept of planet was embedded within a geocentric framework, which included concepts for other heavenly objects, such as 'fixed star' and 'comet'. Or, to use another example, in the Newtonian conceptual framework the concepts of mass, force, acceleration, and energy, among others, are interrelated. Learning to use the concept of force requires an understanding of its connections to other concepts, such as 'mass' and 'acceleration'. In cases like this, the learning of concepts is carried out holistically, via exposure to several concepts, their interconnections, and their collective applications.[8]

To capture this dimension of concepts it is necessary to make an addition to Nersessian's schema: a column specifying the location of a concept in the conceptual framework in which it is embedded, that is, its relations to other concepts in that framework, thus creating a concept map. Furthermore, in the "causal power" column one should include, in addition to the effects associated with a concept, its operational dimension, namely the various ways in which the concept is operationalized in the laboratory or in the field (Arabatzis 2012b).

Following Concepts Around

In addition to studying the representations associated with scientific concepts, one may study concepts by tracking the uses of the corresponding words. In Ian Hacking's aptly chosen words,

> Concepts are words in their sites. Sites include sentences, uttered or transcribed, always in a larger site of neighborhood, institution, authority, language. … [To understand a concept] … one would require a history of the [corresponding] words in their sites.
>
> *(Hacking 1990, p. 359; cf. Kindi 2012, p. 29)*

Following this approach, one can study how a new scientific term (e.g., 'electron') emerged, its relations to other terms (e.g.,'atom'), its domain of applications (e.g., electromagnetic phenomena), and its descriptive and explanatory uses in that domain. This study can reveal how new concepts are formed, so as to perform descriptive and explanatory work within a domain; how they gradually

change through an alteration of their relations to other concepts; and how they are affected by their transfer to other domains.

In studying concepts we have to follow their trajectories, not only in the theoretical environment in which they 'live' but also in the observational, experimental, and measuring practices associated with them.[9] These practices, pace Paul Feyerabend, play a significant role in the specification of the meaning of scientific concepts.[10] Observation and experimentation guide the articulation of concepts, by indicating the kinds of properties that their referents should have in order to account for the observational and experimental situations attributed to them. For instance, the articulation of the concept of the electron in the early 20th century was guided by the experimental phenomena attributed to it. The discrete structure of the hydrogen spectrum, just to mention one example, indicated the discrete structure of the energy levels of the hydrogen atom and, thereby, the quantization of electron orbits within the atom (see Arabatzis 2006).

A focus on practices, both experimental and theoretical, may also elucidate the processes of conceptual change. There are three kinds of conceptual change: new concepts are formed, already available concepts evolve, and older concepts disappear. New concepts are formed in response to problems of an empirical or theoretical character. On the one hand, the resolution of empirical problems, such as the individuation and ordering of phenomena in observational or experimental contexts, often requires new concepts. As I mentioned earlier, the detection and description of several regularities in the domain of static electricity during the 18th century was closely related to the formation of new concepts, such as those of positive and negative electricity (Steinle 2005). On the other hand, the solution of theoretical problems, such as the explanation of novel phenomena may also require new concepts of the entities, properties, and processes underlying those phenomena. For instance, new experimental discoveries in the domains of magneto-optics and cathode rays were explained via the novel concept of the electron (Arabatzis 2006).[11]

The evolution of already available scientific concepts also takes place in response to empirical and theoretical problems. On the one hand, those concepts may change in the process of coming to terms with new observational or experimental information. The refinement of experimental phenomena goes hand in hand with the articulation of the concepts that are used to describe and account for them. On the other hand, already available concepts may change in response to theoretical difficulties, such as the incoherence of a theory. For instance, during the Copernican revolution, the concept of planet changed dramatically, from 'wanderer against the fixed stars' to 'object revolving around the sun', in response, partly, to the incoherence of Ptolemaic astronomy (Kuhn 1957).[12]

The evolution of scientific concepts can occur in various ways. **First**, they can expand by incorporating new properties. For instance, in 1925 the property of spin was incorporated into the concept of the electron. **Second**, concepts can contract by shedding older properties. The property of electron orbits, for

instance, was dropped after the development of quantum mechanics and, thereby, the electron ceased to be conceived as an entity with a well-defined trajectory within the atom. **Third**, the representation of a concept's referent may change altogether. For example, in the late 19th century it was established that atoms have a structure and are, thus, not elementary entities. The novel concept of the atom as a structured entity was at odds with its older version (the atom as an indivisible entity). Furthermore, the new concept violated the very etymology of the term 'atom', which suggests something that cannot be cut into pieces and thus divided. **Fourth**, the relations of a concept with other concepts may be altered. The theory of relativity, for instance, changed the relations between the concepts of mass and energy, via Einstein's famous equation, $E=mc^2$. **Fifth**, the function of a concept may change too. Consider the concept of the ether. In the 19th century the ether functioned as "the supporting medium for electromagnetic radiation" (Badino & Navarro 2018, p. 8). In the early 20th century this function was eliminated and electromagnetic radiation was reconceived as a self-subsisting entity. The concept of the ether persisted, though now the ether functioned as the medium for gravitation (see Einstein 1922).

As long as a scientific term continues to be used in connection with the same class of 'things' (objects, properties, processes, etc.), one may reasonably assume that the corresponding concept has remained the 'same', despite its evolution. Sometimes, though, the referent of a concept may shift. As mentioned earlier, this happened, for instance, in the 16th century, when the objects falling under the concept of planet changed. Before Copernicus, the sun and the moon were considered planets, whereas after the revolution initiated by his work the sun was reconceptualized as a star and the moon as a satellite of the earth.

Finally, scientific concepts sometimes die out, as testified by three well-known cases from the history of the physical sciences: the concepts of phlogiston, caloric, and ether. Those concepts played important roles in 18th- and 19th-century chemistry and physics, but eventually they were overthrown. Phlogiston was displaced by oxygen; caloric gave its place to a conception of heat as a form of motion; and the ether was rendered superfluous by Einstein's theory of relativity.

In sum, tracking the uses of scientific terms enables historians and philosophers of science to trace the (synchronic or diachronic) variations of the corresponding concepts. The coining of a new term indicates the emergence of a new concept. A rearrangement of the relations between terms is, again, indicative of conceptual change. And, finally, a shift in the applications of scientific terms and, more rarely, their extinction are marks of conceptual discontinuity.

Conceptual Change and Its Discontents

The historical variation of scientific concepts has, for some time now, been widely accepted, even among analytic philosophers.[13] As Putnam pointed out,

"Instead of treating concepts as eternal objects, one could consider them as objects that come into existence, serve historically contingent goals, [and] die …" (quoted in Davidson 2001, pp. 178–179). Two philosophical problems have been raised in connection with conceptual change: scientific rationality and scientific realism. Take the former first. If different scientists attach different concepts to a single term, then a breakdown in communication seems inevitable.[14] In such a case, scientists would talk past each other, their seeming disagreements would be illusory and, hence, the rational resolution of scientific disputes would become impossible. Various solutions to this problem have been suggested in the philosophical literature. Among the most promising, I would single out the one proposed by Dudley Shapere (1983) and further developed by Nancy Nersessian (1984). Shapere and Nersessian argued that the problem of scientific rationality had been posed in a misleading way, by focusing exclusively on the beginning and final stages of a long process of conceptual change and by comparing concepts before and after that process. They suggested, instead, a different strategy, which consists in examining the successive stages of that process. When this is done, one will then realize that conceptual change takes place gradually, through an exchange of reasoned arguments and without any communication problems.

With regard to the latter problem, it is easy to understand why conceptual change was considered a major threat to a realist account of scientific progress. If scientific concepts evolved and, thereby, ceased to refer to the same objects, properties, and processes, then the ontology of science would be in flux and no sense could be made of the realist credo "that there are successive scientific theories about the same things: about heat, about electricity, about electrons, and so forth" (Putnam 1975c, p. 197). To resolve this problem, while at the same time acknowledging the frequent occurrence of conceptual change in science, one has to specify under which conditions a concept can continue to persist despite the evolution of its meaning. If those conditions are met, then a concept can evolve and at the same time retain its identity, rather than being replaced by an altogether different one.

How can this happen? As I mentioned earlier, concepts have a domain of application, a function, and a conceptual context (i.e., their relations to other concepts). For a concept to maintain its identity, the term associated with it should continue to be used in the same way: to be applied to the same situations and refer to the same 'things'. Of course, more situations can be added to the domain of applications of a term, without thereby destabilizing the identity of its referent.

The stable identity of a concept can be explained in two ways: first, by the retention of a core of properties attributed to its referent (for details, see Arabatzis 2007); and, second, by the continuity of the epistemic function of a concept, that is, the preservation of a core of problems (e.g., explanatory tasks), which continue to be addressed in terms of the concept in question (cf. Brigandt 2010).

Two Kinds of Scientific Concepts: Hidden-Entity-Concepts versus Manifest-Entity-Concepts

Concepts come in "different varieties" (cf. diSessa & Sherin 1998, p. 1169). To understand their formation, the ways in which they change, and the philosophical problems posed by their variation, it is helpful to draw a distinction between two types of concepts: those referring to manifest entities and those referring to hidden entities. A similar distinction, between observational and theoretical concepts, was originally drawn by logical positivists: observational concepts were taken to refer to observable entities and theoretical concepts to unobservable entities. This way of drawing the distinction did not work, however, because the observational theoretical distinction and the observable/unobservable distinction are not coextensive. For instance, the theoretical concept 'harmonic oscillator' refers to both observable and unobservable entities. Furthermore, there are observational concepts denoting unobservable entities and theoretical concepts denoting observable entities (Putnam 1975a). The concept of planet, for instance, is theoretical, in the sense that it was introduced and defined in the context of astronomical theories. Nevertheless, it refers to observable objects, objects that are accessible to unmediated observation.

Moreover, the epistemological and ontological significance of this distinction is a contested issue in philosophy of science (see, e.g., Matheson and Kline 1988).[15] Anti-realists like Bas van Fraassen attach great weight to it, whereas realists doubt that it can be drawn in an epistemologically meaningful way. Be that as it may, the important point here is not those philosophical issues, but rather the significance of the manifest/hidden distinction for understanding concept formation, concept use, and concept identity.

I think that the processes of concept formation and use are different in these two cases. In the case of manifest entities, concepts are formed for descriptive and classificatory purposes, via a process of abstraction that is based on the similarities between the objects falling under a concept and the differences between those objects and other objects denoted by different concepts.[16] In other words, the formation of manifest-entity-concepts goes hand in hand with the grouping of objects in classes and the discrimination between different kinds of objects (cf. Kuhn 2000, pp. 30–31, 171). Furthermore, the correct use of concepts involves learning to which objects they apply. Their stable identity, thus, derives from the stability of classifications in their domain of application.

In the case of hidden entities, concepts are formed for explanatory purposes, via various forms of abductive reasoning (e.g., analogical or model-based reasoning; see Nersessian 2008). They enable the explanation of regularities and laws via hidden entities and mechanisms, and they are formed in response to theoretical and empirical problem situations. The concept of the electron, for instance, was formed in response to the empirical problems posed by cathode rays and spectroscopy as well as the theoretical difficulties in Maxwellian

electrodynamics (Arabatzis 2006). Furthermore, the correct use of hidden-entity-concepts requires learning the theory in which they are embedded, the problem-solving practices in which they are involved, and the manifestations of their purported referents. Finally, their identity over time and across space remains a thorny issue.[17]

The characteristics of conceptual change are also different in the two cases. In the case of manifest entities, on the one hand, conceptual change is associated with reclassification. For instance, in the transition from Ptolemaic to Copernican astronomy the change in the concept of planet was associated with a reclassification of heavenly objects. The earth was now considered a planet, whereas the sun and the moon lost their planetary status. In the case of hidden entities, on the other hand, conceptual change is associated with changes in their representation and/or the reclassification of the effects attributed to them. The representation of the electron, for instance, changed dramatically from the late 19th century to the late 1920s. In cases of radical conceptual change a concept is replaced by an altogether different one. Something like this happened with the replacement of phlogiston by oxygen in late-18th-century chemistry, when some of the purported effects of phlogiston were reclassified and attributed to different entities, oxygen and hydrogen, respectively (Chang 2012).

Furthermore, the manifest/hidden distinction is important with respect to the problems posed by conceptual change for scientific rationality and scientific realism. The former problem, which concerns the possibility of genuine communication among scientists and the rational resolution of their disputes, is more easily tractable in the case of manifest-entity-concepts, because we have direct access to the individual objects classified by those concepts. This access can enable scientists who associate different concepts with the same scientific terms to overcome communication difficulties. In the case of hidden-entity-concepts, on the other hand, where no such access is possible, problems of communication among scientists who hold different concepts are more difficult to spot and resolve.

The realism problem, once again, takes a different form in the two cases. As regards manifest-entity-concepts, the problem is about the naturalness of our classifications but does not challenge the reality of the objects classified by our concepts. In other words, the question is whether the categories imposed by our concepts reflect a preexisting natural order. Furthermore, in the case of manifest-entity-concepts it is relatively straightforward to tell when their reference changes, because we have direct access to the objects that are grouped together by those concepts. Regarding hidden-entity-concepts, the realism problem is about the very existence of the entities represented by our concepts. Debates concerning the legitimacy of hidden-entity-concepts are, *ipso facto*, ontological debates about their referents. Furthermore, in the case of hidden-entity-concepts it is more difficult to tell whether an evolving concept continues to refer to the same entities.

Concluding Remarks

I have argued that scientific concepts are associated with representations of entities, properties, and processes in nature. They play essential roles in the description, classification, and explanation of phenomena. We can tell whether someone possesses a scientific concept by examining whether he or she can apply the concept correctly. I have pointed out that what underlies this ability remains a contested issue in cognitive psychology and I sketched two ways of representing concepts, put forward by Hilary Putnam and Nancy Nersessian, which can help historians and philosophers of science to make sense of conceptual stability and change. As regards conceptual change, I suggested two conditions which neutralize its unpalatable implications: preservation of salient features of a concept and stability of its referent. Finally, I argued that the distinction between manifest- and hidden-entity-concepts is important for how we study concepts. We can study manifest-entity-concepts by tracking their uses in the observable realm; and we can study hidden-entity-concepts via their theoretical and experimental lives: their embeddedness and integration within a theoretical environment, and the effects associated with them in experimental settings. The stability of the latter enables hidden-entity-concepts to preserve their identity even when they are found within different theoretical environments.

Acknowledgments

I am grateful to Kostas Kampourakis, Vasso Kindi, and Kevin McCain for their constructive and helpful comments.

Notes

1 Cf. Putnam (2004, p. 41): "to ask for the meaning of a word is to ask how it is used, and explanations of how a word is used may often involve technical knowledge of a kind ordinary speakers do not possess".
2 Furthermore, as we see below, there is another objection to the classical view, namely that it does not allow for conceptual change but only for conceptual replacement.
3 Note though that the classical view still persists in the philosophical literature; see, e.g., Kraemer (2018).
4 For an insightful criticism of these views, see Bloch-Mullins (2018).
5 For a detailed review of this literature, see Margolis and Laurence (2014); cf. also Cheon and Machery (2016).
6 For historical reconstructions of the varying representations associated with the concepts of field, electron, and gene, see, respectively, Nersessian (1984), Arabatzis (2006), and Griffiths & Stotz (2007). For an eloquent defense of the concept as use approach see Kindi (2012). Finally, for the embeddedness of scientific concepts in conceptual frameworks, see Andersen, Barker, Chen (2006).
7 Putnam included in the stereotype of 'tiger' "features as … being big-cat-like" and "having black stripes on a yellow ground" (Putnam 1975d, p 267).

8 For the systematic character of concepts and their embeddedness in conceptual frameworks, see Thagard (1990), and Andersen, Barker, Chen (2006).

9 This is a relatively neglected topic, due to the theory-oriented character of much philosophy of science. Cf., however, Arabatzis (2012b), Arabatzis and Nersessian (2015), Feest and Steinle (2016), and Steinle (2012).

10 Feyerabend completely undervalued the importance of observation and experiment for the specification of the meaning of scientific concepts. See Arabatzis (2012b, p. 152).

11 The mechanisms of concept formation in such cases can be very complex. They involve abductive reasoning, which employs, among other things, abstraction, idealization, and modeling. The most detailed and illuminating account of those mechanisms can be found in Nancy Nersessian's *Creating Scientific Concepts* (Nersessian 2008).

12 More recently, the concept of planet changed again, this time in response to new astronomical discoveries. See Brusse (2016).

13 For a history of the problem of conceptual change in philosophy of science, see Arabatzis and Kindi (2013).

14 The qualifier "seems" is meant to indicate that conceptual variation does not necessarily lead to communication difficulties. For instance, as Vasso Kindi has argued, this problem does not arise if we adopt the "concept as use" view (see Kindi 2012, pp. 31–33).

15 In the philosophical literature a distinction is drawn between observable and unobservable entities. My reasons for replacing 'unobservable' with 'hidden' are explained in Arabatzis (2012a). Suffice it to say here that the term 'hidden' carries less epistemological weight than the term 'unobservable': hidden entities could be disclosed under the appropriate circumstances (e.g., through technological innovation). Furthermore, the manifest/hidden distinction lacks ontological significance: purportedly manifest entities (e.g., the Loch Ness monster) may not exist and hidden entities (e.g., the electron) may be real.

16 It should be noted that similarity judgments can be complicated and their precise role in concept formation is still under debate in cognitive psychology. See the illuminating discussion in Bloch-Mullins (2018).

17 Note though that, despite those differences, if a hidden domain has been charted and its entities classified, then the processes of concept learning in the two cases may share similar characteristics (cf. Kuhn 2000; Andersen, Barker, and Chen 2006; and Bloch-Mullins 2018).

References

Andersen, Hanne, Peter Barker, Xiang Chen. 2006. *The Cognitive Structure of Scientific Revolutions*. Cambridge: Cambridge University Press.

Arabatzis, Theodore. 1992. "The Discovery of the Zeeman Effect: A Case Study of the Interplay between Theory and Experiment." *Studies in History and Philosophy of Science* 23: 365–388.

Arabatzis, Theodore. 2006. *Representing Electrons: A Biographical Approach to Theoretical Entities*. Chicago: University of Chicago Press.

Arabatzis, Theodore. 2007. "Conceptual Change and Scientific Realism: Facing Kuhn's Challenge." In S. Vosniadou, A. Baltas, X. Vamvakoussi (eds.), *Reframing the Conceptual Change Approach in Learning and Instruction*. Amsterdam: Elsevier, 47–62.

Arabatzis, Theodore. 2012a. "Hidden Entities and Experimental Practice: Renewing the Dialogue between History and Philosophy of Science." In S. Mauskopf and

T. M. Schmaltz (eds.), *Integrating History and Philosophy of Science: Problems and Prospects*. Dordrecht: Springer, 125–139.

Arabatzis, Theodore 2012b. "Experimentation and the Meaning of Scientific Concepts." In U. Feest and F. Steinle (eds.), *Scientific Concepts and Investigative Practice*. Berlin: De Gruyter, 149–166

Arabatzis, Theodore, and Vasso Kindi. 2013. "The Problem of Conceptual Change in the Philosophy and History of Science." In S. Vosniadou (ed.), *International Handbook of Research on Conceptual Change*. 2nd ed. New York: Routledge, 343–359.

Arabatzis, Theodore, and Nancy J. Nersessian. 2015. "Concepts Out of Theoretical Contexts." In T. Arabatzis, J. Renn, A. Simões (eds.), *Relocating the History of Science: Essays in Honor of Kostas Gavroglu*. Cham: Springer, 225–238.

Badino, Massimiliano, and Jaume Navarro. 2018. "Introduction: Ether – The Multiple Lives of a Resilient Concept." In J. Navarro (ed.), *Ether and Modernity the Recalcitrance of an Epistemic Object in the Early Twentieth Century*. Oxford: Oxford University Press, 1–13.

Bloch-Mullins, Corinne L. 2018. "Bridging the Gap Between Similarity and Causality: An Integrated Approach to Concepts." *British Journal for the Philosophy of Science* 69: 605–632.

Brigandt, Ingo. 2010. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." *Synthese* 177: 19–40.

Brusse, Carl. 2016. "Planets, Pluralism, and Conceptual Lineage." *Studies in History and Philosophy of Modern Physics* 53: 93–106.

Chalmers, Alan. 2009. *The Scientist's Atom and the Philosopher's Stone: How Science Succeeded and Philosophy Failed to Gain Knowledge of Atoms*. Dordrecht: Springer.

Chang, Hasok. 2012. *Is Water H_2O? Evidence, Realism and Pluralism*. Dordrecht: Springer.

Cheon, Hyundeuk, and Eduard Machery. 2016. "Scientific Concepts." In P. Humphreys (ed.), *The Oxford Handbook of Philosophy of Science*. Oxford: Oxford University Press, 506–523.

Davidson, Arnold I. 2001. *The Emergence of Sexuality: Historical Epistemology and the Formation of Concepts*. Cambridge, MA: Harvard University Press.

diSessa, Andrea A., and Bruce L. Sherin. 1998. "What Changes in Conceptual Change?" *International Journal of Science Education* 20: 1155–1191.

Einstein, Albert. 1922. "Ether and Relativity." In A. Einstein, *Sidelights on Relativity*. London: Methuen, 3–24.

Feest, Uljana. 2010. "Concepts as Tools in the Experimental Generation of Knowledge in Cognitive Neuropsychology." *Spontaneous Generations* 4: 173–190.

Feest, Uljana, and Friedrich Steinle. 2012 (eds.). *Scientific Concepts and Investigative Practice*, Berlin: De Gruyter.

Feest, Uljana, and Friedrich Steinle. 2016. "Experiment." In P. Humphreys (ed.), *The Oxford Handbook of Philosophy of Science*. Oxford: Oxford University Press, 274–295.

Griffiths, Paul E., and Karola Stotz. 2007. "Gene." In D. L. Hull & M. Ruse (eds.), *The Cambridge Companion to the Philosophy of Biology*. Cambridge: Cambridge University Press, 85–102.

Hacking, Ian. 1990. "Two Kinds of 'New Historicism' for Philosophers." *New Literary History* 21: 343–364.

Johnston, Mark, and Sarah-Jane Leslie. 2012. "Concepts, Analysis, Generics and the Canberra Plan." *Philosophical Perspectives* 26: 113–171.

Kindi, Vasso. 2012. "Concept as Vessel and Concept as Use." In U. Feest and F. Steinle (eds.), *Scientific Concepts and Investigative Practice*. Berlin: De Gruyter, 23–46.

Kraemer, Daniel Mark. 2018. "Philosophical Analyses of Scientific Concepts: A Critical Appraisal." *Philosophy Compass* 13: e12513.

Kuhn, Thomas S. 1957. *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Cambridge, MA: Harvard University Press.

Kuhn, Thomas S. 2000. *The Road Since Structure: Philosophical Essays, 1970–1993, with an Autobiographical Interview*. Edited by J. Conant and J. Haugeland. Chicago: The University of Chicago Press.

MacLeod, Miles. 2012. "Rethinking Scientific Concepts for Research Contexts: The Case of the Classical Gene." In U. Feest and F. Steinle (eds.), *Scientific Concepts and Investigative Practice*. Berlin: De Gruyter, 47–74.

Margolis, Eric, and Laurence, Stephen. 2014. "Concepts." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), URL = <https://plato.stanford.edu/archives/spr2014/entries/concepts/>.

Matheson, Carl A., and A. David Kline. 1988. "Is There a Significant Observational-Theoretical Distinction?" In E. D. Klemke, R. Hollinger, A. D. Kline (eds.), *Introductory Readings in the Philosophy of Science*. Revised ed. Buffalo, New York: Prometheus Books, 217–233.

Nersessian, Nancy. J. 1984. *Faraday to Einstein: Constructing Meaning in Scientific Theories*. Dordrecht: Martinus Nijhoff.

Nersessian, Nancy. J. 1992. "How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science." In R. N. Giere (ed.), *Cognitive Models of Science*, Minnesota Studies in the Philosophy of Science 15. Minneapolis: University of Minnesota Press, 3–44.

Nersessian, Nancy. J. 2008. *Creating Scientific Concepts*. Cambridge, MA: MIT Press.

Putnam, Hilary. 1975a. "What Theories Are Not." In H. Putnam (ed.), *Mathematics, Matter and Method: Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press, 215–227.

Putnam, Hilary. 1975b. "Language and Philosophy." In H. Putnam (ed.), *Mind, Language and Reality*: *Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press, 1–32.

Putnam, Hilary. 1975c. "Explanation and Reference." In H. Putnam (ed.), *Mind, Language and Reality*: *Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press, 196–214.

Putnam, Hilary. 1975d. "The Meaning of 'Meaning'." In H. Putnam (ed.), *Mind, Language and Reality*: *Philosophical Papers, Volume 2*. Cambridge: Cambridge University Press, 215–271.

Putnam, Hilary. 2004. *Ethics without Ontology*. Cambridge: Harvard University Press.

Shapere, Dudley. 1983. *Reason and the Search for Knowledge: Investigations in the Philosophy of Science*. Dordrecht: Reidel.

Steinle, Friedrich. 2005. "Experiment and Concept Formation." In P. Hájek, L. M. Valdés-Villanueva & D. Westersthal (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*. London: King's College Publications, 521–536.

Steinle, Friedrich. 2012. "Goals and Fates of Concepts: The Case of Magnetic Poles." In U. Feest and F. Steinle (eds.), *Scientific Concepts and Investigative Practice*. Berlin: De Gruyter, 105–125.

Thagard, Paul. 1990. "Concepts and Conceptual Change." *Synthese* 82: 255–274.

7

HOW CAN WE TELL SCIENCE FROM PSEUDOSCIENCE?

Stephen Law

Introduction

What is pseudoscience? Most of us intuitively class more or less the same phenomena together under the umbrella of 'pseudoscience'. Paradigm examples include astrology, Young Earth Creationism, Christian Science, feng shui, homeopathy, flat earthism, and Chinese medicine (though there are certainly some contested borderline cases: not everyone agrees about the status of Freud's psychoanalytic theories, for example). But while it seems most of us recognise pseudoscience when we see, providing an adequate philosophical definition of pseudoscience is not so easy. The aim of this chapter is to survey some of the suggestions that have been made, and to make a recommendation of my own.

Necessary and Sufficient Conditions

Asking 'What is x?' type questions is a traditional philosophical occupation. Philosophers ask: 'What is justice?', 'What is truth?', 'What is the mind?', 'What is science?' etc. Coming up with a philosophically adequate answer to such questions is often assumed to involve producing a list of *necessary and sufficient conditions*.

Some terms can easily be defined in terms of necessary and sufficient conditions. For example: something is a triangle *if and only if* (abbreviated by philosophers as *iff*.) it is a three straight-sided closed figure. Being a three straight-sided closed figure is sufficient to qualify something as a triangle. It is also a necessary condition for something to qualify as a triangle. We can similarly define vixen (something is a vixen iff. it is a female fox) and bachelor (someone is a bachelor iff. they are both unmarried and male).

However, once we switch to traditional philosophical questions such as 'What is truth?', 'What is the mind?', and so on, the task of specifying necessary and

sufficient conditions becomes much more difficult. Indeed, we often quickly run up against *counterexamples* to our proposed definitions.

In fact, this sort of difficulty can arise even when trying to pin down the necessary and sufficient conditions for something to qualify as a chair. Define a chair as an object made for sitting on for example, and philosophical critics will point to counterexamples, for example, (i) objects not made for sitting on that are nevertheless chairs (a conveniently shaped boulder, placed next to a garden table, can become a chair, despite not having been made to be sat on), and (ii) objects that are made for sitting on that are not chairs (e.g., sofas, bicycle saddles). So, our proposed definition of a chair provides neither a necessary *nor* a sufficient condition for something to qualify as a chair. Further refinements to our definition will likely still face counterexamples. For example, if, in order to deal with the sofa counterexample, we suggest that something is a chair iff. it is an object made for *one* person to sit on, the bicycle saddle counterexample still remains, and so does the stool (neither are chairs, but both are made for one person to sit on).

The philosophical activity we are engaged in here – that of trying to pin down the necessary and sufficient conditions for something to be (an) X – can of course be pursued with respect to pseudoscience. Definitions of pseudoscience are offered, but we again run up against counterexamples.

For example, we might be tempted to define pseudoscience as any attempt to explain that appeals to the supernatural. However, while a great deal of pseudoscience is indeed bound up with belief in the supernatural (Christian Science, Young Earth Creationism, and, on some versions, astrology all involve supernatural elements), this suggestion faces obvious counterexamples.

First, note that involving the supernatural is not a necessary condition of a theory qualifying as pseudoscience. Many examples of pseudoscience involve no supernatural elements – flat earthism, for example.

Second, nor is a supernatural dimension sufficient to qualify a theory as pseudoscientific. True, the supernatural is often assumed to be beyond the remit of science to investigate, but this assumption is false. The mere fact that the supernatural is supposedly unobservable is certainly no obstacle to it being scientifically investigated or the focus of a properly scientific theory. Subatomic particles are also unobservable, as is the distant past of this planet, yet both are the proper focus of scientific theories that are well-confirmed.

There's no reason in principle why belief in the effectiveness of prayer couldn't be good science. Some supernatural claims could in principle be scientifically well-supported (which is not to say they are). For example, there have been two multi-million-dollar investigations into whether petitionary prayers for heart patients has some positive medical effect (see Benson et al. 2006 and Krucoff et al. 2005). Both studies found prayer had no effect. However, they *might* have produced good scientific evidence for the effectiveness of petitionary prayer. And I take it that a hypothesis for which there's good scientific evidence is not pseudoscientific.

Let's now turn to Karl Popper. Popper introduced the so-called 'demarcation problem' of identifying what distinguishes pseudoscience from science.

Popper's suggestion, as I explain below, is that, unlike real science, pseudoscience is *unfalsifiable*.

Popper on Falsificationism and Pseudoscience

Popper was a philosopher of science particularly concerned with a notorious puzzle: *Hume's problem of induction*.

Hume's Problem of Induction

Why do we suppose the sun will rise tomorrow? Well, we have seen it rise countless times before, and so we conclude that it will very probably rise tomorrow too. This is an example of inductive reasoning. We move from premises, such as:

The sun rose on Monday.
The sun rose on Tuesday.
The sun rose on Wednesday.
The sun rose on Thursday.
The sun rose on Friday.
The sun rose on Saturday.

To a conclusion, for example:

The sun will rise on Sunday.

The mark of an inductive argument is that its premises are supposed to *support*, but not *logically entail*, the conclusion. Note that there is no logical contradiction involved in supposing the premises of this example are true while the conclusion is false. So the premises don't logically entail the conclusion. Still, we suppose the premises support the conclusion – that they provide grounds for supposing the conclusion is true.

Arguments of this sort are required to provide us with substantive knowledge of the unobserved (for substantive claims about the unobserved are never logically entailed by just statements about what has been observed). Science, in so far as it makes substantive claims about the unobserved (which is what any scientific theory does – it may predict what will happen tomorrow, for example), must then rely on inductive reasoning.

Hume famously questions whether the premises of such an inductive argument give us *any grounds at all* for supposing their conclusions are true. Hume suggests that when we reason inductively, we make an assumption: that *nature is uniform*. We assume that the local patterns we observe are likely to continue over the horizon into the unobserved portions of reality. Without that assumption, thought Hume, such reasoning is unfounded.

So how might we justify the assumption that nature is uniform? That nature is uniform is no a priori *logical* truth (there's no logical contradiction involved in supposing nature is not uniform). But neither can we justify the assumption by appeal to experience, for:

(i) we cannot *directly observe* that nature is uniform throughout (for we can observe only a small fragment of it), and:
(ii) we cannot *infer* that nature is uniform throughout on the basis of the parts we have observed (because that would itself be an as-yet-unjustified inductive argument; we would then be using induction to justify induction – a hopelessly circular justification).

Hume concludes that the assumption that nature is uniform is therefore entirely unjustified and that consequently *inductive arguments fail to provide any justification whatsoever for their conclusions.* We seem forced to accept a very radical scepticism about the unobserved, a scepticism that renders all scientific theories entirely unjustified. If Hume is right, it's as reasonable to believe the sun will rise tomorrow, given what we have observed, as it is to suppose that an enormous bowl of cherries will appear over the horizon instead (see Hume 1739, Book 1, part iii, section 6)

Falsificationism as a Solution to the Problem of Induction

Popper's solution to Hume's problem is ingenious. Rather than attempting to show that we are justified in believing our current scientific theories are true, Popper accepts that we are *not* justified, but *this does not matter.*

According to Popper, science progresses by theories being put forward and then tested. For example, having noticed that this object fell when released, I may develop the theory that all objects fall when released. I can now test this theory, dropping pens, bricks, feathers and so on. If any object fails to fall when released, that logically entails my theory is false. Such an observation *falsifies* my theory. Popper maintains that science progresses, not by theories being inductively confirmed, but by theories being falsified. We may not be justified in supposing our current, unfalsified theories are true, but it's reasonable for us to prefer those theories to those that have been falsified (Popper 2002, p. 72–73).

However, this is not to say that, on Popper's view, all unfalsified theories are equally preferable. Popper maintains we should prefer those unfalsified theories that are more falsifiable over those that are less falsifiable.

A theory can be more falsifiable by being *clearly stated*, preferably in terms mathematically quantifiable and measureable. The theory that all adult dogs are 'heavy-ish' is a vague claim that can be easily protected from falsification by insisting that what the term means has been misunderstood ('No, Chihuahuas *are* heavy-ish!'). The hypothesis that all adult dogs weigh over 3kg, on the other hand, can be straightforwardly falsified with the aid of a scale.

A theory can also be more falsifiable by being *wider ranging*. The theory that all adult dogs weigh more than 3kg is more falsifiable than the hypothesis that all adult Dalmatians weigh more than 3kg because any observation that falsifies the latter will falsify the former whereas the reverse is not true.

On Popper's view, scientists should work with the most falsifi*able* unfalsifi*ed* theories.

Popper on Pseudoscience

Popper suggests that *what distinguishes science from pseudoscience is the fact that the former is falsifiable, whereas the latter is not*.

Popper illustrates his criterion using the examples from Marx, Freud, and Adler. Popper considered both Marx's theory of history and the psychoanalytic theories of Freud and Adler unfalsifiable, but for different reasons. Marx's theory of history is a supposedly scientific theory about how the history of human society unfolds, marked by revolutionary transitions, while Freud and Adler's theories are supposedly scientific theories about our unconscious motivations.

Popper thought the problem with Freud and Adler's psychoanalytic theories is that, whatever human behaviour is observed, it can always be interpreted in such a way as to 'fit' either theory. Popper illustrated this point by considering two hypothetical situations – one in which a man pushes a child into water with the intention of drowning it, the other in which a man sacrifices himself to save a child. Popper claimed each event can easily be explained in either Freudian and Adlerian terms:

> According to Freud the first man suffered from repression (say, of some component of his Oedipus complex), while the second man had achieved sublimation. According to Adler the first man suffered from feelings of inferiority (producing perhaps the need to prove to himself that he dared to commit some crime), and so did the second man (whose need was to prove to himself that he dared to rescue the child). *(2002, p. 46)*

Indeed, Popper found he couldn't think of any human behaviour that couldn't be made to fit either theory.

Popper concluded that both Freud's and Adler's theories were unfalsifiable, and for the same reason. Popper believed that Marx's theory of history was also unfalsifiable, though for a different reason. On Popper's view, Marx's theory – unlike Freud's and Adler's – *started out* as a falsifiable theory. Indeed, it made some risky predictions about how history would unfold. It predicted a revolution would happen in an industrially advanced society such as Britain. However, Marx's prediction turned out to be incorrect (there was a revolution, but not in the way Marx predicted – for example, it occurred in industrially backward Russia). Marx's theory was therefore falsified. However, rather than accept this, Marx's followers

adopted an immunizing strategy, re-interpreting theory and evidence so that the theory continued to fit the evidence after all.

So, Popper frowns on unfalsifiability, considering it sufficient to qualify supposedly 'scientific' theories such as those of Marx, Adler, and Freud as pseudoscience. However, he notes that a theory's failure to be falsifiable can be achieved in different ways. Marx's theory started out falsifiable, was falsified, but was then rendered unfalsifiable by the immunising strategy adopted by its adherents. Freud's and Adler's theories, on the other hand, started out as unfalsifiable – something about the way these theories were *initially* constructed and/or applied ensured that observation could never count against them.

Popper notes that, to those wedded to unfalsifiable belief systems, their truth can appear manifest. Indeed, so obvious is their truth, adherents suppose, that those who fail to recognise it must be suffering from something akin to a *perceptual defect*. The suggestion that the unbelievers are somehow blinded to the manifest truth crops up in a great deal of pseudoscientific thinking. For example, Young Earth Creationist Ken Ham says about those who reject the Bible-literalist account of creation:

> Why can't the humanists, the evolutionists, see that all the evidence supports exactly what the Bible says? It is because they do not want to see it. It is not because the evidence is not there. They refuse to allow the evidence to be correctly interpreted in the light of biblical teaching. *(2012, p. 76)*

The inability of folk to recognise the truth of flat Earthism is often explained as a result of a global conspiracy to hide the truth.

Criticism of Popper's Account

I think there is considerable insight in Popper's thinking about pseudoscience. However, it is flawed. Below are a couple of criticisms.

First, falsificationism appears incorrect as an account of how science progresses. Here's one example. On Popper's view, what really counts in favour of a scientific theory is that it makes risky predictions that turn out to be true. However, a theory might be well confirmed even if it doesn't predict much. To illustrate, consider the theory of evolution, on which new species evolve over time. While that theory does predict a great deal, *one very strong piece of evidence in its favour did not arise from a prediction, let alone a risky one.*

Whales are occasionally discovered with vestigial limbs. That's because whales are mammals that evolved from earlier land-dwelling creatures possessing limbs. The existence of such vestigial limbs was not *predicted* by the theory of evolution: the theory does not say such limbs are likely to be found on whales (to commit yourself to the theory of evolution is not to suppose the existence of such limbs is probable). Still, the discovery of whales with vestigial limbs strongly confirmed

the theory of evolution because, while such limbs may not be probable on that theory, they are still far, more probable on that theory than on alternatives such as Bible literalism. Yet on Popper's account, not only are theories never be confirmed to the extent we can reasonably suppose they are true (Popper accepts Hume's radical scepticism about the unobserved), a theory can *only* be confirmed by way of its making a *risky prediction*.[1] That last claim is incorrect.

Secondly, Laudan notes that levelling the charge of unfalsifiability against a doctrine like Young Earth Creationism 'egregiously confuses doctrines with the proponents of those doctrines' (1983, p. 17). The theory that the universe is just a few thousand years old *is*, in fact, falsified. The fact that followers of the theory remain unwavering in their commitment to it may reveal something about the psychology of its devotees, but it's a fact about them that stands quite independently of whether or not the theory itself is falsifiable/falsified, which it is.

In short, on Laudan's view, in trying to demarcate pseudoscience, Popper mistakenly focuses on the product (on the theories), whereas the real fault lies with the producers – those who promote and defend those theories.

Actually, even Popper acknowledges that sometimes the fault is not with the theory *per se* but rather with the manner in which its proponents defend it. As we noted already, Popper thought Marx's theory of history was falsifiable, and was indeed falsified. What turned Marx's theory into pseudoscience, on Popper's view, was the manner in which it was subsequently defended.

So, in trying to demarcate pseudoscience from science, where should our focus be? On the content of the theories, or on the manner in which those theories are defended and/or supported by their proponents? Even Popper acknowledges our focus should sometimes be on the latter.

However, as Maarten Boudry (2013, p. 91) points out, it's often difficult to tell where a theory ends and the obfuscations of its defenders begin. In the case of Young Earth Creationism, the fault does appear to lie largely with its defenders. However, in other cases – such as Christian Science (see later) – the method that renders the theory pseudoscientific appears to be integral to the theory itself. Christian Science just *is*, in part, a dodgy method.

So, we have now looked at two suggestions regarding how pseudoscience should be defined. We have seen that pseudoscientific theories need not involve the supernatural. We have now also seen that Popper's suggestion that what marks out pseudoscience is its unfalsifiability also runs into difficulties. In particular, failure to make a risky prediction – and thus to be falsifiable – does not necessarily prevent a theory from being scientifically well confirmed. And I take it that any scientifically well-confirmed theory is not pseudoscience.

Family Resemblance

Given the difficulties involved in providing a watertight definition of 'pseudoscience' in terms of necessary and sufficient conditions, some may pessimistically

conclude that the term is, then, just a 'hollow phrase' – a term used by speakers to express nothing more than their disapproval of a theory. Rather than pick out some objective property or kind, the use of 'pseudoscience' is more like our use of 'weed', which does not mark any objective difference between the plants growing in our gardens, but merely reflects our own personal preference as to what we like to see growing there. The philosopher Larry Laudan (1983) famously came to just this conclusion about 'pseudoscience' in part because he noted the kind of difficulties I have outlined above in terms of providing necessary and sufficient conditions.

I am less pessimistic about the term 'pseudoscience'. Certainly, our inability to provide a watertight definition of terms like 'truth', 'the mind', or even 'science', in terms of necessary and sufficient conditions does not, by itself, justify as us in condemning *those* expressions as empty. So why be so quick to condemn the term 'pseudoscience' on that basis?

Perhaps our difficulty in providing a watertight philosophical definition of pseudoscience is due to the concept being what Wittgenstein (1958) terms a *family resemblance concept*?

Wittgenstein remarks that some concepts, such as that of a game, have a 'family resemblance' character. The various members of a family may resemble each other to differing degrees, despite there being no one feature (the big nose, the bushy eyebrows, the lopsided mouth) that they all share. Wittgenstein notes that there appears, similarly, to be no one feature that all games have in common. Some games are competitive, but some are not. Some involve balls, but some do not. And so on. So what is *the one thing that all and only the games have in common?* There need not be anything – just a series of overlapping similarities. Yet the concept of a game is legitimate all the same. Our inability to provide a definition of the sort that we can provide for 'triangle' or 'vixen' – in terms of necessary and sufficient conditions – should not lead us to abandon our use of the term 'game'.

Of course, a critic might argue that even if there's no one thing all and only games have in common, still, it should be possible to specify a precise algorithm that determines what is a game and what is not.

For example, such an algorithm might require that for something to be a game, at least three out of a set of six characteristics must be possessed (but *any* three, so there need be no one characteristic all the games share). Not that such an algorithm *would* provide a necessary and sufficient condition for something to qualify as a game (the condition being that at least three of the six characteristics are possessed).

However, Wittgenstein thought even such an algorithm was unspecifiable when it comes to games. The use of the term 'game', thought Wittgenstein, is such that it is not everywhere marked by sharp boundaries, or even by any boundary at all. Therefore, any attempt to provide a definition in terms of necessary and sufficient conditions must fail – it will involve drawing boundaries where none exist. Yet, for all that, 'game' is a perfectly serviceable term.

Pseudoscience might, similarly, be a family resemblance concept. The suggestion that it is a family resemblance concept has been made by a number of philosophers including Massimo Pigliucci (2013).

Examples of Pseudoscience

To finish, I want briefly to examine a couple of examples of pseudoscience in order to identify certain criteria that I think should probably be included – even only if in a family resemblance way – in any adequate account of what constitutes pseudoscience.

Example 1: Young Earth Creationism

Young Earth Creationists are Bible literalists who believe the entire universe is around 6,000 years old. They also believe God made all species in the week of creation, and that no new species can or has spontaneously evolved. This theory is, of course, false, and faces a mountain of counter-evidence. This includes, for example:

> *The fossil record*. It reveals that new species have evolved.
>
> *The light from distant stars.* Given the speed at which light travels, the universe would have to be much older than six thousand years for the light to have reached us.
>
> *The white cliffs of Dover.* The chalk cliffs are made from the calcium-rich shells of tiny organisms that lived in the sea. It would take much longer than six thousand years for that depth of material to accumulate.

How do Young Earth Creationists typically respond to such evidence? They *explain it away.* The fossil record, for example, is usually explained as a result of the Biblical flood. The sedimentary layers we see in the ground containing fossils were laid down very recently, mostly during the flood, which also drowned many creatures and trapped their bodies in the layers. The ordering of the fossils within the layers is explained as a result of different ecological zones being flooded at different times. Humans appear in only the top most layers because they managed, through their intelligence, to avoid drowning until late in the deluge. The light from distant stars has been explained by appeal to a 'time dilation'. The white cliffs of Dover have been explained as a result of Noah's Flood producing vast blooms of microorganisms. Of course, each of these explanations faces its own evidential challenges, and so further explanations can be and are developed *ad nauseum* to deal with them.[2]

The strategy employed above exploits the more general point that *any theory, no matter how outlandish can, with sufficient ingenuity, always be made consistent with – be made to 'fit' – the available evidence.*

Suppose, for example, that I believe dogs are spies from the planet Venus planning an imminent invasion. There is a mountain of evidence against my

belief – evidence that dogs aren't smart, lack language, that Venus is uninhabitable by dogs, etc. However, that evidence can be explained away. I might maintain dogs hide their intelligence and linguistic abilities from us, and they live on Venus in deep underground bunkers that protect them from the harsh Venusian atmosphere, and so on. By endlessly explaining away such evidence, 'fit' can always be achieved.

Now achieving 'fit' in this manner can be misleadingly packaged as doing genuine science. For isn't science all about developing theories that 'fit' the evidence? And hasn't our Creationist shown that their theory can be made to 'fit' the evidence? The emphasis on achieving 'fit' leads proponents of Young Earth Creationism like Ken Ham to conclude it is, indeed, scientific.

> Increasing numbers of scientists are realizing that when you take the Bible as your basis and build your models of science and history upon it, all the evidence from the living animals and plants, the fossils, and the cultures *fits*. This confirms that the Bible really is the Word of God and can be trusted totally. (My italics)[3]

According to Ham, Young Earth Creationists and evolutionists do the same thing: they take the evidence, and then look for ways to make it fit the axioms of the framework theory to which they have already committed themselves:

> Evolutionists have their own framework … into which they try to *fit* the data. (my italics)[4]

This strategy, which I have previously dubbed 'But it Fits!' (Law, 2011), often crops up in pseudoscientific thinking. One of the obvious problems with it, of course, is that it conflates achieving *consistency with* the evidence with being *confirmed by* that evidence. Any theory, no matter how absurd – even the theory that dogs are Venusian spies – can be made consistent with the evidence. That's not to say it's confirmed by that evidence.

So yes, Young Earth Creationists may be able to show their theory is consistent with the evidence. That's not say their theory is confirmed by that evidence. On the contrary, it's the theory of evolution that is strongly confirmed by the evidence.

Am I suggesting that heavy reliance on the 'But it Fits!' strategy in response to counter evidence is a necessary and/or sufficient condition for a theory, as defended by certain followers, to qualify as pseudoscience? No. Adoption of the 'But it Fits!' strategy is not a *sufficient* condition. To see why it is not sufficient, note that much the same strategy is also employed by other suspect belief systems that are not usually considered pseudoscientific.

For example, consider conspiracy theories (of course, I want to acknowledge that *some* conspiracy theories are both reasonably held and true e.g. Watergate and Iran/Contra). Evidence against a conspiracy theory is often explained away by expanding the scope of the conspiracy to take care of it. For example,

experts challenging the conspiracy theory may subsequently be accused of being implicated in it. Some conclude that because their particular conspiracy theory can, by such means, be shown to be consistent with the evidence, it must be at least as reasonable, given that evidence, as alternatives.

Yet, by adopting the 'But it Fits!' strategy, a conspiracy theorist does not thereby qualify as a pseudoscientist. Adoption of 'But it Fits!' is not *sufficient* to qualify a belief system as pseudoscientific. Nor, as we are about to see, is adoption of 'But it Fits!' a *necessary* condition of pseudoscience. Consider Christian Science, outlined here.

Example 2: Christian Science

Christian Science – a religious movement started in 19th Century New England by Mary Baker Eddy – holds that disease is an illusion: a product of the mind. Practitioners believe illness can be healed through prayer using the exact method and means by which Jesus healed.

How do Christian Scientists know their method works? Because of the supposedly 'scientific' evidence they have amassed in its support over many decades. Tens of thousands of testimonies have been published by Christian Scientists of cases in which their methods have been applied and people have subsequently recovered (see Fraser 1999 for more detail).

Christian Science is pseudoscience but differs from the version of Young Earth Creationism outlined earlier in that *very little attention is given to explaining away the evidence against it*.

Instead of focussing on the 'misses' – cases where Christian Science was applied and failed – Christian Scientists focus almost entirely on recording the 'hits' – on cases where the method was applied and the subject recovered. This approach, which I have elsewhere dubbed *Piling Up The Anecdotes* (Law, 2011) is deemed 'scientific' because, in the minds of Christian Science's followers, a huge amount of data has been built up in support of their theory.

Christian Science is pseudoscience, and indeed, like Young Earth Creationism, it apes the methods of genuine science. However, it makes little use of the 'But it Fits!' strategy. So, reliance on 'But it Fits!' is not a necessary condition of pseudoscience.

Science-Like Features

Is a heavy reliance on Piling Up The Anecdotes *sufficient* to qualify any belief as pseudoscientific? I am not sure it is. Many absurd beliefs similarly rely heavily on the use of anecdotal evidence. The widespread belief that some people are psychic, for example, is almost always justified by appeal to anecdotal evidence. So too is the belief that ghosts are real. Yet I would be disinclined to class every belief in ghosts as pseudoscientific.

However, belief in ghosts certainly can *become* pseudoscientific once adherents start to use ghosts-detecting devices, start more systematically amassing anecdotal evidence and calling it 'scientific', and start developing a kind of ghost mechanics (involving e.g. ectoplasm, orbs, etc.). So, perhaps what is *also* required for a belief or theory to qualify as pseudoscientific is that it *exhibit certain further science-like features.* I will return to this suggestion shortly.

Incidentally, I don't claim that defenders of pseudoscience must employ one or both of 'But it Fits!' and/or Piling Up The Anecdotes. These are two strategies that may be employed in defence and/or support of dubious belief systems. However, other unreliable strategies are also available.

Note, for example, that defenders of pseudoscience often suggest their theory provides the *best available explanation* of what is observed. Argument to the best explanation is quite properly used in science (we suppose it is reasonable to believe in some unobserved phenomena, such as the Big Bang, or electrons, because it provides the best available explanation of what is observed). However, argument to the best explanation is also regularly abused by believers in woo. Those who believe in alien visitation and abduction, miracles, fairies, and even the Resurrection, will often challenge sceptics by saying, 'Explain *that*!', and then note how their opponents struggle to provide an explanation. They conclude their belief provides the best available explanation of what's observed.

Notice that *anything* we observe can be neatly and easily explained by positing a hidden agent with extraordinary powers and a desire to bring it about. Can't explain how your keys – which you are sure were on the sofa – ended up on the table? I can: there are mischievous gremlins in your house who enjoy playing such tricks. Can't explain why the flowers grow? I can: fairies imbued with magical powers force the flowers to bloom each spring. Can't explain Jesus's empty tomb and the testimony of eyewitnesses to a risen Christ? I can: God raised Jesus from the dead. Can't explain why the Twin Towers came down vertically like that, with little puffs of smoke appearing just below the collapsing structure? I can: 9/11 was a controlled demolition by secret, Government-backed forces. These aren't sound applications of argument to the best explanation[5], but they can appear so, particularly in the eyes of believers. We might call employing such poor examples of argument to the best explanation *'Explain THAT!'*

What these three approaches to justifying beliefs have in common is of course that they are unreliable methods so far as arriving at true beliefs is concerned. 'But it Fits!', Piling Up The Anecdotes, and 'Explain THAT!' are not truth-conducive in the way that genuinely rational approaches to justifying beliefs are.

Pseudoscience and Bullshit

Here is my proposed characterisation of pseudoscience. First, I suggest that proponents of pseudoscientific belief systems exhibit certain unreliable approaches to supporting and/or defending the theory, approaches such as 'But it Fits!', Piling

Up The Anecdotes, and 'Explain THAT!', that nevertheless *appear* reliable in the eyes of their proponents. However, as we have seen, it's not just proponents of pseudoscience that employ such approaches – 9/11 Truthers and ordinary believers in ghosts and psychic powers do the same thing.

Second, in addition to relying on strategies such as 'But it Fits!', Piling Up The Anecdotes, and/or 'Explain THAT!', pseudoscience *also* typically involves one or more of the following:

(i) a *claimed* scientific or science-like *methodology* that is, in reality, highly suspect (both Young Earth Creationism and Christian Science claim to be using the scientific method, though their actual methods are highly suspect). Proponents of pseudoscience don't just employ unreliable argumentative methods in supporting and/or defending their belief. They maintain they are systematically employing a methodology that is rigorous and perhaps even properly scientific.
(ii) a dubious form of science-like *mechanics* – positing various forces, entities, fields, stuffs, and/or channels, which can be blocked, enhanced, and so on (Feng Shui and Chinese Medicine do this for example, and so do the Marxist theory of history and the psychoanalytic theories).

I don't claim this characterisation is perfect. You may be able to think of counter-examples. But I suggest it provides a pretty good characterisation of pseudoscience.

My characterisation explains why non-scientific disciplines such as history are not pseudoscience: historians don't (usually) claim to be employing a scientific or science-like method, and what methods they do employ are for the most part, fairly reliable.

It also explains why fraudulent science is not (or is not necessarily) pseudoscience. A scientist who has falsified data to get the result he wants is not practising pseudoscience – his professed methods are genuinely scientific. He has just fiddled the figures.

My suggested characterisation explains why other dubious belief systems – such as 9/11 conspiracy theories, mainstream belief in ghosts and psychic powers, theodicies – aren't pseudoscience, despite their shared reliance on the use of strategies such as 'But it Fits!', Piling Up The Anecdotes, and 'Explain THAT!' For these other belief systems don't usually involve any claimed scientific or science-like method or invoke any dubious science-like mechanics.

Moreover, in so far as those belief systems do start to involve (i) and (ii), they start to look like pseudoscience. Ghost-hunters who call their anecdotal evidence 'scientific' and who build ghost-detecting gizmos designed to detect ectoplasm are engaged in pseudoscience.

My account also explains why some forms of climate change denialism are pseudoscience while others are not. Climate change deniers who justify their denial by appeal to a bunch of anecdotes about places that are colder than average

are not pseudoscientists, but fools. Climate change deniers that fiddle the data to get the result they want are guilty of science fraud, not pseudoscience. However, climate change deniers who claim to be employing the scientific method but who are actually employing a dodgy method are engaged in pseudoscience.

Let's call belief systems that make heavy use of strategies such as 'But it Fits!', Piling Up The Anecdotes, and 'Explain THAT!' *bullshit belief systems*. My suggestion is that pseudoscientific belief systems are a variety of bullshit belief system. What further distinguishes those bullshit belief systems that are pseudoscientific is the fact that they exhibit further science-like features. So, pseudoscience is a subcategory of bullshit belief systems more generally, and indeed gradually shades into other varieties of bullshit.

I am not the first philosopher to link pseudoscience and bullshit. Philosopher Harry Frankfurt (2005) famously defined bullshitters as unconcerned with the truth of what they claim. Bullshitters are not liars, deliberately asserting untruths. Rather, they just don't care whether what they say is true or not. James Ladyman suggests that pseudoscience is a variety of bullshit:

> As a first approximation, we may say that pseudoscience is to science as science fraud is to bullshit. … This is only a first approximation because we usually assume that bullshitters know what they are doing whereas, as pointed out above, many pseudoscientists are apparently genuinely seeking the truth. Just because one's first-order representations are that one is sincerely seeking truth, it may be argued that, in a deeper sense, one does not care about it because one does not heed to the evidence. A certain amount of self-deception on the part of its advocates explains how pseudoscience is often disconnected from a search for the truth, even though its adherents think otherwise. This is important because it means that what makes an activity connected or disconnected to the truth depends on more than the individual intentions of its practitioners. *(2013, p. 52–53)*

Ladyman also suggests that pseudoscience 'involves some kind of emulation of science or some of its characteristics or appearance' (2013, p. 52).

Clearly, I follow Ladyman in supposing that pseudoscience is a variety of bullshit belief, and also in suggesting that what further distinguishes pseudoscience is further science-like features. However, I do not endorse Ladyman's suggestion that what bullshit and pseudoscience tend to have in common is the fact that adherents collectively, in some sense, *don't care about the truth*.

I don't endorse Frankfurt's characterization of bullshit. I think Frankfurt accurately captures a certain sort of bullshitter – for example, the person who, unconcerned with whether or not something is true, nevertheless says it for effect, to self-aggrandise, or persuade, or whatever. However, a great deal of what goes by the title 'bullshit' – in particular, bullshit belief *systems* – are promoted and defended by folk who are, both individually and collectively, desperately concerned with the truth.

Religious and cult belief systems are often bullshit, but to say about those religious and cult communities that at some level they don't care whether what they believe is true is quite a stretch. Indeed, adherents of such belief systems will sometimes stake their own lives, and also the lives of those they love, on the truth of their beliefs. Consider those cultists who commit mass suicide (those that drank the Kool-Aid at Jonestown, for example). Consider the Christian Scientists who have killed their sick children by rejecting conventional medicine and relying on prayer instead. Consider antivaxxer parents. To describe such belief communities as being, in any sense, unconcerned with whether what they believe is true strikes me as, at best, dubious.

Ladyman suggests that such belief communities are not sincerely seeking truth because they 'don't heed to the evidence'. True, a lack of concern with truth would explain why they don't properly heed to the evidence. However, that's not the only explanation available. Such communities may just be deeply ignorant about what truth-conducive methods and heeding to the evidence actually involves. Thus, no matter how sincerely they seek the truth, they may still fail properly to heed to the evidence.

So I suggest the case for saying that, in some sense, such communities aren't sincerely seeking the truth is not well made. We would do better to define bullshit, and also pseudoscience, just in terms of a heavy reliance on methods of justification that are not in fact truth-conducive but can appear so, irrespective of whether or not those employing such methods happen, at any level, and either individually or collectively, to care about whether what they believe is true.

Conclusion

It seems to me that pseudoscience is a legitimate and indeed useful concept. Terms like 'pseudoscience' and 'bullshit' are helpful because they allow us to class together beliefs and belief systems that are distinguished by particularly seductive and common forms of irrationality – forms of irrationality that we need to be on our guard against.

I have here provided (i) a ball-park characterisation of pseudoscience that stresses that pseudoscience is a subcategory of bullshit belief systems more generally, (ii) provided an account of what makes them bullshit belief systems (and rejected Ladyman's and Frankfurt's suggestion), and (iii) provided an account of what distinguishes pseudoscience from other varieties of bullshit.

Notes

1 According to Popper,

> confirmations should only count if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an

> event which was incompatible with the theory – an event which would have refuted the theory.
>
> (2002, p. 47–48)

Note that, given Popper accepts the conclusion of Hume's sceptical argument regarding induction, what Popper here terms a confirmation should not be understood as making the theory probably, or at least more probably, true.

2 The Answers in Genesis website www.answersingenesis.org offers many such explanations.

3 Answers in Genesis website https://answersingenesis.org/who-is-god/creator-god/the-root-of-the-problem/

4 Answers in Genesis website https://answersingenesis.org/human-evolution/neanderthal/what-about-the-neandertal-dna/

5 Why do I say these aren't sound applications of argument to the best explanation? One reason is that they involve a move from: (i) we are justifiably unwilling to endorse any of available mundane explanations for x (because each considered individually seems improbable), to: (ii) we should accept the extraordinary explanation of x as the best available. This move is unjustified.

Consider, for example, a standard method of arguing for the Resurrection. Given the supposed fact that, when we consider each of more obvious mundane explanations for the empty tomb and reports of the risen Christ, we justifiably assign each a low probability (it seems improbable all the witnesses are hallucinating; it seems improbable they are all lying, etc.), it's then suggested that the Resurrection is the best available explanation and should be accepted. In fact, even if sceptics are justifiably unwilling to accept any of the mundane explanations on offer as being correct, they may still justifiably consider it far more probable that one of those mundane explanations, or another explanation they have not yet thought, is correct than that Christ rose from the dead.

References

Benson, Herbert et al (2006) 'Study of the Therapeutic Effects of Intercessory Prayer (STEP) in Cardiac Bypass Patients: A Multicenter Randomized Trial of Uncertainty and Certainty of Receiving Intercessory Prayer'. *American Heart Journal*, 151: 934–42.

Boudry, Maarten (2013) 'Loki's Wager and Laudan's Error', in: M. Pigliucci and M. Boudry (eds.), *Philosophy of Pseudoscience* Chicago: University of Chicago Press. 79–98.

Frankfurt, Harry (2005) *On Bullshit*. Princeton: Princeton University Press.

Fraser, Caroline (1999) *God's Perfect Child: Living and Dying in the Christian Science Church*. New York: Metropolitan Books.

Ham, Ken (2012) *The Lie: Evolution/Millions of Years*. Green Forest, AR: New Leaf.

Hume, David (1739) *A Treatise on Human Nature*. Oxford: Oxford University Press.

Krucoff, M.W. Crater S.W., Gallup D. et al. (2005) 'Music, Imagery, Touch, and Prayer as Adjuncts to Interventional Cardiac Care: The Monitoring and Actualization of Noetic Trainings (MANTRA) II Randomized Study'. *Lancet* 366. 211–17.

Ladyman, James (2013), 'Toward a Demarcation of Science from Pseudoscience', in: M. Pigliucci and M. Boudry (eds.), *Philosophy of Pseudoscience*. Chicago: University of Chicago Press. 45–59.

Laudan, Larry (1983) 'The Demise of The Demarcation Problem', in: R.S. Cohen and L. Laudan (eds.), *Physics, Philosophy, and Psychoanalysis*. Dordrecht: D. Reidel. 111–27.

Law, Stephen (2011) *Believing Bullshit*. Amherst, NY: Prometheus.
Pigliucci, Massimo, (2013) 'The Demarcation Problem', in: M. Pigliucci and M. Boudry (eds.), *Philosophy of Pseudoscience*. Chicago: University of Chicago Press. 9–28.
Popper, Karl (2002) *Conjectures and Refutations*, 2nd Edition. London: Routledge Classics.
Wittgenstein, Ludwig (1958) *Philosophical Investigations*. Oxford, UK: Blackwell.

8

HOW DO WE KNOW THAT 2 + 2 = 4?

Carrie S. I. Jenkins

Introduction

This is a survey chapter about issues in the epistemology of elementary arithmetic. Given the title of this volume, it is worth noting right at the outset that the classification of arithmetic as science is itself philosophically debatable, and that this debate overlaps with debates about the epistemology of arithmetic.

It is also important to note that a survey chapter should not be mistaken for a comprehensive, definitive, or unbiased introduction to all that is important about its topic. It is rather an exercise in curation: a selection of material is prepared for display, and the selection process is influenced not only by the author's personal opinions as to what is interesting and/or worthy, but also by various contingencies of her training, and my survey reflects my training in Anglo-American analytic philosophy of mathematics.

Although I'm surveying an area of epistemology, I will classify approaches by metaphysical outlook. The reason for this is that the epistemology and metaphysics of arithmetic are so intimately intertwined that I have generally found it difficult to understand the shape of the epistemological terrain except by reference to the corresponding metaphysical landmarks. For instance, it makes little sense to say that arithmetical knowledge is a kind of "maker's knowledge" unless arithmetic is in some way mind-dependent, or to classify it as a subspecies of logical knowledge unless arithmetical truth is a species of logical truth.

I will be discussing $2 + 2 = 4$ as an easily-graspable example of an elementary arithmetical truth, our knowledge of which stands in need of philosophical explanation. While some of the surveyed approaches to this explanatory demand proceed by *rejecting* the presumed explanandum – that is, by denying that $2 + 2 = 4$ is known (or even true) – for clarity and ease of expression I will proceed as if $2 + 2 = 4$ is a known truth except when discussing these approaches.

The rest of this chapter proceeds as follows. In the next section, I identify two key challenges for an epistemology of simple arithmetic, and then adduce two constraints on what should count as a successful response. Next, I discuss ways of addressing these challenges, grouped according to their corresponding metaphysical outlook. The subsequent sections survey non-reductive Platonist approaches, look at reductions (often better labelled "identifications"), and consider an array of anti-realist strategies. I conclude with a brief summary, returning to the question of arithmetic's status as science.

Epistemological Challenges

Challenge 1: Abstractness

In this first challenge, we can see immediately how intimately metaphysics is involved in the epistemology of arithmetic. The *locus classicus* for the challenge is 'Benacerraf's dilemma' (Benacerraf 1973). Paul Benacerraf drew attention to two points which together generate the dilemma. The first originates in semantics, and specifically in the fact that arithmetic appears to be *about* objects (numbers). When we talk about arithmetic we appear to refer to numbers, ascribe properties to them, quantify over them, and so on. The second is that if we are to *know* truths about such objects, we must be in some kind of contact with them, and Benacerraf interpreted this as a requirement of *causal* contact. Indeed, he endorsed a causal account of knowledge in general. Such accounts of knowledge are now widely thought to be untenable, but a more sophisticated version of Benacerraf's challenge, developed by Hartry Field (1989), requires only that we be able somehow to explain how we manage to be *reliable* concerning the arithmetical domain, or say how it is that we have mostly true beliefs about them without positing a large-scale coincidence.

Why is this such a big deal? Because of the (prima facie) metaphysics of arithmetic. Arithmetical objects like the numbers 2 and 4, if such things exist, are presumably *abstract objects*. They are not located in space and time; they are not concrete things that we can touch or hear or put under our microscopes; we appear to have no physical contact with them at all. So how do we learn about them? Whether we interpret this question as Benecerraf's demand for a causal account, as Field's demand for an explanation of reliability, or in some other way, it is one of the central challenges in the epistemology of arithmetic.

Challenge 2: A Prioricity

This second challenge is more purely epistemological. We seem to know $2 + 2 = 4$ without *needing* to touch or hear numbers or put them under our microscopes. Indeed, we seem to know it without relying on any kind of empirical evidence at all. That is to say, arithmetical truths appear to be knowable *a priori*. In a nutshell,

the second challenge is simply: *what's up with that?* Or: how is *a priori* knowledge possible?

Other kinds of truths also appear to be knowable *a priori*, such as those of logic and set theory. The same challenge is faced by epistemologists in these areas, and there may be some methodological reasons for preferring a unified account of the *a priori* which accommodates all these cases. However, the simplicity and near-universal knowability of basic arithmetical truths like *2 + 2 = 4* can make puzzlement about such cases feel all the more urgent, and (as we shall see in a moment) especially challenging.

Constraint i: Applicability

Elementary arithmetic is universally applicable. The breadth of its applicability is even more impressive than that of geometry, which requires some – at least hypothetical – space to describe. By contrast, *2 + 2 = 4* is as applicable to poems and headaches as it is to apples and pebbles.

Explaining the applicability of truths about an apparently abstract domain (of numbers) to the physical world (of apples and pebbles) is itself an interesting challenge, but a metaphysical one. For the purposes of epistemology we don't have to provide such an explanation, but we must *leave space* for one. That is to say, whatever story we end up telling about our knowledge of 2 + 2 = 4 had better be one that is compatible with the universal applicability of arithmetic.

Constraint ii: Non-Specialist Knowledge

A significant constraint, albeit one at risk of being overlooked in some accounts (as we see below) is that almost everyone knows that *2 + 2 = 4*, and moreover knows it in a way that appears to differ substantively from ordinary empirical knowledge. The challenges of abstractness and *a prioricity* are just as applicable to this non-specialist knowledge as to the knowledge possessed by mathematicians or philosophers. Accounts of arithmetical knowledge which depend on specific expertise – say, on proofs from axioms that *almost nobody* ever considers – cannot meet this constraint without some fancy footwork. Simple fixes are unpromising. One could claim, for example, that non-specialists' knowledge that *2 + 2 = 4* is derivative knowledge, reliant on the work done by the mathematical experts. But this is contrary to common sense, which tells us that non-specialists are quite capable of knowing *2 + 2 = 4* without the help of mathematicians, as well as that this non-specialist knowledge is *a priori*, which is almost[1] universally regarded as ruling out reliance on someone else's testimony.

The significance of this constraint is part of the reason for this chapter's title. Not only did I deliberately focus on a very widely-known arithmetical truth, I also worded the title in such a way as to invite more explicit reflection on who "we" are: who falls within the scope of the relevant epistemological enquiry?

Some theories of arithmetical knowledge may only be applicable to certain kinds of "we"s. This is not invariably a failing; a philosopher might set out to provide an account only of how *expert mathematicians* know $2 + 2 = 4$, or only of how *expert philosophers of mathematics* know $2 + 2 = 4$. Such accounts do not meet constraint *(ii)* but are not attempting to do so. That said, it is important that the task of accounting for ordinary arithmetical knowledge, the kind possessed by non-specialists, should not fall off the agenda for epistemologists of mathematics considered *en masse*.

Non-reductive Platonisms

The project of accounting for the knowledge of non-experts has venerable roots in the ancient history of our discipline. In Plato's *Meno*, a classic presentation of a Platonic epistemology of mathematics, an uneducated slave-boy's knowledge of geometry is used as a focal example of the kind of phenomenon in need of philosophical explanation. To address this need, Plato's Socrates offers a theory of *a priori* knowledge which appeals to the broader Platonic doctrine of recollection. According to this doctrine (developed further in other dialogues, including the *Republic*, *Phaedo*, and *Phaedrus*), one is able to know certain things *a priori* because one is *remembering* things learned as a disembodied soul, prior to one's (current) incarnation. Souls are in direct contact with idealized abstract objects – the Forms – – and can, when prompted, recall information about them. Taking arithmetical objects such as the natural numbers to be among the Forms, known prior to birth and recollected when prompted by teachers or experience, supplies a simple account of how we – that is, any or all of us – can know *a priori* that $2 + 2 = 4$.

While Plato's mythology of reincarnation is no longer regarded as the basis for a promising account of *a priori* mathematical knowledge, contemporary views are often labelled forms of *Platonism* if they posit mind-independent, abstract mathematical objects. *Non-reductive* Platonisms moreover make no attempt to identify these objects as being of some (relatively) philosophically untroubling variety, but rather embrace their distinctive abstract status. Such Platonisms cleave strongly to the first half of Benacerraf's dilemma: arithmetic *appears* to be about distinctive abstract arithmetical objects because it *is* about such objects.

These views thus confront the challenge of the dilemma's second half in its purest form: how can we account for knowledge of such objects? One famous (or infamous) answer is inspired by philosopher-mathematician Kurt Gödel's remark that "despite their remoteness from sense-experience, we do have something like a perception of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true." He continued: "I don't see any reason why we should have less confidence in this kind of perception, i.e. in mathematical intuition, than in sense-perception" (Gödel 1947, pp. 483–484). While Gödel's primary interest here was in set theory, one might easily extend or adjust the view to postulate "something like a perception" of numbers.

Positing such a faculty (often labelled 'intuition' or 'rational intuition') is not a popular contemporary option, largely because it is felt to be in tension with *naturalism*. Naturalism is generally seen as demanding a scientific worldview grounded in empirical evidence; a faculty of intuition sits uneasily with this, both because it would itself *be* a non-empirical epistemic source, and because there is a dearth of empirical evidence for its existence.

Another non-reductive option is known as 'plenitudinous' Platonism (see Balaguer 1998). According to this view, all consistent mathematical theories are true, and all of their objects exist. So *any* mathematical proposition one believes – as long as it is not inconsistent – is true. Thus, Field's demand for an explanation of the reliability with which we believe truly in the mathematical domain is supposedly put to rest: the accomplishment is trivial, hence there's nothing to explain.[2]

To the extent that this feels unsatisfying, that feeling may flag a problem with Field's move to reposition Benacerraf's challenge as a demand for such an explanation. Recall that what Benacerraf originally sought was a viable pairing of accounts, one semantic and one epistemological, to deal with propositions like *2 + 2 = 4*. Plenitudinous Platonism, at least in its purest form as an ontological theory, supplies neither. Another (related) kind of objection is that it is never an entirely trivial accomplishment to *refer to* or to *know about* some objects; merely postulating the existence of *lots* of objects does not change that.

Non-reductive Platonisms may be attractive insofar as they hold out the prospect of respecting both the abstractness and the *a prioricity* of arithmetic. They also tend to be well-positioned to account for all kinds of arithmetical knowledge, expert and non-expert alike, in ways that may differ in degree but do not require entirely different epistemologies. Their problems, in my opinion, tend to arise in connection with the applicability constraint. It is difficult to explain the applicability of abstract mathematical objects to the physical world, and my diagnosis is that it is precisely this explanatory gap which underwrites widespread suspicion that any genuine attempt to explain how such flesh-and-blood concrete creatures as ourselves could have *knowledge* of abstracta will be "spooky" and insufficiently naturalistic.

In the twentieth century, Willard van Orman Quine made a sustained effort to marry non-reductive Platonism (particularly about sets) with a naturalistic and empiricist epistemology for mathematics. The most influential statement of this project's outlines can be found in Quine (1951), and the core ideas are developed in the work of many later philosophers (see, e.g., Colyvan 2001; Devitt 2005). Quine proposed a form of epistemological *holism*, according to which it is not individual beliefs or hypotheses that are tested against experience, but rather entire worldviews as a package deal. When empirical confirmation is received, it accrues to everything in the package at once, including mathematical beliefs. According to such a view, our knowledge that *2 + 2 = 4* is not in fact *a priori*, but a (long-established) element of our best empirically confirmed overall theory. As

has been emphasized by Devitt in particular, the holist attempts to *explain away* the appearance of *a prioricity*: to defuse *challenge (ii)*, rather than meet it head-on.

The Quinean approach can, however, be understood as a direct attempt to accommodate the abstractness of mathematics – *challenge (i)* – in tandem with the applicability constraint. It's not an *accident* that we choose the mathematics that is applicable to the world as we experience it: mathematics is deployed in our theories precisely to be so applicable, to help us explain and predict. That theories about abstracta should be helpful in these ways may be surprising to philosophers, but they nevertheless appear to be (not merely helpful but) *indispensable* for best science. Whether this truly meets the challenge of applicability is up for debate, however. What this approach explains is *why certain applicable mathematical theories are the ones we adopt*, not *why those particular ones are applicable in the first place.*

The holistic approach may also face questions as to who are the relevant 'we'. Most people do not depend on much advanced mathematics in their understanding of the world, but if the relevant 'we' is set as (say) some expert community of scientists, the question remains of what *their* situation has to do with everyday, non-expert knowledge of $2 + 2 = 4$.

Reductions/Identifications

If one wants to believe in an arithmetical reality – a mind-independent realm for propositions like $2 + 2 = 4$ to be about – without committing to any arithmetical objects that feel too "spooky" or strange for a respectable naturalistic ontology, one can try arguing that arithmetical reality is identical with some realm of reality that is less spooky-sounding, at least on the face of it, and/or some realm of reality already accepted as *bona fide*.

One might, for example, identify particular arithmetical objects (like the number 4) with naturalistically respectable objects. This has the advantage of proving a relatively simple semantics for $2 + 2 = 4$, like non-reductive Platonisms. Another option is to identify arithmetical reality more broadly with some unspooky realm of reality, without making specific object-identity claims. This option leaves more semantic questions up in the air.

Such identification strategies are often called *reduction* strategies, the idea being that arithmetic is *reduced* to a domain of reality that is already acceptable, thus avoiding the 'inflated' ontology of a non-reductive Platonism. This terminology is in some ways unfortunate, since 'reduction' carries connotations of asymmetry that 'identification' does not. If the posited identities really do obtain, such connotations may be misleading (since identity is symmetric).

One well-known (but not popular) identification strategy is logicism, which in broad-brush terms is the view that arithmetic is part of logic. Like many of the views in this chapter, logicism comes in more and less general varieties: one can be a logicist about all of mathematics, or about specific subdisciplines such as arithmetic. Different areas of mathematics offer different prospects and problems

for logicist identification. Arithmetic is, however, the area whose susceptibility to a logicist treatment has been most thoroughly investigated. The classic attempt to establish an identification is Gottlob Frege's two-volume work *Grundgesetze der Arithmetik* (*Basic Laws of Arithmetic*) published in 1893 and 1903. This attempt was doomed when Frege's proposed Basic Law 5 turned out to be inconsistent – as established by Bertrand Russell, it leads to a paradox.[3]

When it comes to epistemology, a logicist has no single, obviously-preferable option, but can at least argue that what appeared to be two sources of epistemological puzzlement are really one. If arithmetic is a branch of logic, then arithmetical knowledge is a species of logical knowledge, and our preferred epistemology of logic becomes an epistemology for arithmetic also. Logicism is also relatively well-placed to make space for the abstractness and *a priori* knowability of arithmetic, at least on the assumption that logic shares these features. With respect to the applicability of arithmetic, the logicist again starts on firm ground: logic is applicable to reasoning about absolutely anything.[4] However, the ubiquity of non-specialist arithmetical knowledge risks throwing a spanner in logicist works. Frege's logicist derivations of the standard axioms for arithmetic (known as the Peano Axioms) are extremely specialized. Almost nobody in the world has learned how to perform these derivations, or anything remotely like them, and yet almost everybody in the world knows that $2 + 2 = 4$.

A second well-known identification strategy is arithmetical *structuralism*, which attempts to establish that arithmetic is the study of certain structures, hoping thus to avoid the non-reductive Platonist's commitment to spooky objects. One formulation is due to Stewart Shapiro (1997), who outlined what is sometimes called an *ante rem* variety of structuralism (to distinguish it from *in rebus* structuralism, of which more in the next section). According to a Shapiro-style structuralist, structures are characterized by structural relations. These structures exist independently of any objects which exemplify their defining relations (hence the label 'ante rem'). Knowing that $2 + 2 = 4$ is thus knowing something about the natural number structure; in particular, it is knowing something about the relations that define certain positions (*2* and *4*) within that structure.

Structures of this kind are still abstracta, and questions remain as to whether they are a sufficiently comprehensible or unspooky kind of abstracta. This form of structuralism is among the most promising of the views surveyed in this chapter with respect to the universal *applicability* of arithmetic, since it makes arithmetic the study of structural relations which absolutely any kind of object can instantiate. (It also tidily explains what application actually consists in, i.e. instantiation.) Non-specialist knowledge is also accommodated, on the plausible assumption that a structure is the kind of thing one can know a little or a lot about, and about which one might learn in various (formal and informal) ways. The view may also appear to fare well in accommodating the abstractness of arithmetic, since it renders arithmetic as the study of abstracta; however, an effective response to Benacerraf's challenge requires in addition some explanation of how this abstract

domain is known to us, and (apparently) known *a priori*. One attempt to develop an epistemology for arithmetic that sits well with structuralism of this bent is found in Jenkins (2008),[5] where I discuss the hypothesis that aspects of the world's arithmetical structure impacts us through sense experience, and thereby epistemically "grounds" our most basic arithmetical concepts, from which we may then recover information about that structure.

Anti-Realisms, Type A (No-Truths) and Type B (No-Objects)

Further alternatives to non-reductive Platonism can be bundled together under the umbrella of *anti-realism* (although that label admits of such broad and variable usage that its application to any of these positions is not always particularly illuminating until further clarification is appended). Broadly speaking, there are three main ways to be an anti-realist about arithmetic:

a. Deny that arithmetical propositions such as $2 + 2 = 4$ are true.
b. Allow that they are true, but deny that they are about objects (such as numbers).
c. Allow that they are true (and perhaps about objects), but deny that their truth is mind-independent.

Option a is perhaps the most counterintuitive. While it may sound extreme, however, it promises a get-out-of-jail-free card to the metaphysician troubled by abstracta, and derivatively (albeit more relevantly for our purposes) to the epistemologist troubled by Benacerraf-style worries. If arithmetic is not a domain of truths, then there are no mysterious arithmetical objects and no mysterious arithmetical knowledge to account for.

Two principal versions of this *no-truths* form of arithmetical anti-realism are fictionalism and formalism. According to the fictionalist, arithmetical propositions like $2 + 2 = 4$ are strictly speaking false, although it might be very useful to treat them as if they were true for certain purposes. In the ontology room, the fictionalist says, we must say that they are false and hence do not commit us to peculiar objects like the number 4.[6] Field is known for defending a form of fictionalism in his 1980 book *Science Without Numbers*, where he argued that the use of arithmetic in scientific applications does not commit us to the truth of the arithmetical propositions so applied.

While the formalist agrees with the fictionalist that arithmetical propositions like $2 + 2 = 4$ are not true, the formalist says they are not false either, but (strictly speaking) meaningless.[7] The activity we call 'arithmetic' is an activity in which we manipulate formal symbols like '2' and '4', and while we have rules for such manipulations which permit the string '2 + 2 = 4' and forbid the string '2+2=5', this is not because of what they *mean*. Neither string, in fact, has any semantic content.

Fictionalists and formalists alike sidestep the difficulties associated with accounting for arithmetic's *a priori* knowability and the abstractness of its subject matter. They can comfortably accommodate the role of non-specialists, too, as inexpert participants in the (fictional or formal) activity of arithmetic. However, such strategies come at a high cost when it comes to the applicability of arithmetic. In a discussion of formalism, in §91 of volume II of the *Grundgesetze*, Frege stated that 'it is applicability alone which elevates arithmetic above a game to the rank of a science' (1893/2013, p. 100). The equivalent issue arises for fictionalism, where it becomes a version of the Putnam-Boyd *no-miracles* argument against anti-realism in the philosophy of science:[8] if the theory isn't (by and large) true, why does it work so well?

Option b is *no-objects* anti-realism about arithmetic. The term *nominalism* is sometimes used for this kind of view (derivatively upon its use to describe the rejection of universals and/or abstract objects in general). Fictionalism and formalism deliver nominalism fairly directly – indeed, the full title of Field's fictionalist 1980 is *Science Without Numbers: A Defence of Nominalism*.

In mathematical contexts, the *no-miracles* argument has been sharpened into what is known as the Quine-Putnam *indispensability* argument against nominalism.[9] It runs roughly as follows: propositions quantifying over numbers are indispensable elements of our best scientific theories, and (applying Quine's criterion of ontological commitment) this means that accepting those theories commits us to the existence of numbers. We accept the theories, so we must accept the numbers.

It is this kind of argument to which Field was responding in his 1980 book, attempting to show that scientific theories can be viably reformulated to avoid quantification over numbers. His response is limited in scope (he only addressed part of Newtonian gravitational theory), and it is controversial whether it succeeds even within its limits – for example, one may dispute whether his convoluted nominalistic proposal is a viable contender for a *good* (never mind our best) scientific theory, given that simplicity is generally considered an important theoretical virtue, and whether Field's proposal sneaks in abstract number-substitutes in the guise of what he called 'space-time regions'.

But one may also be a nominalist for less drastic reasons than being a fictionalist or formalist. Some forms of structuralism are best classified as *type b* anti-realisms (but *not type a*), especially those that reject from their ontologies not only numbers but also structures. These are known as *in rebus* structuralisms because they hold that arithmetic is the study of certain kinds of structural arrangements in which concreta stand, but that no abstract structures exist beyond, or in addition to, the instantiating concrete things. This type of view, however, renders mathematical truth hostage to the factual question of how many concreta exist.[10]

Nominalist structuralisms duck the challenge of explaining how we can have knowledge concerning a realm of abstract mathematical objects by denying that there are any such objects. Such views also inherit some of the advantages of

Platonist *ante rem* structuralism when it comes to accounting for non-expert knowledge, and for the applicability of arithmetic. It should be noted, however, that the modalized versions are less well placed in both regards, as they may end up recommitting to abstracta in their modal ontologies, and face the additional task of explaining why merely possible structures are applicable to the actual world. When it comes to accounting for the apparent *a prioricity* of arithmetical knowledge, however, nominalist structuralists have significant work to do.

Perhaps the appearance of *a prioricity* can be explained away. If this is the task, it is one shared by a quite different form of nominalism, less metaphysically sophisticated than that of the *in rebus* structuralists, but important enough to discuss here. In his 1843 work *A System of Logic*, J.S. Mill famously denied that arithmetical knowledge is *a priori*. He maintained that arithmetic is simply a branch of the empirical, *a posteriori* study of nature, and that arithmetical theorems are among the most general laws of nature. In effect, $2 + 2 = 4$ states that any two objects added to any other two objects makes four objects. But "[a]ll numbers must be numbers of something: there are no such things as numbers in the abstract" (1843, Vol. I, Book II, Chapter VI, §2).

This Millian form of empiricism takes non-specialist knowledge and the applicability of simple arithmetic in its stride, but faces difficulties accounting for the application of arithmetic to extremely large numbers of things that we have never experienced (and that may not even exist). This problem also promises to recur in spades when one moves from arithmetic to consider higher mathematics. It is also important to note in general that directly rejecting the abstractness and *a prioricity* of arithmetic is no epistemological get-out-of-jail-free card, since finding an adequate empirical basis for *all* knowledge of arithmetic (not just knowledge of elementary sums) is a huge task. (Moves to less simplistic empiricist views, such as Quinean holism, are motivated by precisely such considerations.)

Anti-Realisms, Type C (Mind-Dependence)

The third and final kind of anti-realism to be surveyed here is *type c*, or mind-dependence, anti-realism. Several well-known – but very different – positions come under this heading.

The first is Kant's conception of arithmetic, as put forward in his 1781 *Critique of Pure Reason*. Kant says that arithmetic is the form of our 'temporal intuition'. While Kant exegesis is a fraught business, one way of unpacking this has Kant saying that the way we (human agents) experience the world is temporally structured, while the world as it exists in itself – that is, outside of experience – cannot be assumed to share this temporal structure. Time is thus an aspect of *our* ways of thinking, feeling, experiencing and understanding. It is temporal structure so understood that is – both metaphysically and epistemologically speaking – the *basis* of arithmetic, according to Kant.

This Kantian approach has been a strong contender in the epistemology of arithmetic ever since, and for good reason, as can be seen by noting its potential

for accommodating the criteria that structure this chapter. The applicability of arithmetic is built right into the account: arithmetic is the study of the (temporal) structure of experience, so naturally it is applicable to the world as we experience it. Non-specialist knowledge is also straightforwardly accounted for: we all share this temporal form of intuition. Next, arithmetical knowledge is positioned as a form of self-knowledge, since arithmetic is an aspect of our own contribution to the world of experience, and this is widely held to be helpful in accounting for its *a prioricity*: one looks *inward* rather than outward to learn about arithmetic. (However, one might have reservations about this, as certain kinds of self-knowledge are hard to come by, and cannot be secured through introspection.) Whether the proposal must *explain* or *explain away* knowledge of abstract objects such as numbers will depend on how one spells out the details of the Kantian view; its emphasis on time as *structuring* our experience might be developed into something resembling a Platonist or a nominalist form of structuralism.

Wittgenstein (especially in his *Remarks on the Foundations of Mathematics*, a collection of notes published in 1956) proposes a different, and in some ways more radical, kind of mind-dependence. Wittgenstein argues that all of mathematics is created (or invented) by us. In keeping with his broader interest in the nature of rules and rule-following, he treats arithmetic (and mathematics in general) as a domain of normative rules, such as the rule that says when adding 2 and 2 you should give the answer 4. But such rules, he argues, are created by us in the process of practicing arithmetic – that is to say, by calculating, counting, and so on, we generate arithmetical truths. This is a (particularly stark) form of *constructivism* in the metaphysics of arithmetic, which has historical precedents in the work of L. E. J. Brouwer.[11] Brouwer is also known as an *intuitionist* because he held that it is the possibility of constructing a proof in the mind – in one's 'intuition'[12] – that renders a mathematical proposition true.

As far as epistemology goes, constructivist proposals face a suite of advantages and disadvantages. On the one hand, arithmetical knowledge does not involve access to a realm of spooky abstract objects. Whether there are arithmetical objects at all (and whether – and in what sense – they are abstract) depends on the details of one's constructivist metaphysics, but the realm of arithmetic is at least restricted to the realm of the constructed – that is, constructed *by someone* – and is to that extent all "within reach." The *a prioricity* of arithmetical knowledge might be approached in a somewhat Kantian spirit: if arithmetical knowledge is *maker's* knowledge, this will differentiate it in certain ways from the knowledge of *discovered* facts. The applicability of arithmetic needs addressing but need not be an insurmountable challenge. (Why would an invented or constructed arithmetic apply tidily to the world? Perhaps because it was built specifically for that purpose.)

Non-specialist knowledge is more awkward, however, at least for the intuitionist, since most non-mathematicians' knowledge of simple arithmetical truths has little to do with *proof* (or construction) in anything like a mathematician's

sense. Whether or not this is an issue for Wittgenstein's constructivism is hard to pin down, due to his often vague style of writing (and perhaps, depending on one's exegesis, his changes of mind). He tends to use first-person plural language ('we', 'our') without specifying who is included.

A third category of mind-dependence anti-realism is worth canvassing briefly. This is the class of views on which true arithmetical statements are *analytic*: made true by the meanings of words, or by relations between concepts. This approach finds a *locus classicus* in the work of the Logical Empiricists, and particularly Carnap, who argues that we adopt a framework of arithmetical concepts for pragmatic reasons: that is, because such a framework is useful in application to the physical world (see Carnap 1950). The question of whether numbers exist can then be asked as a question *internal* to this framework, in which case the answer is trivially "yes," or it can be asked as an *external* question, in which case it is empty (because talk of numbers only makes sense within the arithmetical framework).

On this view, our *a priori* knowledge of arithmetic is accounted for as knowledge concerning the framework of arithmetical concepts that we have adopted. The applicability of such knowledge is built into the view: the framework has been adopted *because* it is useful. Non-specialist knowledge is also relatively unproblematic, insofar as anyone who is using the arithmetical framework might be supposed to know some elementary facts about the relations between its constituent concepts. The strangeness of knowledge concerning abstract arithmetical objects is purportedly defused: it is a trivial matter to answer *internal* questions about such objects, and those are the only kinds of questions one can sensibly ask about them. (Precisely this kind of *deflation* of arithmetic's ontology, however, may prove a sticking point for those struck by the apparent robustness of arithmetic.)

A more sophisticated descendent of the analyticity view is *neo-Fregeanism*, initially developed by Crispin Wright and Bob Hale (see Wright 1983; Hale and Wright 2001). This work secures a derivation of the standard Peano Axioms for arithmetic from a single premise, known as Hume's Principle, which is then positioned as analytic (or an implicit definition) of *number*. This, in some respects, loops us back around to logicism (and, indeed, this neo-Fregean view sometimes also goes by the name *neo-logicism*): for, like logicists, neo-Fregeans maintain that all that is required to ground arithmetic are logic and definitions.

A priori arithmetical knowledge is accounted for by neo-Fregeans as knowledge of analytic or definitional matters (though the *a prioricity* of logic still requires its own explanation). Because this is a crucial point, much subsequent debate has centered on whether or not the premise known as Hume's Principle *is* in fact analytic in the sense required to deliver these epistemological results.

It is also worth noting that neo-Fregeanism inherits some potentially troubling features of its logicist ancestors, including apparent inapplicability to non-specialist knowledge of arithmetic (which does not seem to proceed from knowledge of anything like Hume's Principle, or the derivations therefrom of the Peano

Axioms). Neo-Fregeanism also supplies no particularly obvious explanation for the applicability of arithmetical knowledge to the physical world.

Conclusions

The question of how we know that *2 + 2 = 4* remains hotly debated in contemporary epistemology of mathematics. As the foregoing discussion evinces, the answers preferred by different philosophical camps depend heavily on what kind of truth (if any) they take *2 + 2 = 4* to be. In particular, metaphysical divisions between non-reductive Platonists, reductionists, and anti-realists of various stripes are reflected in very different approaches to arithmetical epistemology.

In a similar manner, the extent to which philosophers are inclined to classify arithmetic as a *science* tends to be reflected in a corresponding disinclination to emphasize its apparent epistemological differences from the (paradigmatic, or stereotypical) natural sciences. In particular, mathematical naturalists who (following Quine) classify arithmetic firmly as part of science tend correspondingly to downplay or deny the *a prioricity* of arithmetic, emphasizing instead its dependence on empirical confirmation.

Like the question of how we know *2 + 2 = 4*, the question of whether arithmetic is a science or not remains unsettled, lacking even an emergent consensus. To an extent, however, this latter question is one about the contingencies of disciplinary borders and categories, and as such might be most helpfully addressed by sociologists, anthropologists, and historians of academia.

Notes

1 But see Burge (1993) (and elsewhere).
2 Avoiding inconsistency is non-trivial in some cases, but when our attention is focused on propositions like *2 + 2 = 4* we might reasonably waive that concern.
3 This is "Russell's Paradox," the famous problem that arises when we consider a set whose members are *all the sets that are not members of themselves*.
4 See MacFarlane 2015, §4 for a helpful discussion of the notion of *topic neutrality*, which untangles a potential complication here. One kind of topic-neutrality *is* universal applicability; both logic and arithmetic are topic-neutral in this sense. In the other sense, to be topic-neutral is to contain no expression that discriminates between particular objects. Since arithmetic does appear to contain some such expressions (e.g., 'is the number 4'), it is not straightforwardly classifiable as topic-neutral in the second sense. But if it turns out to have been only logic all along, it may be possible to argue that arithmetic is topic-neutral in both senses.
5 See especially Chapters 4 and 5.
6 Fictionalism about arithmetic admits of many possible permutations with respect to the details of the view. For a good overview, see Balaguer (2015).
7 Early versions of formalism, made famous as the subject of Frege's critique in his *Grundgesetze*, were unclear on this point. See section 2 of Weir (2015) for a helpful historical discussion.
8 Developed in, e.g., Putnam (1975) and Boyd (1989).

9 See Colyvan (2001) for a thorough treatment.
10 And because this strikes many as counterintuitive, modalized alternatives have been explored, for example in Geoffrey Hellman's (1989) *Mathematics Without Numbers*. On this approach, arithmetic becomes the study of *possible* structures.
11 See Brouwer 1981 (a collection of material for lectures Brouwer delivered between 1946 and 1951).
12 Cautionary note: the English word 'intuition' is used in translations of both Brouwer and Kant, but it should not necessarily be assumed to mean exactly the same in each case.

References

Balaguer, M. 1998. *Platonism and Anti-Platonism*. Oxford: Oxford University Press.
Balaguer, M. 2015. 'Fictionalism in the Philosophy of Mathematics', in: E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, edn. of summer 2015, accessed 28 June 2018 at https://plato.stanford.edu/entries/fictionalism-mathematics/.
Benacerraf, P. 1973. 'Mathematical Truth,' *Journal of Philosophy* 70, pp. 661–680.
Boyd, R. 1989. 'What Realism Implies and What it Does Not', *Dialectica* 43, pp. 5–29.
Brouwer, L.E.J. 1981. *Brouwer's Cambridge Lectures on Intuitionism*, D. van Dalen (ed.), Cambridge: Cambridge University Press.
Burge, T. 1993. 'Content Preservation', *Philosophical Review* 102, pp. 457–488.
Carnap, R. 1950. 'Empiricism, Semantics and Ontology', *Revue International de Philosophie* 4, pp. 20–40.
Colyvan, M. 2001. *The Indispensability of Mathematics*. New York: Oxford University Press.
Devitt, M. 2005. 'There Is No A Priori', in: E. Sosa and M. Steup (eds), *Contemporary Debates in Epistemology*, Cambridge, MA: Blackwell, pp. 105–115.
Field, H. 1980. *Science Without Numbers: A Defence of Nominalism*. Oxford: Blackwell.
Field, H. 1989. *Realism, Mathematics, and Modality*. Oxford: Blackwell.
Frege, G. 1893. *Grundgesetze der Arithmetik/Basic Laws of Arithmetic*, edn. of 2013 translated by P. Ebert and M. Rossberg, Oxford: Oxford University Press.
Gödel, K. 1947. 'What Is Cantor's Continuum Problem?', *American Mathematical Monthly* 54, reprinted in P. Benacerraf and H. Putnam (eds.) Philosophy of Mathematics, 1964, Cambridge University Press, pp. 258–273.
Hale, B. and Wright, C. 2001. *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon Press.
Hellman, J. 1989. *Mathematics Without Numbers: Towards a Modal-Structural Interpretation*. New York: Oxford University Press.
Jenkins, C. 2008. *Grounding Concepts*. Oxford: Oxford University Press.
Kant, I. 1781. *Critique of Pure Reason*, edn. of 1929, translated by. N. Kemp Smith. Basingstoke: Palgrave.
MacFarlane, J. 2015. 'Logical Constants', in: E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, edn. of summer 2015, accessed 28 June 2018 at https://plato.stanford.edu/entries/logical-constants/.
Mill, J. S. 1843. *A System of Logic, Ratiocinative and Inductive*, Vols. I and II. London: Parker.
Putnam, H. 1975. 'What Is Mathematical Truth?', in his *Mathematics, Matter and Method*, *Philosophical Papers Vol. 1*, Cambridge: Cambridge University Press, pp. 60–78.
Quine, W.V.O. 1951. 'Two Dogmas of Empiricism', *Philosophical Review* 60, pp. 20–43. Reprinted in *From a Logical Point of View*, pp. 20–46.

Shapiro, S. 1997. *Philosophy of Mathematics: Structure and Ontology*. New York: Oxford University Press.

Weir, A. 2015. 'Formalism in the Philosophy of Mathematics', in: E. Zalta (ed.) *Stanford Encyclopedia of Philosophy*, edn. of summer 2015, accessed 28 June 2018 at https://plato.stanford.edu/entries/formalism-mathematics/.

Wittgenstein, L. 1956. *Remarks on the Foundations of Mathematics*, edn. of 1978, translated by G.E.M Anscombe, G. H. von Wright, R. Rhees and G.E.M. Anscombe (eds). Oxford: Basil Blackwell.

Wright, C. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

9

IS SCIENTIFIC KNOWLEDGE SPECIAL?

Plus ça change, plus c'est la même chose

Richard Fumerton

Introduction

Other things being equal, I try to avoid describing the main thesis of a chapter in such a way that it will strike most readers as absurd. Here I make an exception. I will argue that while there are, of course, different sciences defined by their subject matter, there is nothing particularly special about the kind of *knowledge* that people seek in the areas they investigate, and further that there is no strong reason to believe that there is anything special about the kind of *reasoning* that scientists employ. It is likely the same kind of reasoning that ordinary people have been employing for millennia in reaching commonplace conclusions about the world around them.

To make this suggestion about reasoning a bit more palatable, it is necessary to draw a distinction between *fundamental* and *derivative* epistemic principles, and a distinction between evidence and reasoning.

Fundamental vs. Derivative Epistemic Reasoning

As I acknowledge it might seem initially absurd to suppose that there has been no significant progress in scientific reasoning. In today's world we have all sorts of ways of explaining and predicting phenomena that weren't even imagined hundreds of years ago, let alone thousands of years ago. We have made huge advances in technology that have introduced instruments that allow us to detect all kinds of things that were previously undetectable. To take a trivial example, ancient people didn't have electron microscopes. And to take another trivial example, today no-one (or no normal person) tries to predict the outcome of battles by reading the entrails of birds. So, what sane person is going to insist that there haven't been profound changes and advancements in scientific reasoning?

The first step in trying to make sense of my conclusion is to emphasize that it is a thesis about *reasoning*. Most of the reasoning we employ in everyday life is *enthymematic* – we don't bother to state all of premises upon which we rely. I see tracks that look like they were made by deer and infer that a deer has been in the vicinity. I hear thunder and infer that there is lightning. I see puddles on my driveway and infer that it has rained. My wife gives me a look with which I am all too familiar, and I realize that she's upset with me about something. The thermometer outside my house reads 74, and I infer that it is roughly 74 degrees outside. But, I would contend, there are no principles of *reasoning* that license any of the above "inferences." Put in terms of the idea of an argument form, the implicit arguments sketched above have no *legitimate* form. They all have precisely the same form: P; therefore, Q. Alternatively, one could suggest that the "inference" from, say, the color of the litmus paper to the acidity of the solution is real but derivative. We recognize that inference as legitimate, though, only because we infer the legitimacy of inference from premises that involve yet another inference. We have an independent test for the acidity of a solution as well as a test that involves what an acid solution does to the material of which the litmus paper is made. But rather than recognize a "litmus paper" inference, it seems to me clearer to state the point as I first did. There is no inference *at all* from a premise describing the litmus paper to the acidity of the solution. The real inference is from that proposition together with all the critical unstated premises necessary to reach the relevant conclusions.

Now one might immediately object that I'm simply confusing the kinds of conditions necessary for legitimate (valid) *deductive* reasoning with what we might reasonably expect to see corresponding to legitimate non-deductive reasoning. As we tell our students, the mark of a deductively valid argument is that we can determine its validity based on the form of the argument alone. Any argument of the form P and (if P then Q), therefore, Q is valid. We don't even need to think about the content of the premises and conclusion.[1] But there isn't even a candidate for the corresponding *form* of legitimate non-deductive reasoning, at least on *some* conceptions of what constitutes legitimate non-deductive inference.

There is a sense in which this might be true, for example, if certain externalist[2] accounts of justified belief are correct. Consider, for example, a relatively straightforward reliabilism of the sort Alvin Goldman put forth in his now classic "What is Justified Belief?" Goldman's view is a version of foundationalism.[3] The foundationally justified beliefs that he recognized are those that are formed by unconditionally reliable processes whose input is something other than a belief.[4] Inferentially justified beliefs, by contrast, result from belief-forming processes – their inputs include beliefs, and the output beliefs are justified only if the input beliefs are justified and the process is *conditionally* reliable – usually results in true output beliefs when the input beliefs are true.[5] As I have often pointed out, there are no *a priori* restrictions on what might turn out to be a contingently reliable belief-forming process (though there are, of course, belief-forming processes that are necessarily unreliable, and others that are necessarily reliable.)[6] As a result

there are no uncontroversial examples of noninferentially justified beliefs on an externalist's view. It all depends on whether the belief forming process involves beliefs as input.

Goldman's view is a view about what makes a belief justified. But it also suggests that in the case of belief-dependent processing, the epistemic agent engages in a kind of reasoning – a kind of "movement," at least, from the input beliefs to the output beliefs.[7] The reliabilist can decide whether the inputs to a belief-independent process should include not only conscious beliefs but unconscious beliefs, or even dispositions to believe whose ground is playing a causal role in producing the relevant output (see Fumerton, 2019). These decisions will, no doubt, affect how often a belief will enjoy non-inferential justification. But the important point for us here, is that there is no reason to suppose that the relevant processing (whether it is noninferential or inferential) can be captured by any sort of *form* the reasoning has. It is not too strong to suggest that there is no critical relation that holds between the premises and conclusion of the relevant reasoning process. The processing will presumably be of some *type* of other,[8] but whatever type is relevant won't be revealed by putting the input and output into an argument form where there are premises and a conclusion.

What is true of reliabilism is true of most other externalist accounts of justified belief and the legitimacy of input/ouput relations. This is not the place to evaluate externalism. I have raised elsewhere (1995) what I take to be the critical problems faced by reliabilist accounts of philosophically relevant justification. In what follows I'll begin by contrasting reliabilism with more traditional ways of thinking about non-deductive inference.

Consider, for example, enumerative induction. These inductive arguments do indeed have a form. As Bertrand Russell (1912, Ch. VI) suggested they typically take one of the following two forms:

A.
1a) All (most) observed F's have been G
Therefore,
2a) All (most) F's are G,
or,
B.
1a) All (most) observed F's have been G
2b) a is F
Therefore,
3b) a is G

Let's not worry for a moment about whether this oversimplifies the nature of inductive reasoning. Nelson Goodman (1955) convinced many that the legitimacy of inductive reasoning trades on the critical notion of whether we are dealing with

a "projectible" predicate. And whether or not a predicate is projectible (on his view) won't be a matter detectable by form. But again, leaving that aside, the above arguments have a form. And because they do, Russell looked for a principle (the principle of induction) that might assert the legitimacy of arguments with that form.

Or consider reasoning to the best explanation or what C. S. Peirce (1938) also called abductive reasoning. As Peirce described this reasoning, we make a surprising observation, make an assertion about what would explain this observation, and (tentatively) reach as our conclusion that the assertion is true. Put in terms of the form suggested by Peirce the abductive argument would look like this:

C.
1c) O (the observation/potential explanandum)
2c) If E (the potential explanans) had been the case O would also be the case
Therefore,
3c) E

Almost everyone would agree that we can do better than this at representing the form of reasoning to the best explanation.[9] There will be indefinitely many arguments of this form, with true premises, and we obviously need some way of selecting from among all of the possible explanans. We probably need something that looks more like this:

D.
1c) O (the observation/potential explanandum)
2d) If E were the case that would *better* explain O than any competing explanation.
Therefore,
3c) E

Or better still:

E.
1c) O (the observation/potential explanandum)
2e) If E were the case that would be a more likely explanation of O than the disjunction of all competing explanations.
Therefore,
3c) E

It's worth noting as an aside that (as with any non-deductive reasoning), the reasoning in question could alternatively be put in the form of a deductively valid argument.

F.

1c) O (the observation/potential explanandum)
2f) There is an explanation for O
3f) E is the correct explanation for O
Therefore
3c) E

There may be no benefit in putting the argument this way, for what one gains by way of having premises that entail a conclusion, one loses by way of increased epistemological pressure to establish the relevant premises. I mention the above only because I have argued elsewhere that it might be an illusion to suppose that reasoning to the best explanation is a sui generis kind of reasoning.[10]

Bayesians, of course, have their favorite form of non-deductive reasoning, captured by Bayes' theorem, which states that the probability of A given B is equal to the probability of B given A times the probability of A, divided by the probability of B. Put more formally:

G.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)},$$

Some Bayesians argue as if the principle is the be all and end all of all non-deductive reasoning, but any useable theory needs an account of how one understands and how one gets the non-relativized probabilities (the problem of the priors), and, though I won't argue the point here, it is not clear to me that any plausible account is forthcoming.

A, B, C, D, E and G all have a recognizable form, and one can argue as to which if any of these inferences are sanctioned by *correct* non-deductive principles.

In the history of philosophy, there were some radical empiricists who seemed to suggest that there was only one sort of legitimate non-deductive reasoning – that represented by either A or B above. But that view has an extraordinarily difficult time avoiding skepticism. Unlike David Hume, most contemporary philosophers follow Thomas Reid's (1854, ch. 4) suggestion that if a view implies radical skepticism that's a clear indication that the view is false. Roderick Chisholm (1966, ch. 4) was, perhaps, as clear as anyone that his epistemology begins with an almost existential choice: Do whatever you need to do side with "commonsense" against skepticism. Concluding that induction won't take you very far, Chisholm introduced all sorts of epistemic principles that sanction reaching various commonsense conclusion about the external world and the past. In that vein one might again argue for certain forms of legitimate non-deductive argument. Simplifying greatly, one might suggest that the following argument has premises that make probable its conclusion:

1) I seem to remember having had experience E
   Therefore,
2) I had experience E

Or trying to reach a conclusion about the external world, one might license inferences of the following sort:

1) It appears to me as if there is something that is F
   Therefore,
2) There is something that is F

To be at all plausible one would probably place further restrictions on what the relevant property variable ranges over. So we might be well advised to insert the adverb "phenomenally" to modify F in the inference form described above.

Michael Huemer (2001) embraced phenomenal conservatism, a view according to which the fact that something seems to one to be the case provides prima facie justification for one to believe that it is the case.[11] Huemer explicitly denied that the seeming *needs* to be described in a premise which one justifiably accepts in order for one to reach the relevant conclusion, but he would also probably recognize as legitimate (by non-deductive standards) an argument of the form:

1) It seems to S that P
   Therefore,
2) P

As with all non-deductive arguments, from the fact that the premises make probable the conclusion it doesn't follow that the premises conjoined with any other proposition also make probable the conclusion. That is, of course, one of the key differences between premises making probable a conclusion and premises entailing a conclusion.

Two More Preliminary Observations and Two Concessions

It might be true that the way we reason hasn't changed over millennia even if people have often engaged in reasoning from false or unjustified premises. Ancient people who tried to predict the outcome of battles by reading the entrails of birds almost certainly recognized that they were relying on suppressed premises. Most obviously, they probably realized that they needed a justified belief that there is some sort of correlation between the entrails being bloody and their fortunes going badly.[12] We'll discuss this issue more below.

The second important concession to make, however, is that there has been enormous progress made in formalizing various forms of reasoning. Consider, for example, the development of deductive logic. Obviously, students in logic

courses are learning something, and they are learning something valuable about the nature of deductively valid reasoning. But it would surely be a mistake to infer that because they make discoveries about how one does and should reason, they weren't actually engaged in such reasoning before they made those discoveries. People have been reasoning in accordance with *modus ponens* and *modus tollens*[13] long before they learned to formalize those rules. People have been reasoning inductively long before they read Russell's (1912) discussion of induction. I would even argue that people have reasoned in accordance with Bayes' theorem even if the formalization of that reasoning would strike them as unfamiliar or, even, not particularly intuitive.

Consider the following two analogies. The rules of syntax we follow when we speak are very complex, and it is no insignificant challenge to give a formal description of those rules. Philosophers of linguistics spend careers arguing about just what counts as syntactically well-formed and what doesn't. Fortunately, however, we don't need to wait until we discover the rules of grammar to speak grammatically. Six-year old children who have been raised by people who speak relatively well will themselves speak relatively well. I would say something very similar about the semantic rules we follow. People use meaningfully expressions like "red," "know," "good," "cause," and "right." But as philosophers discover rather quickly, it is notoriously difficult to figure out just how those expressions are used. Put another way, it is notoriously difficult to figure out what semantic rules we follow in using the words of our language.

The examples of syntactic and semantic rules are useful to consider in recognizing a way in which the formalization of reasoning can be enormously helpful. I suggested above, that even young children can speak and write grammatically. But as we all know from reading student papers, these abilities come in degrees, and when a sentence or string of sentences become complicated it is easy to make mistakes. Those who have learned how to "diagram" sentences will probably work their way through complex sentence structures in ways that those who haven't might not. And in complicated descriptions of the world, those who study the meaning of terms and sentences might be able to avoid errors that those who haven't probably won't. The same is true of both deductive and non-deductive logic. I don't want to minimize at all the value of coming to know what rules govern syntax, semantics, and reasoning. I only want to insist that *knowing what* those rules are is not a necessary condition for successfully *following* those rules.

One might argue, that while the above claims are at least initially plausible, we are missing an obvious way of demarcating scientific reasoning from more mundane everyday reasoning in which people engage. Perhaps, scientific reasoning is in some way more rigorous precisely because the scientist has thought about and at least formed justified beliefs about what the correct rules of reasoning are. As a result, they, like those well-trained in grammar, are less prone to engage in fallacious reasoning.

I don't think the above claim is correct. It might be true that *philosophers* of science (or more generally, epistemologists) have tried (with questionable success) to uncover the correct rules of inference. But a scientist isn't a philosopher of science, and I would no more trust a scientist's view about epistemic rules than I would trust a scientist's view about the philosophical problem of perception. The distinction between knowing what the rules of evidence are and following those rules applies to scientists just as much as it applies to ordinary people.[14]

Plus ça change, plus c'est la même chose (the more it changes, the more it is the same thing)

So if the internalist's approach to understanding non-deductive reasoning is correct, what reason is there for supposing that people reason any differently today that they did one hundred, one thousand, or one million years ago? A thousand years ago, people didn't reach conclusions about the acidity of a solution from the color of litmus paper in that solution. For one thing, they didn't have litmus paper. But, my suggestion is, people *today* don't infer that a solution is acidic from the fact that the litmus paper turned red – at least not from that fact alone! They realize that they need at least one additional premise – the premise that there is a strong correlation between the color of the litmus paper and the acidity of the solution! And that premise would, of course, need to be justified if they are to reach a justified belief in a conclusion based on that premise. How did people devise that method of testing solutions for acidity? I don't really know, but I would imagine that they employed something like Mill's (1843) methods.[15] They would, of course, need some independent test for acidity, but that they almost certainly had.

What about our entrails readers, our astrologers, and those who consulted oracles? It is almost certain that they also realized that their reasoning was enthymematic – that they were relying implicitly on premises that asserted causal connections between the characteristics of entrails, the positions of planets, and the reliability of oracles for the respective conclusions they reached. Did they explicitly *formulate* the relevant premises, let alone challenge the justification they might have had for believing them? I don't know. Probably many did not. But the point is that once the relevant questions were asked, they would almost certainly have recognized the legitimacy of the questions. At least they would have if they were the slightest bit rational.

I have no interest in being an apologist for the rationality of human beings. In the past and to this day there are all sorts of people who reason from various premises to a conclusion without having the slightest reason to believe the relevant premises. People who concluded that someone was a witch because that person floated instead of drowning in water clearly needed some reason to think that such facts were relevant indicators of demonic associations – they *knew* they needed such evidence. They didn't have any reason to support the crucial premises and were for that reason irrational. But there is no reason to suppose that they were committed to some sort of *reasoning* that has since been abandoned.

But Don't We Have New Ways of Supporting Premises?

But don't the above observations suggest, at the very least, that people in the past who accepted such bizarre premises must have employed some sort of equally bizarre reasoning in formulating the premises (however mundane the reasoning might have been from those premises to a conclusion). Perhaps. But we would need to look at the cases one by one. Often, many people reach conclusions by relying on the testimony of others, where the testimony here involves nothing more than hearing what other people say. Again, I think that relying on testimony involves enthymematic reasoning with unstated premises indicating that what most people confidently say has a high probability of being true (Fumerton 2006). The more one learns about the world, the more cautious one becomes with respect to relying on testimony. At the very least, one learns that caution is sometimes the best course of action given certain subject matters of testimony. While he overstated his conclusion slightly, René Descartes (1960, p. 8) was close to being right when he observed of philosophy that "it has been studied for many centuries by the most outstanding minds without having produced anything which is not in dispute and consequently doubtful and uncertain." Unless one has reason to think that one person's "testimony" is better than another, one is hardly in a position to rely on testimony.[16] But the disagreement that characterizes philosophy is hardly restricted to that field. It is endemic to economics, sociology, psychology, political science, and even physics once it ranges beyond the practical to the theoretical.

Still progress has been made. And surely that progress was possible only through new ways of reasoning. Again, I'm not sure. We must, of course, distinguish the context of discovery from the context of justification. But even discovery doesn't consist in wild guesses. It relies on imagination, imagination fueled by, among others things analogy. Grover Maxwell's (1962) old "science fiction" story of the discovery of "crobes" isn't far off the mark when it comes to how someone might get a fruitful idea. Maxwell, you probably recall, described a scientist trying to figure out how and why disease was spread at a distance in hospitals. The scientist postulated tiny organisms (he called them "crobes") too small to see, but capable of carrying disease from one place to another. Maxwell told the story in the context of evaluating the question of whether there is a viable distinction between the theoretical and the observable. But what is important to me here is that our scientist is almost certainly employing a kind of analogical reasoning. People already knew that disease was spread by vermin (like rats). The theorist couldn't find any visible vermin and postulated an invisible one. But analogical reasoning of this sort is really just a form of inductive reasoning. We find that in a lot of cases an organism spreads disease from one place to another. We find again a spread of disease and infer that an organism is responsible. It took the invention of the compound microscope to confirm the hypothesis, but the hypothesis had inductive support before then.

The point is that inductive reasoning is potentially very powerful. It can take us not only from observed correlations in the past to projected correlations in an as

yet unobserved future. But it can also take us from observed correlations to similar correlations in an as yet unobserved, possibly never to be observed, world of theoretical entities. I can't convince you that all of science proceeds this way without going through the development of all scientific hypotheses. I'm content to make the more modest claim that there is no reason to believe that this hasn't been the *modus operandi* of people trying to make new discoveries from time immemorial.

Conclusion

In this chapter, I have argued that there is no strong reason to suppose that over the centuries we have changed the way in which we reason. At least that is so if we make a critical distinction between enthymematic reasoning and fundamental reasoning. We have found new instruments with which to measure and predict all manner of natural phenomena. But we have found those instruments using the same fundamental kinds of deductive and non-deductive reasoning that conscious beings have used since the beginning of time. There is a sense in which that doesn't denigrate at all the various sciences. They have used the inferential tools at their disposals to discover those new correlations that allow us to reach conclusions we were previously unable to make. We owe them great appreciation for those advances. But we don't need to pretend that their discoveries, important as they might be, employed anything but the same reasoning that allows the rest of us to navigate the world.

Notes

1 That's probably not quite right. When in one of his movies John Wayne says "That man is no man" he presumably isn't contracting himself. He is, rather, equivocating on the meaning of "man." In evaluating apparently deductively valid arguments we need to guard against such equivocation.

2 Epistemologists understand the internalism/exeternalism controversy in epistemology in different ways. The terminology itself suggests that the internalist is committed to the view that whether there is justification for one to believe some proposition P depends soley on the internal states of that person. The externalist, by contrast, thinks that whether or not one is justified in believing P might depend on such external factors as the causal history of a belief (what caused one to have the belief). But this way of understanding the controversy just scratches the surface. We need to figure out what makes a state of a person an internal state – there is another controversy about this in the philosophy of mind. Still other internalists seem to think that the key issue is whether or not one ties justification to access. So on one internalist view, X can justify someone S in believing P only if S is, in some sense aware, or has the capacity to be introspectively aware, of the fact that S is in X and that being in X makes likely P. The short answer to the question of how to define the internalism/externalism debate is that there is no short answer.

3 A foundationalist is committed to the view that if there are any justified beliefs, they can all be inferred through a chain of inference from beliefs that are justified without inference.

4 Jennifer Wilson Mulnix (2008) points out a complication. When we "intuit" what we believe, the reliabilist will presumably describe the input that results in the metabelief that we have the belief in question as the very belief that is the content of the metabelief. But this is still a foundationally justified belief – the process is unconditionally reliable and the epistemic status of the first level belief is irrelevant to the epistemic status of the metabelief.

5 The reliabilist will almost certainly move to something more sophisticated than this. At the very least, reliability will be spelled out in terms of ratios of true to false output beliefs if the process were employed indefinitely many times.

6 *Modus ponens* is necessarily conditionally reliable. An inference from (P and Q) to not-Q is necessarily conditionally unreliable.

7 On one way of thinking about it, the reliabilist is trying to combine two obvious ideas. One is that one can have an inferentially justified false belief. The other is that justification should have something to do with truth. Reliable belief-producing processes (crudely understood) guarantee that most justified beliefs are true. But reliability also allows for the fact that a reliable process can produce some false beliefs. All this is much more complicated – there are concepts (like the concept of a reliable process) that need to be defined.

8 The reliabilist needs a solution to the famous generality problem – the problem of deciding which of the infinitely many types a process belongs to is the one critical to evaluating the reliability of "the" process resulting in the output beliefs.

9 One of the best discussions of reasoning to the best explanation in recent years is McCain (2014).

10 See, for example, Fumerton (1980). Of course, it should be noted that others suggest precisely the opposite – that inductive reasoning is disguised reasoning to the best explanation, see Harman (1965). And one might even suggest that deductively valid reasoning is always inductive reasoning (see Mill 1843).

11 Huemer argues that what seems to be the case is different from what one believes or even is inclined to believe. Consider the Muller-Lyer illusion where even when one knows that the line segments are of equal length, one line seems to be shorter than the other.

12 For a defense of this sort of position, see Huemer (2002).

13 *Modus ponens* is the general rule that one can deduce Q from P and (if P then Q). *Modus Tollens* is, I think, the philosopher's favorite form of reasoning: *Modus Tollens* states that from not-Q and (if P then Q) one may infer not-P.

14 When one reads popular accounts of highly theoretical physics one starts to wonder if those engaged in such speculation have any idea of what would constitute a legitimate inference. On the other hand, without being a theoretical physicist, it's hard to know whether the "accessible" accounts of such theories are accurate.

15 Mill's methods include classic enumerative induction; the method of agreement:

> If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause (or effect) of the given phenomenon; the method of difference: If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon; the method of residues: Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the

remaining antecedents; and the method of concomitant variations: Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner – is either a cause or an effect of that phenomenon or is connected with it through some fact of causation. For a more detailed discussion of these methods (and the question of whether some may be reduced to others), see Donner and Fumerton (2009, pp. 168–173).

16 For a defense of this view, see Fumerton (2010). For a range of other views, see the papers contained in Warfield and Feldman (2010).

References

Descartes, Rene. 1960 [1641]. *Discourse on Method and Meditations*. Translated by Laurence Lafleur. Indianapolis, IN: Bobbs-Merrill.

Donner, Wendy and Richard Fumerton. 2009. *Mill*. Malden, MA: Wiley Blackwell.

Fumerton, Richard. 1980. "Induction and Reasoning to the Best Explanation." *Philosophy of Science* 47: 523–38.

Fumerton, Richard. 1995. *Metaepistemology and Skepticism*. Lanham, MD: Rowman and Littlefield.

Fumerton, Richard. 2006. "The Epistemic Role of Testimony: Internalist and Externalist Perspectives," in Jennifer Lackey and Ernest Sosa (eds.), *The Epistemology of Testimony*. New York: Oxford University Press: 77–92.

Fumerton, Richard. 2010. "You Can't Trust a Philosopher," in: Ted Warfield and Richard Feldman (eds.), *The Epistemology of Disagreement*. New York: Oxford University Press: 91–111.

Fumerton, Richard. 2019. "Inferential Justification and the Problem of Unconscious Inference," in: Kevin McCain and Ted Poston (eds.), *The Mystery of Skepticism: New Explorations*, Boston, MA: Brill.

Goodman, Nelson. 1955. *Fact Fiction and Forecast*. Indianapolis, IN: Bobbs-Merrill.

Harman, Gilbert. 1965. "The Inference to the Best Explanation." *Philosophical Review* 74: 88–95.

Huemer, Michael. 2001. *Skepticism and the Veil of Perception*. Lanham, MD: Rowman and Littlefield.

Huemer, Michael. 2002. "Fumerton's Principle of Inferential Justification." *Journal of Philosophical Research* 27: 329–340.

Maxwell, Grover. 1962. "The Ontological Status of Theoretical Entities," in Herbert Feigl and Grover Maxwell (eds.), *Scientific Explanation, Space, and Time: Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press: 181–192.

McCain, Kevin. 2014. *Evidentialism and Epistemic Justification*. New York: Routledge.

Mill, J. S. 1843. *A System of Logic*. London: Parker.

Mulnix, Jennifer Wilson. 2008. "Reliabilism, Intuition, and Mathematical Knowledge." *Filozofia* 62(8): 715–723.

Peirce, C. S. 1938. *Collected Papers*. Charles Hartshorne and Paul Weiss (eds.) Cambridge, MA: Harvard University Press.

Reid, Thomas. 1854. *Essays on the Intellectual Powers* in *The Works of Thomas Reid*, 4th edition, Sir William Hamilton (ed.) London: Longmans, Green and Company.

Russell, Bertrand. 1912. *The Problems of Philosophy*. London: Williams & Norgate.

10

CAN SCIENTIFIC KNOWLEDGE BE MEASURED BY NUMBERS?

Hanne Andersen

Introduction

Scientific knowledge is an important basis for our society. But not all scientific knowledge is equally important, and not all scientists are equally good scientists. In deciding which scientists to hire, which journals to acquire for a library, or which publications to read, various numerical indicators are often used to measure quality, impact, or relevance. This chapter provides a short overview of the most popular indicators, such as the h-index and the journal impact factor, and describe how they have been developed. On this basis, I discuss what the various indicators actually measure, where and to what extent it makes sense to draw on such numerical indicators, and where they may lead us astray.

Economic Indicators

Over the last century, science has become increasingly important to society. Today, most countries allocate a substantial amount of their state budget to the production of new scientific knowledge and to the education of new generations of scientists.[1] As public investment in science has increased, politicians, taxpayers, agencies and managers have become increasingly interested in how the funds allocated for research are spent, and what society gets in return.

Interest in measuring research and development activities (R&D) can be found already in the first half of the 20th century (Godin 2002). But it was especially after World War II that policy makers began seeing a strong linkage between economic performance and performance of science and technology. This linkage was explicitly expressed in the seminal report *Science: The Endless Frontier* that science advisor Vannevar Bush delivered to US President Truman shortly after the end World War II. There, Bush described how.

> Advances in science when put to practical use mean more jobs, higher wages, shorter hours, more abundant crops, more leisure for recreation, for study, for learning how to live without the deadening drudgery which has been the burden of the common man for ages past. Advances in science will also bring higher standards of living, will lead to the prevention or cure of diseases, will promote conservation of our limited national resources, and will assure means of defense against aggression.
>
> *(Bush 1945)*

In order to achieve these objectives, Bush concluded that the flow of new scientific knowledge had to be both continuous and substantial.

Seeing science as a major driver of economic growth, it became important to monitor national and international performance. During the 1950s, many countries began making surveys of their R&D activities, but differences in the concepts and methods employed made comparisons difficult. By the late 1950s, the Committee for Applied Research of the European Productivity Agency began discussing various definitions and methods, and by the early 1960s this developed into the so-called *Frascati Manual* that outlined a set of standardized definitions of how to understand research and development, distinguish between different sectors, and how to measure personnel and expenditure devoted to these categories (de la Mothe 1992).[2]

Much emphasis in these attempts at monitoring national performance in science and technology has been on input indicators, especially the financial and human resources devoted to R&D. Hence, the *Frascati Manual* stipulated how to define basic research, applied research, and development of products and processes; and how to measure the allocation of monetary and human resources by disciplines or industrial sectors. This focus on input fits well with a science policy that expected increased activity in science to lead more or less automatically to an increase in economic performance and improved standards of living (Godin 2005, ch. 7).

Output indicators, on the other hand, were more difficult to define. One of the first attempts at measuring the outcome of science came from the historian of science Derek de Solla Price. In his two monographs *Science Since Babylon* (Price 1961) and *Little Science, Big Science* (Price 1963), Price wanted to develop a science of science; to find the underlying principles and laws that could guide science policy. For example, examining a variety of variables, including the number of journals, the number of published papers, the number of graduates, the operating energy of particle accelerators, or the number of discovered chemical elements, he concluded that science grew exponentially. But this result also included a dire warning for the future. As Price graphically pointed out, "To go beyond the bounds of absurdity, another couple of centuries of 'normal' growth of science would give us dozens of scientists per man, woman, child and dog of the world population" (Price 1961, p. 13). Hence, similar to many phenomena in biology

and epidemiology, Price predicted the future development of science to follow a sigmoid or logistic curve. According to this view, future experiences of shortage in manpower or funding would not be an "incidental headache" that could be easily cured. Instead, Price warned, "We must not expect such growth to continue, and we must not waste time and energy in seeking too many palliatives for an incurable process" (Price 1961 p. 117).

While Price's work was groundbreaking in its attempt at measuring the growth of science, it also illustrated that the output and outcome of science could be understood in many different ways. Nevertheless, science policy makers who wanted to ensure that the investment in science was made in a way that maximized the returns to society needed some kind of indicators to measure output and outcome. Hence, countries and organizations began supplementing the input indicators on expenditure and manpower with indicators of the direct results of the scientific activity, such as patents, as well as the effects that science had on society, such as high-technology trade or balance in payment for patents, licenses, and know-how.[3]

Bibliometric Indicators

Whereas science policy focuses on the economic indicators, the primary interest within academia itself is typically on bibliometric indicators, based on publication numbers and citation counts. Methodologically, there is a major difference in the accessibility of these two indicators. An author can always count the number of publications that he or she has authored and document this number by presenting a publication list with the bibliographic information of each publication. In contrast, citations of a publication may be made by many different people and through many different channels, and so there is no unique overview of these in the same way an author has an overview of his or her own publications.

Citation Databases

An important prerequisite for indicators that involve the number of citations is databases that track citations of academic publications. In creating such databases, the ideal of making an exhaustive database that includes all of the world's academic journals and books needs to be balanced against what is feasible. Decisions on what to include or not to include will depend on the intended use of the database. If the aim is to use citations as an indicator of overall influence, it is desirable to include as many journals as possible for as long as possible. If, instead, the aim is to use citations as an indicator of the expected use of a journal in a library, it may suffice to include only the major channels. Especially before the digital age, when building a database of citations required going through the print volumes of each issue of each journal and manually creating a record for each citation, pragmatic decisions needed to be made about which journals to include and for how many previous years.

Once such a database is established, individual publications can be uniquely identified by bibliographic information, and the number of citations of each publication can be counted. However, since there may be multiple authors with identical names, it is difficult to identify authors in a unique way. In most cases, they can be distinguished from one another by adding the institutional affiliation to the identifier, but that also requires a procedure for keeping track of changes in affiliations as people change jobs. Before the digital age, this could be challenging, but now services such as ORCID (Open Researcher and Contributor Identifier) offers solutions to name ambiguities by assigning unique identifiers to authors (Haak et al. 2012).

The first attempt at producing a database of scientific citations was the Scientific Citation Index (SCI), created by Eugene Garfield in the late 1950s and launched as a commercial product by a company called the "Institute for Scientific Information" in 1964. From the outset, the index was developed as a tool for assessing journals as well as scholars and publications, and this came to influence the construction of the database in important ways.

Initially, Garfield had seen the index as an "association-of-ideas" index that provided a complete listing of all the publications that had ever referred to a particular publication (Garfield 1955). In Garfield's view, this would minimize the risk of drawing on fraudulent, incomplete, or obsolete research. However, this use of the index was later downplayed, and the ideal of exclusiveness relinquished.

Instead, to make the creation of the database practically feasible, Garfield followed the pragmatic approach of the existing subject indexing services that analyzed only those journals that were considered most important within a particular discipline. Garfield justified this selectivity by empirical studies showing that about 20% of the journals covered by the database received about 80% of the indexed citations, while about 40% of the journals published about 80% of the indexed publications (Garfield 1990, 1996). Based on these results, he argued that even if the database were limited to the top 500 journals, it would still provide a comprehensive coverage of the most important publications.[4]

In making decisions about which journals to include in the database, Garfield argued that citation data, expert judgment, and journal standards should all be taken into account. Of these factors, the most basic criterion that a journal must fulfill in order to be included in the index is that it is published according to schedule. Garfield's argument for this criterion was that it was "unethical and unacceptable for publishers to allow journals to appear chronologically late" (Garfield 1990). This is not a criterion based on the quality of the content. But because average citation rates over a strict 2-year window was the basis for the index assessment of journals, timely publication was crucial for the reliability of this measure.

Originally, the SCI covered only the natural and medical sciences, while an Arts and Humanities Index as well as a Social Science Index were later added to what is now known as the Web of Science (WoS). Journals from these fields have only gradually been added to the database, and coverage in the WoS is still much

stronger in the natural and biomedical sciences than in the humanities and social sciences.

In recent years, various alternatives to the WoS have emerged, such as Scopus provided by the publishing house Elsevier. Different databases may differ in which journals they cover, how far in the past they go to draw content to index, whether they index books, which publishing houses they cover, and whether they index additional types of publications such as conference series. This means that different databases may yield very different citation counts for the same publication or the same author. Similarly, as a database includes, for example, new types of material, or extends its historical coverage, additional citations may be added. Hence, when two consecutive searches in a citation database show an increase of citations of a particular publication or author, this does not necessarily reflect new citations but may instead reflect recent additions of old citations to the database.

Citation databases such as the WoS or Scopus have detailed policies for which publication channels to include or exclude. This is different from Google Scholar that uses automated software to index publications found on websites, including preprint servers, university repositories, personal web pages, and so on. In most fields, Google Scholar provides a much broader coverage than the standard citation databases, however what has been covered is less transparent. It is therefore a matter of debate whether citation counts from Google Scholar bear the same claim to legitimacy as citation counts from the WoS or Scopus. On the one hand, Google Scholar includes citations of a non-academic nature, such as those from blogs, preprint servers, and similar venues. On the other hand, such citations may still indicate a relevant form of impact.[5]

Studies of Citations and Stratification

Since the emergence of the SCI, it has been a matter of debate as to how far citations can be used as indicators for the impact or quality of individual researchers. When the SCI first emerged, sociologists of science interested in stratification and reward soon began investigating correlations between authors' numbers of publications and citations and their recognition as reflected by, for example, their having received prestigious awards or positions, or being widely known among peers. For example, studying a population of 120 physicists, the sociologists Cole and Cole examined whether the doctrine of "publish or perish" actually holds, i.e., whether scientists who publish many trivial papers are rewarded, while scientists who publish only a few high quality papers are not (Cole and Cole 1967). Comparing publications and citations to rewards and recognition they found that awards, prestigious positions, and being well-known to peers was more strongly correlated with the number of citations than the number of publications, especially for scholars employed at prestigious institutions. Cole and Cole therefore argued that citations seemed to be a better indicator of researchers' overall success than the sheer number of publications.

However, the overall correlation between citations and reward varied considerably, depending on reward type. Hence, citations can be useful for examining overall stratification patterns in a population, but when applied to the individual, predictions of recognition from the number of citations are quite uncertain. Some sociologists of science who used citation data to study stratification therefore at the same time warned against using citation data in promotion and hiring processes (Wouters 1999, 102).

Nevertheless, bibliometric tools have become increasingly popular in assessing individual researchers, and a plethora of different metrics have been developed. The following section therefore describes some of the most popular indicators, before returning to various points of criticism.

Individuals' Productivity and Impact

The most basic measure in assessing individual scientists is the number of authored publications. Usually, this is seen as an indicator of the individual researcher's productivity. However, if understanding productivity as efficiency in converting input into output, the number of publications needs to be seen in relation to the amount of time that has been spent on producing the publications. Further, many studies document how variation in time expenditure and research output are influenced by such factors as gender, parental status, or ethnicity, and how such differences may affect the career patterns of underrepresented groups (Bellas et al. 1999).

Second, publications vary in kind. Agencies and institutions interested in assessing researchers may therefore request publication numbers to be divided into different categories according to their particular interpretation of publications as research output. How different types of publications are weighed against each other may vary from field to field and from institution to institution. For example, monographs tend to play an important role in the humanities and social sciences (Ochsner et al. 2016; Williams et al 2018), while conference publications tend to have a high status in computer science (Freyne et al. 2010).

Third, many publications have several authors who have shared the workload of producing them. To adjust for this, a single-author equivalent number of publications can be computed by dividing each publication by the number of authors and then summing the fractional author counts (Carbone 2011). However, this assumes that labor has been uniformly distributed among all co-authors. Alternatively, a weighted fractional output (WFO) can be computed by weighing each author's fraction of the paper according to the institutional distribution of the authors and their order on the byline (Abramo, D'Angelo, and Rosati 2013a, Abramo, D'Angelo, and Rosati 2013b).[6] Yet, this measure also builds upon fixed assumptions about the distribution of labor between different types of authors and collaborations, where, in practice, this varies considerably from case to case.

Sometimes, co-authors may be included on the byline although their contribution to the publication has been only marginal or even non-existent. Such

gift authorship is often granted as a favor or out of courtesy, such as, for example, when the head of a laboratory is listed as co-author on all papers produced by employees at the lab. Sometimes gift authorships are granted in an attempt to increase the visibility of a paper by adding a famous scholar as co-author. However, in both cases gift authorships make it unclear who is responsible for the research reported in the publication, and it confers credit to people who have not earned it. Gift authorships are therefore considered a questionable research practice. In an attempt to promote responsible authorship practices, the International Committee of Medical Journal Editors (ICMJE) has issued a set of guidelines that stipulates what is required to be listed as co-author to an academic paper. According to these guidelines, authorship should be based on the following four criteria:

- substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; and
- drafting the work or revising it critically for important intellectual content, and
- final approval of the version to be published and
- agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved" (ICMJE 2017, p. 2).

These guidelines have gradually been adopted by many journals, also outside the biomedical disciplines. However, studies also show that the Vancouver guidelines are often violated. For example, in a survey conducted by Martinson et al. (2005) among more than 3,000 US scientists 10% reported they had inappropriately assigned authorship credit. Similarly, in a survey among hyperprolific authors, Ioannidis et al. (2018) found that more than two-thirds of the respondents failed to meet all the criteria stipulated by the Vancouver guidelines more than 25% of the time, and some of the respondents argued that there should be different levels of authorship.

The measures of productivity described above are all based on treating publications as an output parameter. But ideally, the output of research is new knowledge, or new ideas. Publications are merely a dissemination channel for this knowledge, and how much new knowledge individual publications present varies considerably. Sometimes authors may distribute their recent results in several separate papers instead of a single, comprehensive publication, or they may publish the same article more than once. Such practices, often referred to as "salami publication" and "duplicate publication", are often considered questionable research practices.[7]

Since a large output of publications is not necessarily equivalent to a large output of new knowledge, additional measures are needed in order to assess whether a scientist has contributed something of value. Often, the number of citations is used for this purpose.

Sometimes the number of citations of a publication is seen as a measure of the publication's quality. However, scientific results can be of a very high quality without necessarily gaining a high number of citations. There are several reasons for this. First, they may not be discovered by other scientists, for example because they are published in a publication channel that only a few libraries subscribe to. Second, their importance to other areas of science may not be discovered until a long time after they were published. Third, they may be brilliant within a very small field, with only a few other researchers to produce citations.

Citations can also be seen as reflections of influence on subsequent research, or impact. On this interpretation of citations, the number of citations to publications that a researcher has authored or co-authored is seen as an indicator of the researcher's total impact. However, impact on subsequent research can mean many different things, and these are not necessarily valued in the same way. For example, a methodological refinement may be widely adopted and therefore receive a very high number of citations, but it is often seen as less important than, for example, new conceptual developments.

Whether the number of citations is seen as an indicator of quality or impact, it has been argued that ranking scientists by citation numbers alone does not distinguish between scientists who have published just a few highly cited publications and scientists who have a stable production of well-cited publications. The so-called h-index was introduced to distinguish between these two situations (Hirsch 2005). The h-index is defined as the number of publications, n, that an author has published which have each been cited at least n times. The n most cited papers are called the author's h-core. Thus, if an author has an h-index of 25, this means that his or her 25 most cited publications each have 25 citations or more.

However, as argued by Gingras (2014b), heterogeneous indicators combining characteristics often fail to be sensitive to variations in the phenomenon that these indicators purport to measure. For example, a researcher who has published ten papers that have each received ten citations, a researcher who has published hundred papers that have each received ten citations, and a researcher who has published ten papers that have each received hundred citations will each have an h-index of ten – although they are quite different with respect to their productivity and impact, and will probably not be ranked as equal in an assessment.

The difficulty in including both productivity and impact in the same indicator is reflected by the plethora of alternatives that have been proposed in order to represent some particular aspect of productivity and impact that the h-index seems to ignore.[8]

First, the h-index ignores citations outside the h-core, so an author who has published ten publications that have each received ten citations and an author who has published ten publications that have each received hundreds of citations will have the same h-index. Some alternative indicators therefore grant more weight to highly cited papers. This includes the g-index (Egghe 2006) that is defined as the largest number, g, for which the g most cited publications from an author

have together received at least g^2 citations, and the e-index (Zhang 2009) that is defined as the square root of the excess citations of the publications included in the author's h-core.

Second, since the h-index is purely cumulative, it favors researchers who have published many papers over a long time. The m-quotient (Hirsch 2005) adjusts for academic age by dividing the h-index by the number of years that have passed since the author's first publication appeared, while the hIa-index (Harzing, Alakangas, and Adams 2014) adjusts for both co-authorship and career length by normalizing citations for each paper by dividing the number of citations by the number of authors for that paper, calculating the h-index from these normalized citation numbers (called the hI-norm), and then dividing by the number of years that the author has been publishing. In this way, the m-quotient and the hIa-index both attempt to adjust for the length of the author's career, but by doing so they also penalize authors who published their first paper very early in their career.

As this brief description of just a few of these alternative indicators indicates, there is no unique way of balancing productivity and impact that fits with all intuitions of how researchers compare to one another.

Measuring Scientific Literature

Bibliometric indicators are not only used to measure researchers, but they are also used to measure scientific literature. Most visibly, citations are used to compute the so-called journal impact factor (IF) that is used to rank scientific journals. Less visibly, citations are also used to rank individual publications when searching for literature in academic databases or by using Google Scholar to search for literature on the internet.

Citations and the Journal Impact Factor

The journal impact factor is computed by the WoS as the average number of citations per article over the last two years. In this way, the IF can be seen as an indicator of the average impact of publications in the journal, but sometimes the IF is also interpreted as an indicator of the journal's importance or quality. Rankings of journals within a field according to their IF, for example, offered by the WoS, contributes to this image of the IF as an instrument for journal quality assessment. This interpretation also sometimes causes authors to flag the IFs of the journals in which they have published. However, the IF merely conveys a computed average for the entire journal and cannot be used as an indicator of the impact or quality of an individual publication (McNutt 2014).

Impact factors vary considerably across fields. First, citation practices vary from field to field. In some fields it is important to cite a broad range of recent literature, while in others it is valued to cite only a few classic publications. Second, how quickly published articles are cited varies, and therefore the fixed two-year

window favors some fields over others. Further, citations are only counted from journals included in the WoS, and this favors English-language publications over non-English language publications, as well as the natural and health sciences over the humanities and social sciences. This all contributes to the differences in what is considered a high or a low impact factor. For example, in medicine, some journals have IFs above 50; in philosophy of science, they rarely exceed 1.[9]

The impact factor can also be manipulated in several different ways (Falagas and Alexiou 2008; Archambault and Larivière 2009). Increasing the share of review articles, publishing those articles that are most likely to receive citations in the first issue of a volume, or having articles online for a long time before print publication and final volume assignment are all editorial strategies that can be used increase the impact factor of a journal. Encouraging authors to cite recent papers from the journal is another way in which editors may improve the impact factor of their journal.

Citations and Literature Searches

Citation numbers have also come to play a crucial, but largely invisible, role with respect to how scientific literature is located. When an electronic database or search engine is used for searching for scientific literature, the results of this search need to be displayed in some order. This holds whether the results come from a search engine that searches the internet broadly, such as Google Scholar, or search functions in specific databases covering, for example, a specific academic publisher or publications within a particular discipline.

Many databases allow the user to choose between different ranking algorithms, for example whether they want results sorted by publication date, number of citations, number of downloads, or relevance. However, relevance is an elusive concept. Although most databases offer to rank results according to relevance, they rarely describe how the measure of relevance is computed. This leaves opaque for the user the basis on which individual publications have acquired their position on the list. For example, on the "About Google Scholar" web page (https://scholar.google.com/intl/en/scholar/about.html) it is stated that "Google Scholar aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature". No further information is provided about how these various factors enter the ranking algorithm. Computer scientists attempting to reverse-engineer Google Scholar's ranking algorithm have found that citation counts had the highest weight among the potential factors they examined, and that other factors included whether the search term occurred in the title and whether the article was published recently (Beel and Gipp 2009). On this basis, they concluded that Google Scholar is more suitable for finding standard literature than for finding recent trends or for finding publications advancing alternatives to mainstream views.

As for other bibliometric measures, these rankings can be manipulated. For example, Beel and Gipp (2009) concluded from their studies of Google Scholar's ranking algorithm that an article will be more retrievable in Google Scholar if its authors refrain from a strict terminology and instead alternate between as many synonyms as possible.

Criticism of Bibliometric Indicators

Indicators are variables used to measure a phenomenon by means of proxies when the phenomenon itself cannot be measured directly (Lazarsfeld 1958). In designing an indicator, it needs to be clear what phenomenon the indicator purports to measure, and the indicator must be sensitive to variations of the phenomenon being measured (see Gingras 2014a, b for details). However, for bibliometric indicators this has turned out to be highly complex.

As described in previous sections, it is not well defined what the number of publications and citations measure. Although the number of publications is related to productivity in some sense, at the same time, as a measure of productivity, it has two important shortcomings: first, it focuses on quantity regardless of quality, and second, if productivity is to be understood as efficiency, it is unclear how to adjust for time or workload spent in producing the output.

Similarly, although the number of citations is related to impact, importance or quality in some sense, at the same time, as an indicator it also has a number of shortcomings. First, citations indicate reception. In contrast, the quality of a publication or of a scholar is often understood as inherent quality that is independent of whether the publication has been received or the scholar has been noticed by others. Citations should therefore be seen as an indicator of impact rather than of quality. Second, even if understanding citations as an indicator only of impact, it is important to remember that impact reflects how scientific achievements are received. It is therefore also dependent on the recipients. For example, a result that is ahead of its time may remain more or less uncited for a long time. Third, some achievements which have a high impact, such as, for example, new laboratory techniques or new experimental procedures, may not necessarily be as highly regarded as new theoretical or conceptual developments.

In general, indicators that correlate to some degree with the phenomenon they purport to measure may be useful for vindicating other observations regarding the phenomenon, or for investigating the phenomenon statistically in a large population. Their real shortcoming is their probabilistic nature when applied to an individual case. For example, the number of publications and citations may be used to investigate gender differences in how rewards are distributed within a population (see, e.g., Wennerås and Wold 1997; Larivière et al. 2013), but it is difficult to use the number of publications and citations to prove bias in an individual case – although examples of such cases have been seen (Wade 1975).

A separate line of criticism concerns the adverse effects of using bibliometric measures to assess scholars. The increasing focus on the number of publications and citations when, for example, distributing research grants or selecting candidates for positions, produces a strong incentive for scholars to boost these numbers. The resulting "publish or perish" culture is often seen as a driver for questionable research practices such as salami publication, duplicate publications, and gift authorships (Anderson et al. 2007; Franzoni et al. 2011), or for participation in "citation rings" in which authors mutually credit each other's work (Biagioli 2016). For some, this just calls for metrics that are more detailed. For example, examining the increase in the number of authors who publish more than one publication per week on average, Ioannidis et al. (2018) have argued that "if adding more authors diminished the credit each author received, unwarranted multiauthorship might go down". Others seek solutions in which qualitative information is brought back into the assessment procedure. For example, some research foundations and agencies ask applicants to supplement the traditional publication lists with a selected list of the 5–10 most important papers, or to provide a description of the most important scientific achievements that a scholar has produced during his or her career.

Due to these shortcomings of bibliometric indicators, scholars from history and philosophy of science (e.g., Gingras 2014a), from mathematics and statistics (e.g., Adler et al. 2009), and from scientometrics, science policy, and related fields (e.g., Hicks et al. 2015; Stephan et al. 2017) call for caution in using bibliometrics for the assessment of individual researchers or individual institutions. Finally, in reaction to the increasing use of bibliometrics in hiring and promotion cases, scholars from various fields have taken initiatives to produce recommendations and guidelines that institutions can accede. Thus, the San Fransisco Declaration on Research Assessment (DORA) presents a set of recommendations for agencies, institutions, and researchers regarding the use of metrics in research assessment, including being explicit about the criteria used in evaluating the productivity of individuals, using a broad range of impact measures and considering the value of all research outputs, and highlighting that "the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published" (see sfdora.org). Similarly, the Leiden Manifesto for Research Metrics present ten principles to guide research evaluation, including that indicators should never substitute informed judgement, but that quantitative evaluation should be used primarily to support qualitative, expert assessment (Hicks et al. 2015).

Notes

1 Data on R&D expenditure for individual countries as well as aggregated world averages can be found at the web page of the World Bank: https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS.

2 For the current version of the *Frascati Manual*, see www.oecd.org/sti/inno/Frascati-Manual.htm.
3 See Godin (2005), ch. 7 for a detailed account.
4 In their description of the journal selection process for the Web of Science, Clarivate still refers to Garfield's Law of Concentration as a principle stating that the core literature for all scholarly disciplines may be concentrated in a relatively small number of journals (https://clarivate.com/essays/journal-selection-process/; accessed October 20, 2018). However, it should be noted that since citation practices vary considerably between fields, a core literature defined from citations risks overlooking small and specialized fields whose research results are published primarily in specialized journals.
5 Some empirical studies indicate that citation analyses conducted using Scopus, Google Scholar and the WoS produce equivalent results. (Harzing and Alakangas 2016; Meho and Yang 2007; Harzing 2013). On the other hand, it has also been shown how Google Scholar can be manipulated to report very high h-indexes by creating a web site with a large number of short articles citing each other; see Labbé (2010) or http://bibliometrie.wordpress.com/2011/05/12/ike-antkare-i-dont-care, accessed Nov. 20, 2018.
6 Abramo and collaborators define the weights such that

> if first and last authors belong to the same university, 40% of the publication is attributed to each of them; the remaining 20% are divided among all other authors. If the first two and last two authors belong to different universities, 30% of the publication is attributed to first and last authors; 15% of the publication is attributed to second and second-last author; the remaining 10% is divided among all others.
>
> *(Abramo, D'Angelo, and Rosati 2013b, p. 201)*

7 The perception of these practices has varied over time. For example, before the digital age, when it could be much more difficult than it is today to locate literature on a particular topic, it was in many fields seen as a legitimate practice to publish the same result in similar forms for different audiences. Today, this may still be legitimate, but there has been an increased focus on the importance of making such duplications transparent through appropriate cross-referencing.
8 For an overview of the many different measures that have been proposed in reaction to the H-index, see, e.g., Garner et al. (2018); Bornmann, Mutz, and Daniel (2008); Harzing (2016); Bornmann et al. (2011).
9 Several attempts have been made to provide journal rankings that can adequately include journals in the humanities and social sciences, including the European Reference Index for the Humanities (ERIH). However, the validity and reliability of this categorization of journals have been questioned (Adler and Harzing 2009), and a large number of journals in the history of science protested against the ERIH ranking when it was first introduced (Andersen 2009). On a national scale, similar attempts at ranking journals into two or three tiers have been made in Norway and Denmark, but a comparison between the categorizations made in the two systems reveals substantial differences in how individual journals are categorized.

References

Abramo, G, CA D'Angelo, and F. Rosati. 2013a. "Measuring institutional research productivity for the life sciences: The importance of accounting for the order of authors in the byline." *Scientometrics* 97:779–795.

Abramo, Giovanni, Ciriaco Andrea D'Angelo, and Francesco Rosati. 2013b. "The importance of accounting for the number of co-authors and their order when assessing research performance at the individual level in the life sciences." *Journal of Informetrics* 7 (1):198–208.

Adler, R., J. Ewing and P. Taylor. 2009. "Citation statistics: A report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)." *Statistical Science* 24(1): 1–14.

Adler, Nancy J., and Anne-Wil Harzing. 2009. "When Knowledge Winds: Transcending the Sense and Nonsense of Academic Rankings." *Academy of Management Learning and Education* 8 (1):72–85.

Andersen, Hanne et al. 2009. "Editorial. Journals under threat: A joint response from history of science, technology and medicine editors." *Centaurus* 51 (1):1–4.

Anderson, M., E.A. Ronning, R.D.Vries, and B.C. Martionson. 2007. "The perverse effect of competition on scientists' work and relationships." *Science and Engineering Ethics* 13: 437–461.

Archambault, Éric, and Vincent Larivière. 2009. "History of the journal impact factor: Contingencies and consequences." *Scientometrics* 79 (3):635–649.

Beel, Jöran, and Bela Gipp. 2009. "Google Scholar's ranking algorithm: An introductory overview." *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, Vol. 1, pp.230–241, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics.

Bellas, M.L, and R.K. Toutkoushian. 1999. "Faculty time allocations and research productivity: Gender, race and family effects." *Review of Higher Education* 22(4): 367–385.

Biagioli, M. 2016. "Watch out for cheats in citation game". *Nature* 535: 201.

Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2008. "Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine." *Journal of the Association for Information Science and Technology* 59 (5):830–837.

Bornmann, Lutz, Rüdiger Mutz, Sven E Hug, and Hans-Dieter Daniel. 2011. "A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants." *Journal of Informetrics* 5 (3):346–359.

Bush, Vannevar. 1945. *Science, The Endless Frontier: A Report to the President*: US Govt. printing Office.

Carbone, Vincenzo. 2011. "Fractional counting of authorship to quantify scientific research output." arXiv preprint arXiv*:1106.0114*.

Cole, Stephen, and Jonathan R Cole. 1967. "Scientific output and recognition: A study in the operation of the reward system in science." *American Sociological Review* 32(3):377–390.

de la Mothe, John. 1992. "The revision of international science indicators: The Frascati manual." *Technology in Society* 14 (4):427–440.

Egghe, Leo. 2006. "An improvement of the h-index: The g-index." ISSI Newsletter. 2.

Falagas, Matthew E, and Vangelis G Alexiou. 2008. "The top-ten in journal impact factor manipulation." *Archivum immunologiae et therapiae experimentalis* 56 (4):223.

Franzoni, C., G. Scellato and P. Stephan. 2011. "Changing incentives to publish". *Science* 333: 702–703.

Freyne, J., L. Coyle, B. Smyth and P. Cunningham. 2010. "A quantitative evaluation of the relative status of journal and conference publications in computer science". *Communications of the ACM* 53(11): 124–132.

Garfield, E. 1955. "Citation indexes for science." *Science* 122:108–111.
Garfield, E. 1990. "How ISI selects journals for coverage: Quantitative and qualitative considerations." *Current Contents* 22:5–19.
Garfield, E. 1996. "The significant scientific literature appears in a small core of journals." *The Scientist* 10 (17):13–15.
Garner, Rebecca M, Joshua A Hirsch, Felipe C Albuquerque, and Kyle M Fargen. 2018. "Bibliometric indices: Defining academic productivity and citation rates of researchers, departments and journals." *Journal of Neurointerventional Surgery* 10(2):102–106.
Gingras, Y. 2014a. *Bibliometrics and Research Evaluation: Uses and Abuses* Cambridge MA: MIT Press.
Gingras, Y. 2014b. "Criteria for evaluating indicators," in: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, edited by Blaise Cronin and Cassidy R Sugimoto. Cambridge MA: MIT Press.
Godin, B. 2002. "The number makers: Fifty years of science and technology official statistics." *Minerva* 40 (4):375–397.
Godin, B. 2005. *Measurement and Statistics on Science and Technology*. New York: Routledge.
Harzing, Anne-Wil. 2013. "A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel prize winners." *Scientometrics* 94 (3):1057–1075.
Harzing, Anne-Wil. 2016. "From h-index to hIa: The ins and outs of research metrics." *Research in International Management*. https://harzing.com/publications/white-papers/from-h-index-to-hia. Accessed 20 November 2018.
Harzing, Anne-Wil, and Satu Alakangas. 2016. "Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison." *Scientometrics* 106:787–804.
Harzing, Anne-Wil, Satu Alakangas, and David Adams. 2014. "hIa: An individual annual h-index to accommodate disciplinary and career length differences." *Scientometrics* 99 (3):811–821.
Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah De Rijcke, and Ismael Rafols. 2015. "The Leiden Manifesto for research metrics." *Nature* 520 (7548):429.
Hirsch, Jorge E. 2005. "An index to quantify an individual's scientific research output." *Proceedings of the National academy of Sciences of the United States of America* 102 (46):16569.
Haak, Laurel L, Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner. 2012. "ORCID: A system to uniquely identify researchers." *Learned Publishing* 25 (4):259–264.
ICMJE. 2017. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. www.icmje.org/recommendations/.
Ioannidis, J.P.A., R. Klavans, and K.W. Boyack. 2018. "The scientists who publish a paper every five days." *Nature* 561: 167–169
Labbé, Cyril. 2010. "Ike antcare one of the great stars in the scientific firmament". RR-LIG-008. https://hal.archives-ouvertes.fr/hal-01354123.
Larivieré, V., C. Ni, Y. Gingras, B. Cronin, and C.R. Sugimoto. 2013. "Global gender disparities in science." *Nature* 504: 211–213
Lazarsfeld, Paul F. 1958. "Evidence and inference in social research." *Daedalus* 87 (4):99–130.
Martinson, Brian C., M.S. Anderson, and R. de Vries. 2005. "Scientists behaving badly." *Nature* 435: 737–738.
McNutt, M. 2014. "Editorial: The measure of research merit." *Science* 346: 1155.
Meho, Lokman I, and Kiduk Yang. 2007. "Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar." *Journal of the Association for Information Science and Technology* 58 (13):2105–2125.

Ochsner, M, S. Hug, and I. Galleron. 2016. "The future of research assessment in the humanities: Bottom-up assessment procedures." *Palgrave Communications* 3:17030. DOI: 10.1057/palcomms.2017.20

Price, D.J.de S. 1963. *Little Science, Big Science*. New York: Columbia University Press.

Price, Derek John de Solla. 1961. *Science since Babylon*. Vol. 47. New Haven & London: Yale University Press.

Stephan, P., R. Veugelers, and J. Wang. 2017. "Blinkered by bibliometrics." *Nature* 544: 413–414.

Wade, N. 1975. 'Citation analysis: A new tool for science administrators'. *Science*, 188: 429–432.

Wennerås, C., and A. Wold. 1997. "Nepotism and sexism in peer-review." *Nature* **387**: 341–343.

Williams, G., A. Basso, I. Galleron, and T. Lippiello. 2018. "More, Less or Better: The Problem of Evaluating Books in SSH Research." In: *The Evaluation of Research in Social Science*, edited by A. Bonaccorsi. Berlin: Springer.

Wouters, P. 1999. *The Citation Culture*. Amsterdam: Universiteit van Amsterdam.

Zhang, Chun-Ting. 2009. "The e-index, complementing the h-index for excess citations." *PLoS One* 4 (5):e5429.

PART III

Does Bias Affect Our Access to Scientific Knowledge?

11

WHY DO LOGICALLY INCOMPATIBLE BELIEFS SEEM PSYCHOLOGICALLY COMPATIBLE?

Science, Pseudoscience, Religion, and Superstition

Andrew Shtulman and Andrew Young

Introduction

Humans' understanding of science is at once impressive and appalling. Humans, as a species, have uncovered the hidden causes of most natural phenomena, from rainbows to influenza to earthquakes. Unobservable causal agents, like germs and genes, have been discovered and studied and are now familiar to everyone, scientists and nonscientists alike. Even children are familiar with germs and genes, despite our ignorance of these entities for the majority of human history. On the other hand, individual humans often lack an understanding of core scientific ideas – ideas that most educated adults have encountered in books, museums, and classes but still fail to understand. National polls in the United States and other countries have revealed that millions of people believe that dinosaurs coexisted with humans, that atoms are smaller than electrons, and that the earth's continents are fixed in place. Likewise, millions are skeptical that genetically modified foods are safe to eat, that climate change is caused by humans, and that humans evolved from non-human ancestors (National Science Board, 2018; Pew Research Center, 2015).

Exposure to scientific ideas does not guarantee their comprehension or acceptance. While there are several reasons why scientific ideas remain elusive, one primary reason is that they conflict with the explanations we devise on our own about how the world works (Carey, 2009; Shtulman, 2017; Vosniadou, 1994). These explanations, termed "folk theories" or "intuitive theories," are typically constructed in childhood prior to any formal instruction in the relevant domain. They are derived from a combination of inputs – innate concepts, empirical observations, culturally transmitted beliefs – and they serve the same function as scientific theories, namely, furnishing us with systematic and coherent inferences about natural phenomena (though see DiSessa, 2008, for an alternative view of how conceptual knowledge is structured).

Intuitive theories allow us to interpret and intervene on the phenomena they cover, but they also act as an impediment to learning more accurate theories of those phenomena. In the domain of evolution, for instance, children form creationist theories of the origin of species that impede learning about common descent (Blancke, De Smedt, De Cruz, Boudry, & Braeckman, 2012), and they construct essentialist theories of biological adaptation that impede learning about natural selection (Shtulman & Calabi, 2012). Intuitive theories impede the learning of scientific theories because they carve the world into entities and processes that do not actually exist – entities and processes that better align with how we perceive reality than with reality itself (Thagard, 2014). Learning a scientific theory thus requires learning a new ontology, or abstract causal framework.

Learning a new ontology can be quite difficult (Slotta & Chi, 2006), but it is not the only difficulty posed by intuitive theories. Another difficulty is avoiding the influence of intuitive theories even after one has learned the new ontology. Several lines of research indicate that intuitive theories are never fully replaced by scientific theories. Rather, the two theories coexist in the mind of the learner, providing competing interpretations of the same phenomena (Barlev, Mermelstein, & German, 2017; Foisy, Potvin, Riopel, & Masson, 2015; Goldberg & Thompson-Schill, 2009; Kelemen, Rottman, & Seston, 2013; Merz, Dietsch, & Schneider, 2016; Shtulman & Valcarcel, 2012).

Consider illness. The scientific explanation for illness is germs – microscopic organisms that invade a body and hijack its resources to further their own replication and survival – but learning about germs does not displace other, more intuitive ways of thinking about illness. We also explain illness as the result of behaviors that are not actually associated with germ transmission or germ reproduction, such as going out into the cold without a jacket or going to sleep with wet hair (Au et al., 2008). We may evoke supernatural causes as well, pointing to karma if we are Indian (Raman & Gelman, 2004), witchcraft if we are African (Legare & Gelman, 2008), or God if we are Judeo-Christian (Laurin & Kay, 2017).

This chapter discusses several phenomena for which scientific explanations coexist with non-scientific ones. We explore a range of nonscientific explanations, including religious explanations (e.g., attributing illness to God), superstitious explanations (e.g., attributing illness to witchcraft), and pseudoscientific explanations (e.g., attributing illness to behaviors unrelated to germs). We argue that the ubiquity of coexisting explanations across cultures and domains implies that coexistence is an inherent feature of conceptual representations and a regular impediment to understanding science. We conclude by considering several questions about the origin and dynamics of coexistence that may shed further light on our understanding and acceptance of scientific explanations.

Coordinating Multiple Representations of the Natural World

Natural phenomena can be mentally represented in several ways. Sometimes these representations are compatible with one another, and sometimes they

are not. Representations at different levels of abstraction may be compatible, as when we represent the diffusion of a gas at both the macroscopic level (in terms of pressure and volume) and the microscopic level (in terms of molecular interactions; Chi, Roscoe, Slotta, Roy, & Chase, 2012). Likewise, representations that evoke different scales of causation may be compatible, as when we represent sexual behavior as both an evolved adaptation (for perpetuating one's genes) and an environmentally-triggered response (in the presence of potential mates; Tinbergen, 1963).

Representations that conflict are those that evoke mutually incompatible ontologies – ontologies that operate at the same level of abstraction and on the same scale of causation. Those who hold incompatible ontologies are sometimes aware of the conflict, but in many cases that conflict is implicit, revealed only when we are asked to reason about the ontologically relevant phenomena under time pressure or cognitive load. The fact that we are often unaware of holding mutually incompatible ontologies underscores the pervasiveness of this phenomenon and raises questions about the psychological status of scientific explanations, which are almost always learned after a religious, superstitious, or pseudoscientific explanation. Such explanations may vary in their surface-level features, but they share the deeper commonality of arising from an intuitive theory that is ontologically distinct from scientists' current theory of the domain.

Coexistence of Science and Pseudoscience

Explanations for natural phenomena that do not conform to science but also do not evoke supernatural causes are termed here "pseudoscientific." These explanations are often endorsed by children, who construct them prior to formal schooling, and they were once endorsed even by scientists, prior to the discoveries that displaced them (Shtulman, 2017). Consider intuitive models of the solar system. Everyday observation of the sun, moon, and earth suggests that the sun and moon are in motion but the earth is not. These observations motivate a geocentric model of the solar system, in which day and night are caused by the sun and moon orbiting the earth in alternation. Most children hold this model, as did most adults centuries ago (Vosniadou & Brewer, 1994).

Today, most adults know that the sun is at the center of the solar system, not the earth, and that day and night are caused by the earth's motion, not the sun's or the moon's. Under time pressure, however, adults reveal evidence of harboring geocentric models. In recent studies by Shtulman and colleagues (Shtulman & Harrington, 2016; Shtulman & Valcarcel, 2012), college-educated adults were asked to verify two types of scientific statements: those that accord with intuition and those that conflict with it. The statements covered ten domains of knowledge, including astronomy. In the domain of astronomy, participants verified statements about planets, stars, lunar phases, the seasons, and the solar system. Participants' verifications for intuitive statements, like "the moon revolves around the earth," were compared to their verifications for closely-matched counterintuitive

statements, like "the earth revolves around the sun." Overall, participants were less accurate at verifying counterintuitive statements relative to intuitive ones, and when they verified counterintuitive statements correctly, they took longer than when verifying intuitive statements of the same form.

Similar results have been documented in the domain of biology, with respect to adults' conceptions of life. Biologists identify life with the capacity to engage in metabolic processing, but young children identify life with self-directed motion (Piaget, 1929). That is, young children construct intuitive theories of life that correctly classify animals as alive (because animals move on their own) but incorrectly classify plants as not alive (because plants do not move on their own, at least not to the naked eye). By age ten, most children have learned to associate life with metabolic activities rather than motion (Stavy & Wax, 1989), but this knowledge does not erase the previous misconception that only moving things are alive. Under time pressure, adolescents and adults often misclassify plants as not alive. They also misclassify nonliving objects that move on their own, like the sun and the clouds, as alive (Babai, Sekal, & Stavy, 2010; Goldberg & Thompson-Schill, 2009; Young et al., 2018).

An even more striking demonstration of the resilience of motion-based, or "animistic," theories of life comes from studies of how Alzheimer's Disease affects biological reasoning (Zaitchik & Solomon, 2008). When individuals with Alzheimer's Disease are asked to name some things that are alive, they frequently mention animals but rarely mention plants. When asked about the life status of natural phenomena, like fire and wind, they typically judge them to be alive, even when they are given no time limit for responding. And when asked for a definition of life, they cite the capacity for motion more often than metabolic activities, like breathing or growing. These impairments are not just the result of age; elderly adults who are not afflicted by Alzheimer's Disease cite plants as examples of living things, judge natural phenomena as not alive, and define life in metabolic terms. The cognitive impairments wrought by Alzheimer's Disease strip away scientific knowledge of life, revealing an intuitive theory of life constructed decades earlier, when these elderly adults were children.

Coexistence of Science and Religion

A dominant source of non-scientific explanations is religion. Religious explanations for natural phenomena typically evoke supernatural agents (like gods, spirits, and ancestors), which, in turn, evoke our intuitions about agents in general – our theory of mind (Heiphetz, Lane, Waytz, & Young, 2016). Consider the difference between scientific and religious explanations for why organisms are adapted to their environment. The scientific explanation – evolution by natural selection – views adaptation as the selective propagation of randomly-occurring mutations across many generations of an interbreeding population, whereas the most popular religious explanation – creationism – views adaptation as the product of a divine

creator. Evolutionary explanations for adaptation require coordinating several unfamiliar processes: mutation, heredity, differential survival, and differential reproduction. Creationist explanations, on the other hand, typically tap into a single, well-understood process: intentional design.

Because creationist explanations are intuitively compelling, they are difficult to dispel. Interventions that have proven successful at teaching evolutionary principles rarely uproot inclinations toward creationism. For instance, museum exhibits that succeed at increasing visitors' scientific understanding of micro-evolutionary change have no effect on their endorsement of creationist explanations for those changes (Spiegel et al., 2012). Likewise, storybooks that succeed at teaching elementary schoolers selection-based explanations for the origin of biological traits have no effect on their endorsement of creationist explanations for those traits (Shtulman, Neal, & Lindquist, 2016). If people are allowed to endorse both evolutionary and creationist accounts of biological change, they do.

In this same vein, people who endorse evolutionary explanations for life can be induced to doubt those explanations in anxiety-provoking situations, such as when contemplating their own mortality. In a study by Tracy, Hart, and Martens (2011), participants read and evaluated two passages: an argument in favor of an evolutionary explanation for life, written by biologist Richard Dawkins, and an argument in favor of a creationist explanation, written by the intelligent design proponent Michael Behe. Half of the participants were primed to think about their mortality prior to reading the passages, and half were primed to think about an unpleasant experience other than death. The mortality prime decreased participants' ratings of the quality and truthfulness of the evolutionary passage and increased their ratings of the quality and truthfulness of the creationist passage, relative to the non-mortality prime. These changes held regardless of how educated the participants were, how religious they were, and how strongly they accepted evolution prior to the study.

Similar results have been obtained in comparing people's endorsement of religious and scientific explanations for the origin of the universe: God vs. the Big Bang. In a study by Preston and Epley (2009), participants read a passage about the Big Bang that either affirmed or challenged the theory's validity. They then completed a speeded categorization task in which the concepts *God* and *science* were implicitly primed. In this task, participants categorized adjectives like "excellent" and "awful" as positive or negative as quickly as possible. On some trials, the adjectives were preceded by the word "science" for 15 milliseconds or the word "God" for 15 milliseconds – too quickly for participants to consciously register.

Participants who read the passage that affirmed the validity of the Big Bang were faster to respond to positive adjectives than negative adjectives when those adjectives were preceded by the word "science," whereas participants who read the passage that challenged the validity of the Big Bang were faster to respond to positive adjectives than negative adjectives when preceded by the word "God." In

other words, priming participants to think of the Big Bang as valid rendered their implicit associations with science more positive and their implicit associations with God less positive, whereas priming participants to think of the Big Bang as invalid had the opposite effect. These findings imply that people have access to both religious and scientific explanations and can be induced to shift their evaluations of those explanations by subtle contextual cues. These findings also imply that people view religious and scientific explanations as conflicting, because priming participants to value one explanation led them to devalue the other.

Coexistence of Science and Superstition

The two types of non-scientific explanations discussed thus far – pseudoscientific explanations and religious explanations – differ in their form of causation (natural vs. supernatural), as well as their relation to cultural institutions. Religious explanations are embedded in a coherent, institutionally-endorsed narrative about the origins of the world and humans' place within it, whereas pseudoscientific explanations are typically constructed ad hoc and are not part of the doctrines or teachings of any institution. Superstitious explanations fall between these two extremes. They evoke supernatural causes, like religious explanations, but they are constructed and transmitted through informal channels, like pseudoscientific explanations.

Illness is a domain in which superstitious explanations proliferate, possibly because of the anxiety aroused by existential threats to oneself and one's loved ones. The particular superstitions vary by culture. South Africans appeal to curses cast by jealous neighbors and displeased ancestors, at least for serious illnesses like AIDS (Legare & Gelman, 2008). South Asians appeal to imminent justice, or the conviction that bad things happen to bad people (Raman & Gelman, 2004). Vietnamese individuals appeal to evil spirits and magic spells, fixating on omens of misfortune such as broken mirrors, haunted houses, or graveyards (Nguyen & Rosengren, 2004). Critically, appeals to superstition do not occur in isolation; they occur alongside appeals to biological factors, such as contact with a disease-infected person or disease-infected object. Individuals who appeal to superstition also typically know a fair amount about the transmission, symptoms, and treatment of the target disease (Legare & Gelman, 2008). Superstition is embraced in spite of, not in place of, biological knowledge.

Teleology is another form of cognition that can take on supernatural overtones. Teleology is explaining something in terms of its end, purpose, or goal (Lennox & Kampourakis, 2013), as when we appeal to sight as the explanation for eyes or flight as the explanation for wings. Kelemen (1999) has shown that children are more "promiscuous" with their teleological explanations than adults are. Whereas both children and adults provide teleological explanations for human artifacts (e.g., pencils are "for writing") and biological parts (e.g., ears are "for hearing"), only children provide teleological explanations for whole organisms (e.g., birds

are "for flying") and naturally occurring objects (e.g., clouds are "for raining"). Children become more selective in their use of teleology by early adolescence, but that selectivity is tenuous.

When college-educated adults are asked to judge the acceptability of teleological explanations under speeded conditions, they tend to accept explanations they would normally reject, such as "birds are for flying" and "clouds are for raining" (Kelemen et al., 2013). Moreover, just as Alzheimer's patients willingly endorse animistic conceptions of life, they also willingly endorse teleological conceptions of nature. Alzheimer's patients claim that teleological explanations for natural phenomena, like "rain exists so that plants and animals have water for drinking," are not only acceptable but are actually preferable to mechanistic explanations, like "rain exists because water condenses in clouds and forms droplets" (Lombrozo, Kelemen, & Zaitchik, 2007).

Teleology is also regularly evoked to explain the events in one's own life. Most college-educated adults eschew the possibility that life events transpire at random and believe instead that "everything happens for a reason" (Banerjee & Bloom, 2014; Norenzayan & Lee, 2010; Svedholm, Lindeman, & Lipsanen, 2010). Emotionally significant events (e.g., meeting a future spouse) and statistically unlikely events (e.g., holding a royal flush in poker) are attributed to fate and assigned meaning, even by adults who are not religious and do not believe in God. These adults deny that supernatural agents are responsible for life events, yet they cannot shake the idea that such events portend larger patterns of meaning.

Coexistence Is a Cognitive Default

The studies reviewed earlier indicate that scientific explanations coexist with non-scientific ones in a variety of domains, from astronomy to evolution to illness. Coexistence has been observed in other domains as well, including motion (Foisy, et al., 2015), matter (Potvin, Masson, Lafortune, & Cyr, 2015), electricity (Masson, Potvin, Riopel, & Foisy, 2014), cosmography (Carbon, 2010), and neuroscience (Preston, Ritter, & Hepler, 2013). In any domain where intuitive theories precede scientific theories, the former appears to survive the latter. This finding has been observed using behavioral methods, such as those described above, as well as neurocognitive methods. Using function magnetic resonance imaging (fMRI), researchers have observed that physics experts' ability to judge intuitively plausible events, like a heavy object falling to the ground faster than a lighter object, as physically impossible requires heightened activity in the anterior cingulate cortex and prefrontal cortex (Foisy et al., 2015; Masson et al., 2014). These areas of the brain are involved in inhibition, and their activation suggests that physics experts must inhibit latent misconceptions in order to respond in accordance with known physical principles.

The coexistence of scientific and non-scientific explanations appears to be pervasive across cultures as well. This phenomenon has been observed most

extensively in American and European samples but has also been observed in samples from China (Rottman, Zhu, Wang, Seston Schillaci, Clark, & Kelemen, 2017), India (Raman & Gelman, 2004), Vietnam (Nguyen & Rosengren, 2004), Mexico (Rosengren, Miller, Gutiérrez, Chow, Schein, & Anderson, 2014), South Africa (Legare & Gelman, 2008), Madagascar (Astuti & Harris, 2008), and Vanuatu (Watson-Jones, Busch, Harris, & Legare, 2017). People in different cultures construct different intuitive theories, but the resilience of intuitive theories in the face of scientific theories appears to be universal.

Conflict between intuitive and scientific theories has been observed across the lifespan as well, in children (Legare, Evans, Rosengren, & Harris, 2012), adolescents (Babai et al., 2010), young adults (Shtulman & Valcarcel, 2012), and elderly adults (Barlev, Mermelstein, & German, 2018). It has even been observed in populations with extensive scientific knowledge, including high school science teachers (Potvin & Cyr, 2017) and college science professors (Shtulman & Harrington, 2016). Under time pressure, biology professors are prone to judge plants as not alive (Goldberg & Thompson-Schill, 2009), and physics professors are prone to endorse teleological explanations for natural phenomena (Kelemen et al., 2013). This finding – that even professional scientists harbor non-scientific explanations – suggests that coexistence is an inevitable byproduct of acquiring more than one representation of the same domain. Professional scientists may deploy scientific theories on a daily basis, but that practice does not appear to erase, or even weaken, intuitive theories of the same phenomena.

Questions About the Origin and Nature of Explanatory Coexistence

The phenomenon of coexisting explanations has been well documented, but its causes and consequences are not well understood. Here we consider questions about the origin of coexistence and its effects on everyday reasoning, with the goal of identifying directions for future research.

Does Coexistence Require Belief?

In some cases of coexistence, individuals explicitly endorse incompatible explanations, whereas in others, individuals show evidence of mentally representing incompatible explanations but endorse only one. When South Africans point to both witchcraft and unprotected sex as reasons for contracting AIDS or when museum visitors endorse both creationism and evolution as explanations for the origin of species, they are exhibiting an explicit form of coexistence reasoning. On the other hand, when biologists take longer to classify plants as alive than to classify animals as alive or when physicists endorse teleological explanations for natural phenomena under time pressure, they presumably do not endorse the non-scientific ideas their behavior has betrayed. Biologists and physicists may represent

nonscientific ideas at an implicit level, but they have the knowledge and knowhow to reject those ideas at an explicit level.

That said, biologists and physicists were once children who lacked scientific knowledge and believed nonscientific ideas, accepting those ideas as true descriptions of reality. Does a nonscientific idea have to be believed, at some point in development, to survive the acquisition of a scientific alternative? Or can an idea that was entertained but never accepted as true still cause cognitive conflict in the relevant domain?

Research on the coexistence of scientific and supernatural explanations suggests that coexistence can, in fact, occur in the absence of belief. Atheists, after all, show signs of representing the supernatural ideas that all life events are meaningful (Banerjee & Bloom, 2014), that people continue to think and feel after they have died (Bering, 2002), and that animals, plants, and other natural kinds were purposely created by some kind of being (Järnefelt, Canfield, & Kelemen, 2015). These findings suggest that the nonscientific ideas prevalent in one's culture may be mentally represented as viable alternatives to science, even if those ideas are not personally endorsed.

However, it's not clear that the atheists in these studies have always been atheists. Additional research is needed to verify that explanations one has never endorsed can indeed compete with the explanations one currently endorses. Such research could introduce participants to novel nonscientific explanations (e.g., magnet therapy for treating chronic pain) and then manipulate the believability of those explanations, though such a manipulation would likely require sustained reinforcement of the target explanation to prove effective. Another possibility would be to explore the onset of coexistence in cultures that differ in their baseline levels of acceptance for some nonscientific explanation (e.g., creationism, as endorsed in Scandinavia vs. the Middle East), with the goal of disentangling the roles of personal acceptance and cultural acceptance on the cognitive conflict induced by coexisting explanations.

Does Coexistence Require Comprehension?

In research by Shtulman and colleagues (Shtulman & Harrington, 2016; Shtulman & Valcarcel, 2012), the presence of coexistence was explored in ten domains: astronomy, evolution, fractions, genetics, germs, matter, mechanics, physiology, thermodynamics, and waves. Using a speeded sentence-verification task (described above), Shtulman and colleagues documented cognitive conflict between intuitive and scientific theories in all ten domains. In some domains, the relevant scientific concepts are acquired early in life, such as physiology (Hatano & Inagaki, 1994) and matter (Smith, 2007), whereas in others the relevant concepts are acquired late in life, such as evolution (Shtulman & Calabi, 2013) and mechanics (Halloun & Hestenes, 1985). The discovery of coexistence in the latter domains was unexpected, given that most college-educated adults exhibit only partial understanding

of these domains on unspeeded, comprehensive assessments. Nevertheless, partial understanding appears to be sufficient for creating cognitive conflict between scientific and intuitive theories.

Consistent with this finding, research in science education has found that coexistence emerges early in instruction. Studies that have explored the efficacy of various teaching interventions have found that improvements in scientific reasoning rarely force a decrease in intuitive reasoning (Coley, Arenson, Xu, & Tanner, 2017; Nehm & Reilly, 2007; Schneider & Hardy, 2013; Shtulman et al., 2016; Spiegel et al., 2012). Successful instruction appears to increase the number of reasoning strategies rather than the accuracy of a single strategy, and this result can occur after a single lesson (see Siegler, 1998, for parallel findings in the development of procedural knowledge). A single lesson may not be enough time for students to fully comprehend a new scientific explanation, but it may be enough for students to appreciate the utility of that explanation.

Utility has been cited as the prime reason intuitive theories persist in the face of scientific ones (Ohlsson, 2009; Shtulman & Lombrozo, 2016), and utility may also be the reason that scientific theories begin to conflict with intuitive theories before they are fully understood. Further research is needed to explore the conditions under which a scientific explanation is transformed from a hypothetical idea to a viable alternative. The utility of a scientific explanation may have to cross some threshold before it begins to conflict with a more intuitive explanation. Alternatively, the utility of an intuitive explanation may have to drop below some threshold before a scientific explanation can begin to compete with it.

Are Coexisting Explanations Activated Serially or In Parallel?

To date, the most common measure of coexistence is a decrease in the speed or accuracy of scientific reasoning when that reasoning conflicts with intuitive reasoning (Babai et al., 2010; Barlev et al., 2017, 2018; Foisy et al., 2015; Goldberg & Thompson-Schill, 2009; Kelemen et al., 2013; Merz et al., 2016; Potvin et al., 2015; Potvin & Cyr, 2017; Rottman et al., 2017; Shtulman & Harrington, 2016; Shtulman & Valcarcel, 2012). Findings of this nature imply that intuitive responses have to be inhibited in order for scientific ones to be articulated, but the dynamics of this process are not yet understood. Scientific responses may be activated in parallel with intuitive ones, or they may be activated only after the intuitive ones have been inhibited.

One reason to favor a parallel-activation account is that the conflict between science and intuition is seemingly impervious to expertise (Goldberg & Thompson-Schill, 2009; Kelemen et al., 2013; Shtulman & Harrington, 2016). If experts routinely deploy scientific concepts, then those concepts should become closely associated with science-relevant contexts and should not have to await activation following the inhibition of erroneous ideas. Indeed, interventions that increase people's accuracy at verifying counterintuitive scientific ideas have no

effect on the speed of those verifications (Young et al., 2018), implying that the activation of erroneous ideas is inevitable.

On the other hand, research on the processing dynamics of a similar task – the Cognitive Reflection Test (CRT; Frederick, 2005) – reveals that intuitive reasoning has to be inhibited before analytic reasoning can be engaged. The CRT measures a person's tendency to reflect on the validity of intuitive, yet inaccurate, responses and override those responses in favor of more accurate ones. Consider this item: "A bat and a ball cost $1.10 in total. The bat costs $1 more than ball. How much does the ball cost?" Many adults provide the intuitive response of 10 cents, defaulting to simple subtraction, yet the correct answer is 5 cents (because the bat must cost $1.05 if their sum is $1.10 and their difference is $1.00). When the response options "5 cents" and "10 cents" are displayed on opposite sides of a computer screen and must be selected using a mouse, respondents' mouse trajectories reveal an initial pull toward the intuitive option even when the correct option is ultimately selected (Travers, Rolison, & Feeney, 2016). This result implies that the intuitive response is activated first, and the correct response is activated second, following inhibition of the intuitive response. Measures of online processing, such as mouse tracking or eye tracking, could clarify whether the conflict between intuitive theories and scientific theories follows the same pattern as the conflict between intuitive responses and analytic responses on the the CRT.

What Is the Role of Executive Function in Prioritizing Science?

When cognitive conflict arises, we typically resolve that conflict though executive function (Koechlin & Summerfield, 2007). Executive function refers to a suite of domain-general abilities – working memory, inhibitory control, comprehension monitoring, set shifting – and its operation has been linked to science learning. Children with higher executive function construct biological theories of life, death, and the body earlier than those with lower executive function (Zaitchik, Iqbal, & Carey, 2014), and they also learn more when directly instructed on these topics (Bascandziev, Tardiff, Zaitchik, & Carey, 2018). Conversely, the loss of executive function has been linked to the loss of scientific knowledge and the reemergence of childlike misconceptions, such as the misconception that the sun and the wind are alive but plants are not (Tardiff, Bascandziev, Sandor, Carey, & Zaitchik, 2017).

Executive function may also be linked to the prioritization of scientific theories over intuitive theories when those theories compete to provide inferences about the same phenomena. Inhibitory control is an aspect of executive function, and brain networks implicated in inhibitory control are activated when science experts access counterintuitive scientific ideas, as noted earlier (Foisy et al., 2015; Masson et al., 2014). Behavioral measures of inhibitory control have not, however, revealed consistent associations between inhibition and the ability to prioritize scientific responses over intuitive ones. At least three studies (Barlev et al.,

2017; Barlev et al., 2018; Kelemen et al., 2013) have failed to observe correlations between speeded scientific-reasoning tasks and the Stroop task – a measure of cognitive control in which participants must name the color of ink used to print words denoting a different color (e.g., "red" printed in blue ink) – though another study (Vosniadou et al., 2018) did document such correlations.

This pattern of results suggests that inhibitory control may be less important than other aspects of executive function for prioritizing scientific responses over nonscientific ones. Alternatively, inhibition may be important, but the Stroop task measures the wrong type of inhibition. The Stroop task measures inhibition of perceptual information, whereas a task that measures the inhibition of conceptual information, such as the CRT, may be more appropriate. Individuals with high CRT scores do perform better on tests of science understanding than those with low CRT scores (Shtulman & McCallum, 2014; Young & Shtulman, 2018), but it's unclear whether cognitive reflection is needed to prioritize scientific ideas over intuitive ideas or merely to learn scientific ideas in the first place.

Conclusions

Psychologists have long observed the prevalence and popularity of pseudoscientific, religious, and superstitious explanations, but those explanations were presumed to occupy the minds of people ignorant of science or actively opposed to science. That presumption has now been overturned. Findings from cognitive psychology, developmental psychology, cognitive neuroscience, and science education indicate that scientific explanations coexist with nonscientific ones in the same minds – even the minds of the most scientifically literate adults. The coexistence of scientific and nonscientific explanations appears to be an inherent feature of how humans represent and reason about the natural world. Studying this phenomenon promises to refine our theories of conceptual representation, as well as improve the teaching and learning of counterintuitive scientific ideas.

Acknowledgements

This research was supported by a James S. McDonnell Understanding Human Cognition Scholar Award to Andrew Shtulman.

References

Astuti, R., & Harris, P. L. (2008). Understanding mortality and the life of the ancestors in rural Madagascar. *Cognitive Science, 32*, 713–740.

Au, T. K. F., Chan, C. K., Chan, T. K., Cheung, M. W., Ho, J. Y., & Ip, G. W. (2008). Folkbiology meets microbiology: A study of conceptual and behavioral change. *Cognitive Psychology, 57*, 1–19.

Babai, R., Sekal, R., & Stavy, R. (2010). Persistence of the intuitive conception of living things in adolescence. *Journal of Science Education and Technology*, *19*, 20–26.

Banerjee, K., & Bloom, P. (2014). Why did this happen to me? Religious believers' and non-believers' teleological reasoning about life events. *Cognition, 133*, 277–303.

Barlev, M., Mermelstein, S., & German, T. C. (2017). Core intuitions about persons coexist and interfere with acquired Christian beliefs about God. *Cognitive Science*, 41, 425–454.

Barlev, M., Mermelstein, S., & German, T. C. (2018). Representational coexistence in the God concept: Core knowledge intuitions of God as a person are not revised by Christian theology despite lifelong experience. *Psychonomic Bulletin and Review*, 25, 2330–2338.

Bascandziev, I., Tardiff, N., Zaitchik, D., & Carey, S. (2018). The role of domain-general cognitive resources in children's construction of a vitalist theory of biology. *Cognitive Psychology, 104*, 1–28.

Bering, J. M. (2002). Intuitive conceptions of dead agents' minds: The natural foundations of afterlife beliefs as phenomenological boundary. *Journal of Cognition and Culture, 2*, 263–308.

Blancke, S., De Smedt, J., De Cruz, H., Boudry, M., & Braeckman, J. (2012). The implications of the cognitive sciences for the relation between religion and science education: The case of evolutionary theory. *Science and Education, 21*, 1167–1184.

Carbon, C. C. (2010). The Earth is flat when personally significant experiences with the sphericity of the Earth are absent. *Cognition, 116*, 130–135.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Chi, M. T. H., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive Science, 36*, 1–61.

Coley, J. D., Arenson, M., Xu, Y., & Tanner, K. D. (2017). Intuitive biological thought: Developmental changes and effects of biology education in late adolescence. *Cognitive Psychology, 92*, 1–21.

DiSessa, A. A. (2008). A bird's-eye view of the "pieces" vs. "coherence" controversy (from the "pieces" side of the fence). In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 35–60). New York: Routledge.

Foisy, L. M. B., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education, 4*, 26–36.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25–42.

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science, 20*, 480–487.

Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics, 53*, 1043–1055.

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition, 50*, 171–188.

Heiphetz, L., Lane, J. D., Waytz, A., & Young, L. L. (2016). How children and adults represent God's mind. *Cognitive Science, 40*, 121–144.

Järnefelt, E., Canfield, C. F., & Kelemen, D. (2015). The divided mind of a disbeliever: Intuitive beliefs about nature as purposefully created among different groups of non-religious adults. *Cognition, 140*, 72–88.

Kelemen, D. (1999). The scope of teleological thinking in preschool children. *Cognition, 70*, 241–272.

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*, 1074–1083.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences, 11*, 229–235.

Laurin, K., & Kay, A. C. (2017). The motivational underpinnings of belief in God. *Advances in Experimental Social Psychology, 56*, 201–257.

Legare, C. H., & Gelman, S. A. (2008). Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science, 32*, 607–642.

Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development, 83*, 779–793.

Lennox, J. G., & Kampourakis, K. (2013) Biological teleology: The need for history. In K. Kampourakis (Ed.), *The philosophy of biology: A companion for educators* (pp. 421–454). Dordrecht: Springer.

Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations for patients with Alzheimer's disease. *Psychological Science, 18*, 999–1006.

Masson, S., Potvin, P., Riopel, M., & Foisy, L. M. B. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, *8*, 44–55.

Merz, C. J., Dietsch, F., & Schneider, M. (2016). The impact of psychosocial stress on conceptual knowledge retrieval. *Neurobiology of Learning and Memory, 134*, 392–399.

National Science Board (2018). *Science and engineering indicators*. Arlington, VA: National Science Foundation.

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57*, 263–272.

Nguyen, S., & Rosengren, K. (2004). Causal reasoning about illness: A comparison between European- and Vietnamese-American children. *Journal of Cognition and Culture, 4*, 51–78.

Norenzayan, A., & Lee, A. (2010). It was meant to happen: Explaining cultural variations in fate attributions. *Journal of Personality and Social Psychology, 98*, 702–720.

Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*, 20–40.

Pew Research Center (2015). *Public and scientists' views on science and society*. Washington, DC: Pew Research Center.

Piaget, J. (1929/2007). *The child's conception of the world*. New York: Routledge.

Potvin, P., & Cyr, G. (2017). Toward a durable prevalence of scientific conceptions: Tracking the effects of two interfering misconceptions about buoyancy from preschoolers to science teachers. *Journal of Research in Science Teaching, 54*, 1121–1142.

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education, 13*, 21–43.

Preston, J., & Epley, N. (2009). Science and god: An automatic opposition between ultimate explanations. *Journal of Experimental Social Psychology, 45*, 238–241.

Preston, J. L., Ritter, R. S., & Hepler, J. (2013). Neuroscience and the soul: Competing explanations for the human experience. *Cognition, 127*, 31–37.

Raman, L., & Gelman, S. A. (2004). A cross-cultural developmental analysis of children's and adults' understanding of illness in South Asia (India) and the United States. *Journal of Cognition and Culture, 2*, 293–317.

Rosengren, K. S., Miller, P. J., Gutiérrez, I. T., Chow, P., Schein, S., & Anderson, K. A. (2014). Children's understanding of death: Toward a contextual perspective. *Monographs of the Society for Research in Child Development, 79*, 1–162.

Rottman, J., Zhu, L., Wang, W., Seston Schillaci, R., Clark, K. J., & Kelemen, D. (2017). Cultural influences on the teleological stance: Evidence from China. *Religion, Brain and Behavior, 7*, 17–26.

Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology, 49*, 1639–1649.

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York: Basic Books.

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution* (pp. 47–65). Cambridge, UK: Oxford University Press.

Shtulman, A., & Calabi, P. (2013). Tuition vs. intuition: Effects of instruction on naïve theories of evolution. *Merrill-Palmer Quarterly*, *59*, 141–167.

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science, 8*, 118–137.

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner & A. Baron (Eds.), *Core knowledge and conceptual change* (pp. 49–67). Oxford: Oxford University Press.

Shtulman, A., & McCallum, K. (2014). Cognitive reflection predicts science understanding. Proceedings of the 36th Annual Conference of the Cognitive Science Society, 2937–2942.

Shtulman, A., Neal, C., & Lindquist, G. (2016). Children's ability to learn evolutionary explanations for biological adaptation. *Early Education and Development, 27*, 1222–1236.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition, 124,* 209–215.

Siegler, R. S. (1998). *Emerging minds: The process of change in children's thinking*. Oxford: Oxford University Press.

Slotta, J. D., & Chi, M. T. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction, 24*, 261–289.

Smith, C. L. (2007). Bootstrapping processes in the development of students' commonsense matter theories: Using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction, 25*, 337–398.

Spiegel, A. N., Evans, E. M., Frazier, B., Hazel, A., Tare, M., Gram, W., & Diamond, J. (2012). Changing museum visitors' conceptions of evolution. *Evolution: Education and Outreach, 5*, 43–61.

Stavy, R., & Wax, N. (1989). Children's conceptions of plants as living things. *Human Development, 32*, 88–94.

Svedholm, A. M., Lindeman, M., & Lipsanen, J. (2010). Believing in the purpose of events – why does it occur, and is it supernatural? *Applied Cognitive Psychology, 24*, 252–265.

Tardiff, N., Bascandziev, I., Sandor, K., Carey, S., & Zaitchik, D. (2017). Some consequences of normal aging for generating conceptual explanations: A case study of vitalist biology. *Cognitive Psychology, 95*, 145–163.

Thagard, P. (2014). Explanatory identities and conceptual change. *Science and Education, 23*, 1531–1548.

Tinbergen, N. (1963). On aims and methods of ethology. *Ethology, 20*, 410–433.

Tracy, J. L., Hart, J., & Martens, J. P. (2011). Death and science: The existential underpinnings of belief in intelligent design and discomfort with evolution. *PloS ONE, 6*, e17349.

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the cognitive reflection test. *Cognition, 150*, 109–118.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45–69.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science, 18*, 123–183.

Vosniadou, S., Pnevmatikos, D., Makris, N., Lepenioti, D., Eikospentaki, K., Chountala, A., & Kyrianakis, G. (in press). The recruitment of shifting and inhibition in online science and mathematics tasks. *Cognitive Science*.

Watson-Jones, R. E., Busch, J. T. A., Harris, P. L., & Legare, C. H. (2017). Does the body survive death? Cultural variation in beliefs about life everlasting. *Cognitive Science, 41*, 455–476.

Young, A., Laca, J., Dieffenbach, G., Hossain, E., Mann, D., & Shtulman, A. (2018). Can science beat out intuition? Increasing the accessibility of counterintuitive scientific ideas. *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, 1238–1243.

Young, A., & Shtulman, A. *Children's cognitive reflection predicts conceptual understanding in science and mathematics*. Manuscript under review.

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development, 85*, 160–175.

Zaitchik, D., & Solomon, G. E. A. (2008). Animist thinking in the elderly and in patients with Alzheimer's disease. *Cognitive Neuropsychology, 25*, 27–37.

12

DO OUR INTUITIONS MISLEAD US?

The Role of Human Bias in Scientific Inquiry

Susan A. Gelman and Kristan A. Marchak

Introduction

Science is carried out by scientists, and scientists are themselves human. History is littered with examples of human reasoning biases infiltrating the scientific enterprise. At times in human history, egocentric assumptions and self-interest led scientists to posit that Earth is the center of the universe, that some races are genetically inferior to others, or that female hysteria is a constitutional disorder. Scientists are infamous for not letting go of their own cherished theories even in the face of counter-evidence (Kuhn, 1996). Given this grim history, it is important to take a direct look at how human bias can arise not just out of egocentrism (as with earth-centric theories of astronomy) and self-interest (as with racist, sexist, or classist theories of human difference), but also as a consequence of heuristics in how information is processed and represented.

In this chapter we examine this question by focusing on a single reasoning bias, **psychological essentialism**, that shows a lack of fit with important, broadly agreed-upon scientific phenomena. A seemingly sensible starting assumption – that the world has real, discoverable structure – has problematic consequences for a surprising breadth of scientific fields, questions, and approaches. We sketch out, through a series of examples, the consequences of essentialist reasoning for the work of scientists, as well as for how non-scientists incorporate and understand scientific concepts and evidence. Specifically, we focus on two key scientific topics (biological change and genes). In each case, we review missteps, in science and/or current lay understandings, that can be traced back to essentialism. We conclude by asking whether there are ways that scientists, and those who communicate science to the lay public, can reduce these errors and distortions. Is it possible for the scientific enterprise to proceed apart from human bias – that is, are there methods and procedures that can insulate us from essentialist bias?

What Is Psychological Essentialism?

The term "essentialism" is used loosely and variously to cover a range of concepts that are importantly distinct from one another (Wilkins, 2013). Table 12.1 provides an overview of different senses of essentialism, varying in ontological type, specificity, and where the essence is located. In this context, we discuss essentialism that is causal, placeholder, and representational; this sense maintains consistency with how essentialism is discussed in the literature as well as our own prior work (Gelman, 2003).

At heart, essentialism is a realist assumption about categories. Just as individual objects (e.g., a squirrel, a diamond) are assumed to be natural entities that exist in the world, discontinuous with their surroundings and independent of our own thoughts or existence, so too are categories (e.g., squirrels, diamonds). The essentialist view is that humans do not construct these categories or arbitrarily determine their boundaries; rather, we discover them. This realist assumption about categories has two core components: a belief that certain categories are natural kinds, with indefinitely many rich and deep similarities shared among category members, and a belief that there is an internal essence shared by all members of a natural kind that *causes* members of that kind to be what they are. The belief in a causal essence is unspecified (i.e., the belief that there is an essence, without necessarily knowing what that essence is) but it is not blank; essentialism presupposes that the essence is internally located, inherent, biological, and causing a multitude of effects (Ahn et al., 2001).

Essentialism cannot be tested by asking people to report on the essence itself, given that it is a placeholder concept. Nonetheless, essentialism can be inferred from behaviors that reflect a set of interrelated component assumptions about categories, including: strict boundaries, within-category homogeneity, causal features, stability, and inductive potential (Gelman, 2003, 2004; Gelman, Heyman and Legare, 2007; Rhodes and Mandalaywala, 2017). These components are listed in Table 12.2. They can be tremendously useful in organizing category knowledge,

TABLE 12.1 Different senses of 'essentialism'

	Current sense	*Alternate senses*
Ontological type	Causal (essence has direct consequences for category features and structure)	Sortal (essence provides a definition) Ideal (essence has no real-world instantiation; also known as Platonic ideal)
Specificity	Placeholder (details are unknown, perhaps unknowable)	Specific (details are known)
Location	Representational (in human concepts, language, culture)	Metaphysical (in the world)

TABLE 12.2 Components of essentialism and associated biases

Component of essentialism	*Description*	*Sample bias*
Strict boundaries	Category boundaries are objective and absolute	Treating overlapping categories as discrete
Homogeneity	Variability within a category does not exist, or is only superficial	Underestimating variability within a category
Inherent causes	All members of a category share a single internal essence	Causes inhere in individuals, not larger units (structures, systems, populations)
Stability	Category identity cannot change; features are constant over change	Resistance to processes involving change
Inductive potential	Category members are alike in non-obvious ways	Treating superficial differences as deeply predictive

and in generating new inferences and discoveries. The assumption that turtles share deep similarities, that studying the physiology of one tapir will license broad inferences about the physiology of other tapirs, that superficial appearances can be misleading (e.g., pyrite and gold may look similar, but are deeply different) – all of these are useful ways of thinking about the world that accord with scientific discoveries. Nonetheless, there are situations in which our essentialist assumptions do not match the variability, change, and complexity of scientific topics. Some of these biases are briefly listed in Table 12.2.

Essentialism is core to how adults and children alike construe a range of categories in the natural world. Preschoolers treat certain categories as having an underlying nature that is internal, innate, and immutable (Gelman, 2003, 2004). This is seen in word learning, inductive inferences, explanations, and identity judgments (Gelman and Davidson, 2013). Children expect that members of a category will share internal, non-obvious, or causal similarities, even in the face of superficial dissimilarities. Category members are thought to have innate potential that resists environmental influences (Gelman and Wellman, 1991). Hearing that a pterodactyl is a "dinosaur", that a swaddled baby is a "boy", or that a child received the heart of a "monkey" leads to the inference that the pterodactyl does not live in a nest, that the baby will grow up to like football regardless of its upbringing, and that the donated heart will increase the child's tendency to eat bananas (Meyer et al., 2017; Taylor, Rhodes and Gelman, 2009). Essentialism is found across cultures (Deeb et al., 2011; Moya, Boyd and Henrich, 2015), and applies to biological categories, a subset of social categories, and attributes such as intelligence or mental illness. Essentialism does not imply that items truly possess essences, but rather that people treat them so. It does not require specialized knowledge, as the "essence"

may be an unspecified placeholder (e.g., a belief that boys and girls deeply differ in unknown respects).

Essentialism and Biological Change

One of the most fundamental yet difficult issues that scientists and laypeople alike have struggled with is how to understand and explain biological change. Change is constant and inevitable in the world of living things. Individual organisms continuously undergo internal and interactive processes that modify their every component, from epidermis to neural connections in the brain. Likewise, at the level of the species, natural selection reveals a process of evolution over generations. These processes raise philosophical puzzles: which came first, the chicken or the egg? At what point does human life begin? How is identity maintained over change, and where does identity reside? We suggest that some of the persistent difficulties in reasoning about change have their roots in essentialist assumptions about the nature of biological kinds. The inevitability, enormity, and centrality of biological change is at odds with the essentialist axioms of stability and immutability, and may lead to false starts, dead ends, and misconceptions in the history of science, and in science education. In this section we discuss two such biological change processes, metamorphosis and evolution, and how they are influenced by essentialism. Whereas metamorphosis involves changes in an individual, evolution involves changes in a species. Some of the same sorts of issues arise in both, as we shall explain.

Development and Metamorphosis

There are many natural changes that animals undergo during their lifespan (e.g., increases in size due to growth or changes in color due to seasonal variation). Yet, some of the most puzzling changes involve transformations in which the appearance of an animal is dramatically different across life stages – the quintessential example being metamorphosis. For scientists and laypeople alike, these types of biological changes are particularly perplexing because they stand in sharp contrast to our underlying essentialist belief that an individual's features (and, by extension, category membership) should be stable. This may lead to either of two misunderstandings about metamorphosis: (a) denial of the persistence of an individual through these transformations, or (b) assuming a greater degree of continuity through change than truly exists. There is evidence for both of these misconstruals in early scientists' views of metamorphosis and also in lay beliefs about these transformations, specifically those of young children.

First, it appears that people may treat the life stages present in metamorphosis as distinct, unrelated kinds and reason that an individual organism cannot persist through this kind of change. In the earliest scientific theory of metamorphosis, William Harvey (1651/1847) proposed that metamorphosis involved an organism that changed 'race' (species), specifically that an individual's original matter

decomposed and from this matter a new animal was spontaneously generated – a belief we might see as analogous to an individual changing from cat to dog (Cobb, 2007; see also Bruguière, Perru and Charles, 2018). This construal of metamorphosis appears to be the result of a difficulty in overcoming an essentialist expectation of category immutability.

A strikingly similar set of beliefs can be seen in how young children reason about these transformations. When young children are asked to select between two possible outcomes of change, they are more likely to select a category-match (e.g., a similar looking caterpillar) than a continuity-match (e.g., a butterfly), suggesting that they see an organism's category membership as stable over time (Herrmann et al., 2013; Rosengren et al., 1991). Further, young children's judgments of identity appear to be tied to an individual's category membership: when five-year-olds are asked to reason about the persistence of an individual following metamorphosis (e.g., "Is this ANNIE?"), they do not judge the caterpillar and the butterfly to be the same individual, suggesting a reliance on an object's persisting kind to reason about its identity (Marchak, 2017). Consistent with this interpretation, Marchak also found that highlighting a common category through change helped children to reason about metamorphosis as involving a persisting individual. Specifically, five-year-olds were more likely to judge a caterpillar and a butterfly to be the same individual when they were asked to reason about the transformation using a category label that applied to the organism both before and after the transformation (i.e., "insect") than one that applied only to the pre-transformation organism (i.e., "caterpillar"). Taken together, these results show that our intuitions about metamorphosis may be shaped by our underlying essentialist belief that an organism's category membership should be stable over time.

Second, children also assume greater continuity through the process of development than is truly appropriate. For instance, when asked to reason about people before conception, young children judge an individual's bodily and mental states to be stable and present before birth (Emmons and Kelemen, 2014). More research is needed to explicitly test whether analogous assumptions of continuity are found in reasoning about metamorphosis. For example, as a young child, one of the authors of this chapter thought that the caterpillar remained intact, forming the thorax and abdomen of the butterfly – that is, that a butterfly was essentially a caterpillar that sprouted wings. Recent scientific studies of metamorphosis, in contrast, show that unlike our essentialist assumption there is limited continuity through change – the body parts of the caterpillar are decomposed and new structures are formed (Lowe et al., 2013).

Evolution

Polls regularly indicate that only about half of U.S. adults say they believe in Darwin's theory of evolution, and even among those who accept evolution, misunderstandings abound (Rosengren et al., 2012). Although there are multiple

factors that contribute to this resistance (e.g., religious teachings, teleological biases, and difficulty considering complex processes and deep time; Rosengren et al., 2012), essentialist reasoning is likely one contributor. Specifically, Gelman and Rhodes (2012) argued that four of the components of essentialism sketched out in Table 12.2 pose obstacles to understanding and acceptance of evolutionary theory: strict boundaries, homogeneity, inherent causes, and stability. (Gelman and Rhodes also discussed a fifth obstacle regarding ideal essences, which is outside the scope of this chapter.) This section provides a brief overview and update of Gelman and Rhodes's argument.

Strict Boundaries

Acceptance of evolution requires understanding that boundaries between species are neither strict nor absolute. New generations continually bring changes, and over many generations boundaries themselves may shift. In contrast, essentialism entails an intensification of category boundaries. Children treat boundaries differently when reasoning about essentialized categories (e.g., the boundary between lions and tigers) versus non-essentialized categories (e.g., the boundary between cups and bowls). Although children and adults understand that hybrids exist as special cases (e.g., a mule is the offspring of a female horse and a male donkey), they view boundaries for most animal kinds as absolute rather than a matter of degree (Rhodes and Gelman, 2009a) and as objective rather than a matter of convention (Rhodes and Gelman, 2009b). The link between essentialism and strict boundaries can also be seen in individual differences: adults who endorse higher levels of essentialism are also more likely to support boundary-enhancing public policies, such as building a wall along national boundaries, even after controlling for education level, conservatism, religiosity, and LGBTQ attitudes (Roberts et al., 2017). Thus, a belief in absolute and objective boundaries may result in outright rejection of evolutionary processes. In the words of Gelman and Rhodes, "If an animal cannot be a semi-X, then how can one understand the evolutionary change from X to Y?" (2012, p. 10).

Homogeneity

One of the greatest challenges that laypeople face in understanding evolution is appreciating the importance of variability (Speth et al., 2014). Both adults and children may deny variability within a kind, underestimate how much variability exists within a kind, or treat variability as merely superficial (e.g., outward appearances may differ across individuals, but genes are identical). Although children are more prone to these errors than adults (e.g., Rhodes and Brickman, 2010), adults fall back on this error under speeded conditions (Shtulman and Harrington, 2016). A tendency to gloss over differences in favor of central tendencies can be found in well-known psychological processes such as prototype formation and stereotyping (Murphy, 2002) as well as in the structure of language itself. That is, in all the

world's languages, generalizations are typically framed using generic expressions, such as "Lions have manes" or "Sharks attack humans", even though only male lions have manes and most sharks don't attack humans (Moravscik, 1994).

Without a full appreciation of variation, natural selection is either rejected or misunderstood. A compelling demonstration of the latter comes from research examining what people think is changing over evolutionary time (Shtulman, 2006; Shtulman and Schulz, 2008). Although some adults endorse the basically correct Darwinian "variational" account, in which the distribution of traits within a population shifts over generations (e.g., a given trait becomes more widely distributed within successive generations, such that members of the kind will differ from one another in whether or not they possess the trait in question), others endorse an incorrect "transformational" account, in which the entire species gradually changes over generations (for example, if a moth population shifts from white-winged to black-winged, successive generations will be gradually darker) (Shtulman and Schulz, 2008). Importantly, adults who endorsed the transformational account showed lower levels of understanding evolution. The authors suggested that essentialism "leads individuals to devalue within-species variation and, consequently, to fail to understand natural selection" (p. 1049). A difficulty appreciating within-kind variation also has consequences for misconstruing evolution in the case of antibiotic resistance. For example, over one-third of undergraduate college students, including advanced biology majors, agreed with the statement, "Individual bacteria are genetically similar and equally likely to be killed by an antibiotic" (Richard, Coley and Tanner, 2017). This misunderstanding leads to attributing antibiotic resistance to non-evolutionary factors, such as the bacteria acting in goal-directed ways to meet their needs.

Inherent Causes

Population thinking can be a difficult level of analysis for people to grasp, as it conflicts with a heuristic assumption that causes inhere in individuals (Cimpian and Salomon, 2014), as well the essentialist assumption that an organism's features are due to its inner essence (e.g., a 'dog essence' within each dog that causes it to have its morphological and behavioral features). The work cited above by Shtulman and colleagues demonstrates not only a tendency to assume homogeneity, but also a tendency to focus on individuals versus populations as the locus of evolutionary change. Viewing the individual as the source of change can also lead to misconstruing evolutionary change as goal-driven, such that animals will develop the features they need, and even pass them along to their offspring (Evans et al., 2009; Ware and Gelman, 2014).

Stability

The essentialist assumption of stability does not deny the existence of change, but rather treats such change as superficial rather than relevant to an animal's

underlying essence. In direct contrast to this assumption, natural selection results in altogether new species. Today we see this resistance to species change most dramatically in young children, who seem to treat animal kinds as constants in an unchanging world. They explicitly deny that animals can undergo change (Samarapungavan and Wiers, 1997), and have difficulty understanding that there ever was 'a very first tiger' – instead assuming that members of a species always existed (Evans, Mull and Poling, 2002). However, this is not strictly a problem that children face. Adults – who have learned about evolutionary change – also resist full acceptance of the changes entailed by evolutionary theory. Some adults dispute scientific conclusions and deny the possibility of evolution; others accept that evolution occurs but limit it to featural modifications within unchanging species (i.e., acceptance of micro-evolution but not macro-evolution); others claim that it applies to non-human animals, but not people (Rosengren et al., 2012).

Overall, essentialism seems to be a formidable obstacle to understanding evolution. Even those students who use scientific concepts to explain evolutionary change are prone to essentialist explanations, which co-exist alongside more accurate explanations (Opfer, Nehm and Ha, 2012), demonstrating that essentialist notions are not easily overcome by the introduction of scientific constructs.

Essentialism and Genes

Transformative discoveries in the biological sciences over the past several decades have revealed the role of genes in the inheritance of numerous human traits. Most formally educated adults in the United States and other industrialized countries learn about genes in school and from the media (Thomas, 2000), and thus are aware of genes as biological entities that are linked to personal characteristics. Laypeople may be exposed to evidence for the heritability of a wide range of human attributes (including disease and temperament, for example), they may learn the genetic bases of biological sex (which informs the social construction of gender), and they may have heard that there are genetic contributions to phenotypic aspects of human variation (which inform the social construction of race). Genes are certainly involved in all of these domains, in the sense that they (along with environmental factors) are always a component of complex human behavior. These scientific advances have far-reaching implications for a wealth of issues (health care, public policy, etc.) and thus are crucial for the public to understand.

At the same time, genes are often understood *not* as interacting within a highly complex, multi-factor biological system that is susceptible to environmental and epigenetic influences (Jamieson and Radick, 2013), but rather in a manner that is strikingly similar to how people construe essences. This tendency is called genetic essentialism (Kampourakis, 2017; Nelkin and Lindee, 1995), which Dar-Nimrod and Heine have characterized in this way (2011, p. 801):

> The defining elements of psychological essentialism (i.e., immutable, fundamental, homogeneous, discrete, natural) are similar to the common lay perception of genes. Such similarity suggests that members who are assumed to share a distinct genetic makeup are also assumed to share their essence. People's understanding of genes may thus serve as an essence placeholder, allowing people to infer their own and others' abilities and tendencies on the basis of assumed shared genes. The tendency to infer a person's characteristics and behaviors from his or her perceived genetic makeup is termed genetic essentialism.

Genetic essentialism and its influence on lay perceptions of science are especially timely, given the proliferation of genetic explanations in recent years. On the basis of sometimes scant evidence, genes have been offered as potential explanations for a wide variety of human variation in perceptions, attitudes, and behaviors (Moore, 2013) – including criminality, mental illness, intelligence, gender-linked attributes, racial differences, luck, promiscuity, and success in life (e.g., Heine et al., 2017; Kendler, 2006; Shostak et al., 2009).

A number of the themes reviewed in the prior section on biological change re-emerge when considering the implications of genetic essentialism. To the extent that genes are construed in an essence-like manner, they may be viewed as resulting in categories that have strict boundaries and rich inductive potential, as being homogeneous within a category, as constituting an inherent cause, and as being stable over time. For one, people often seem to assume that there is a single "gene for" attributes with complex origins (e.g., optimism, autism, musical talent, schizophrenia) (Kendler, 2006). For example, 76% of a national U.S. sample incorrectly endorsed the claim, "Single genes directly control specific human behaviors" (Christensen et al., 2010), even though mental and physical traits cannot be directly inferred from one's genes (Bishop, 2009). Second, genetic essentialism implies that genes are destiny, and that environmental factors and personal control have little to no influence on phenotype. Accordingly, adults often mistakenly interpret genetic attributions as implying that a characteristic is fated, deterministic, and beyond environmental manipulation, treatment, or personal control (Gould and Heine, 2012; Kampourakis, 2017; Moore, 2013). For example, people who read an essay suggesting that there is a gene for obesity are more likely to 'over-indulge' when given an opportunity to eat a snack of cookies (Dar-Nimrod et al., 2014). In reality, of course, genes are not destiny (Nelkin and Lindee, 1995).

Genetic attributions can also foster a belief that categories are discrete, leading people to underestimate within-category genetic variability and exaggerate between-group genetic differences (Plaks et al., 2012). For example, people more often than not endorse the statement that "members of a given race are always more genetically homogeneous than members of different races" (Christensen et al., 2010), even though the degree of genetic variability among people of a given race is just as high as the degree of genetic variability across races (Templeton, 1998). Genetic attributions also imply powerful (and inaccurate) causal consequences,

such that receiving genes from a criminal may increase a person's tendency to display criminal tendencies (Meyer et al., 2013). As another example, genetically modified organisms are generally safe (Panchin and Tuzhikov, 2017), but the idea of combining genes from kinds that are viewed as having distinct essences often strikes people as unnatural, inherently dangerous, and even disgusting (Blancke et al., 2015). Fear of GMOs may likewise reflect an essentialist assumption that natural processes are more trustworthy than artificial human intervention.

Exposure to scientific findings about genetics does not necessarily reduce or eliminate essentialist beliefs, and media portrayals may even confirm or heighten them (Dar-Nimrod and Heine, 2011). For example, some years ago, a scientific study suggesting a genetic marker for male sexual orientation (Hamer et al., 1993) received enormous attention in the media, where this finding was characterized as revealing "the gay gene" (Conrad and Markens, 2001). Similarly, reading an essay that attributes gender differences in mathematics to genes reduced women's performance on a standardized math test, thus apparently confirming a 'genetic-as-fate' misconception (Dar-Nimrod and Heine, 2006). Likewise, 8th graders who heard genetic explanations of race in a classroom context held more strongly essentialized views of race (Donovan, 2014). More generally evidence indicates that genetic attributions often reinforce social inequalities and predict stereotyping and prejudice (Keller, 2005). In the study of genetics, science and lay beliefs can be mutually reinforcing – portrayals of genetic findings in the media may oversimplify and reflect scientists who embody essentialist assumptions in their research, and they may communicate these findings in ways that increase essentialist misconceptions (Conrad, 1997).

More research is needed to examine whether children treat genes in an essentialist manner from their first introduction to the concept, or whether certain kinds of experiences (formal instructional or informal conversational) introduce these beliefs. Some have argued that simple Mendelian examples of genes foster some of the most common misconceptions (Jamieson and Radick, 2013), whereas others discuss the role of popular media portrayals (e.g., Nelkin and Lindee, 1995), commercially available genetic testing (Heine, 2017), or political motivations (Shostak et al., 2009; Suhay, Brandt and Proulx, 2017) in propagating distorted (essentialist) views of genetic science.

Conclusions

In this chapter, we have provided several examples of how essentialism distorts reasoning about biological entities and processes. We examined biological change in both individuals (metamorphosis) and kinds (evolution) as well as genetic science. Across these varied phenomena, five essentialist assumptions about natural kinds distort how scientists and laypeople alike make sense of the natural world: these are strict boundaries, homogeneity, inherent causes, stability, and inductive potential.

This review is not meant to be exhaustive. Another broad class of examples that underlie any scientific investigation are debates about taxonomy, such as how best to classify psychiatric illnesses (Borsboom et al., 2016), emotions (Barrett, 2017b), or race (Pauker et al., 2016). Despite widely varying content across these domains, they share fierce disagreements that turn on whether one embraces or rejects essentialism. Such disputes are not merely academic; they have direct consequences for people's lives, in how medical, legal, and political systems are structured. In the extreme, and most horrifically, consider the consequences of essentialist theories of race for eugenics, embraced by prominent scientists of the 19th and 20th centuries, such as Galton and Terman.

Essentialist assumptions may also limit who engages in scientific practice to begin with. For example, beliefs about how gender maps onto talent and ability constrain both children's and adults' expectations about who is well-suited to science (Bian, Leslie and Cimpian, 2017; Leslie et al., 2015). By the age of six, children endorse the stereotype that men have greater innate intellectual abilities than women. This attitude is important to consider because it influences young girls' interest in engaging in activities such as science that require one to be 'smart'. Though the future for women in science may appear bleak, there is promising new evidence that children are now more likely to draw a woman when asked to draw a scientist than they were fifty years ago, suggesting that overall attitudes may be changing (Miller et al., 2018). The reach of essentialism thus has implications for science education, for everyday practice, and for the conduct of science itself.

Although we have focused primarily on lay beliefs within the biological domain, Lisa Barrett suggested that essentialism may in fact be the starting-point for all sciences – not only in biology but also in physics and chemistry:

> … the history of science can be read as a long, slow march away from essentialist thinking, discovering that universal laws are actually contextual (e.g., in physics, with the discovery of quantum mechanics) and discovering that variation is meaningful and is not in error (e.g., in biology, with Darwin's [1859/1964] *On the Origin of Species*, and then again a century later with the study of epigenetics and genomics).
>
> (Barrett, 2017a, p. 22)

In the field of psychology, too, essentialist thinking often appears to underlie claims about the fundamental nature of humans (a point argued by Oyama, 1985 and Siegler, 1996).

If indeed scientists are susceptible to the same reasoning biases as less informed adults, is it possible for the scientific enterprise to proceed apart from such errors? How does a biased mind itself detect human bias? Are there methods and procedures that can insulate us from essentialism? This is a difficult issue with no easy solution. One obstacle is that essentialism is deeply engrained in human cognition and may itself be a consequence of the ease and

automaticity of our categorization capacities. From early infancy, we have an impressive capacity to extract complex patterns from environmental stimuli, swiftly and automatically (e.g., Saffran and Kirkham, 2018). Because we are generally unaware of our own role in detecting these patterns, we may construe them as existing out in the world rather than as constructed by internal processes. In Barrett's words, "The human brain is so effective at creating similarities that it fails to recognize its own contributions to category formation. The result is naive realism" (2017a, p. 20). As noted earlier, naive realism is itself a core assumption that underlies essentialism. A second obstacle is that empirical evidence alone may not always be sufficient to counteract essentialism. In the extreme case, Barrett argues that essentialist construals are non-falsifiable: "Psychological essentialism permits scientists to posit a hypothetical or unseen essence in the absence of any evidence whatsoever of what the essence might be" (2017a, p. 22). Even in the face of apparent counterevidence, scientists may hypothesize that counterexamples are 'mere' exceptions, that messiness in the data reflects noise in the emitted signal rather than noise in the underlying construct, that measurements weren't precise enough, or that the technology isn't sufficiently advanced.

Nonetheless, we suggest that a comparative approach holds great promise in challenging essentialism and bringing to light alternative frameworks – whether one compares across ages, across cultures, or across species (see also Gelman, 2019). Examining belief systems other than one's own throws into sharp relief different conceptual perspectives, challenging one's assumptions about which concepts are necessary and which concepts are foundational. Our own work has focused on children. Seeing the biases in childhood allows one to see more clearly – and guard against – the same biases in adults. In a sense, children provide a magnifying glass to the essentialist biases that adults share. We suggest that the puzzle of trying to understand minds other than one's own, may actually help solve the puzzle of how the human mind can break out of its limitations to study itself. In this way, scientists can engage in the sort of "cognitive conflict" that Kampourakis (2014) has argued helps students learn by challenging and replacing their existing frameworks.

References

Ahn, W.K., Kalish, C., Gelman, S.A., Medin, D.L., Luhmann, C., Atran, S., Coley, J.D. and Shafto, P., 2001. Why essences are essential in the psychology of concepts. *Cognition*, *82*(1), pp. 59–69.

Barrett, L.F., 2017a. Categories and their role in the science of emotion. *Psychological Inquiry*, *28*(1), pp. 20–26.

Barrett, L.F., 2017b. *How emotions are made: The secret life of the brain*. Boston, MA: Houghton Mifflin Harcourt.

Bian, L., Leslie, S.J. and Cimpian, A., 2017. Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355* (6323), pp. 389–391.

Bishop, D.V.M., 2009. Genes, cognition, and communication. *Annals of the New York Academy of Sciences*, *1156*(1), pp. 1–18.

Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J. and Van Montagu, M., 2015. Fatal attraction: the intuitive appeal of GMO opposition. *Trends in Plant Science*, *20*(7), pp. 414–418.

Borsboom, D., Rhemtulla, M., Cramer, A.O.J., Van der Maas, H.L.J., Scheffer, M. and Dolan, C.V., 2016. Kinds versus continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological Medicine*, *46*(8), pp. 1567–1579.

Bruguière, C., Perru, O. and Charles, F., 2018. The concept of metamorphosis and its metaphors. *Science and Education*, 27(1–2), pp. 113–132.

Christensen, K.D., Jayaratne, T.E., Roberts, J.S., Kardia, S.L.R. and Petty, E.M., 2010. Understandings of basic genetics in the United States: results from a national survey of black and white men and women. *Public Health Genomics*, *13*(7–8), pp. 467–476.

Cimpian, A. and Salomon, E., 2014. The inherence heuristic: an intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*(5), pp. 461–480.

Cobb, M., 2007. *The egg & sperm race: The seventeenth-century scientists who unravelled the secrets of sex, life and growth*. London: Pocket Book.

Conrad, P., 1997. Public eyes and private genes: historical frames, news constructions, and social problems. *Social Problems*, *44*, pp. 139–154.

Conrad, P. and Markens, S., 2001. Constructing the 'gay gene' in the news: optimism and skepticism in the US and British press. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness, and Medicine*, *5*(3), pp. 373–400.

Dar-Nimrod, I., Cheung, B.Y., Ruby, M.B. and Heine, S.J., 2014. Can merely learning about obesity genes affect eating behavior? *Appetite*, *81*, pp. 269–276.

Dar-Nimrod, I. and Heine, S.J., 2006. Exposure to scientific theories affects women's math performance. *Science*, *314*(5798), pp. 435–435.

Dar-Nimrod, I. and Heine, S.J., 2011. Genetic essentialism: on the deceptive determinism of DNA. *Psychological Bulletin*, *137*(5), pp. 800–818.

Deeb, I., Segall, G., Birnbaum, D., Ben-Eliyahu, A. and Diesendruck, G., 2011. Seeing isn't believing: the effect of intergroup exposure on children's essentialist beliefs about ethnic categories. *Journal of Personality and Social Psychology*, *101*(6), pp. 1139–1156.

Donovan, B.M., 2014. Playing with fire? The impact of the hidden curriculum in school genetics on essentialist conceptions of race. *Journal of Research in Science Teaching*, *51*(4), pp. 462–496.

Emmons, N.A. and Kelemen, D., 2014. The development of children's prelife reasoning: evidence from two cultures. *Child Development*, *85*(4), pp. 1617–1633.

Evans, E.M., Spiegel, A.N., Gram, W., Frazier, B.N., Tare, M., Thompson, S. and Diamond, J., 2009. A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of Research in Science Teaching*, *47*(3), pp. 326–353.

Evans, E.M., Mull, M.S. and Poling, D.A., 2002. The authentic object? A child's-eye view. *Perspectives on Object-Centered Learning in Museums*, pp. 55–77.

Gelman, S.A., 2003. *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gelman, S.A., 2004. Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), pp. 404–409.

Gelman, S.A., 2019. What the study of psychological essentialism may reveal about the natural world. In A. Goldman and B. McLaughlin (eds.), *Metaphysics and Cognitive Science*. New York: Oxford University Press, pp. 314–335.

Gelman, S.A. and Davidson, N.S., 2013. Conceptual influences on category-based induction. *Cognitive Psychology*, *66*(3), pp. 327–353.

Gelman, S.A., Heyman, G.D. and Legare, C.H., 2007. Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, *78*(3), pp. 757–774.

Gelman, S.A. and Rhodes, M., 2012. Two-thousand years of stasis. In Rosengren, K.S., Brem, S.K., Evans, E.M. and Sinatra, G.M. (eds.), *Evolution challenges: Integrating research and practice in teaching and learning about evolution*. Oxford: Oxford University Press, pp. 200–207.

Gelman, S.A. and Wellman, H.M., 1991. Insides and essences: early understandings of the non-obvious. *Cognition*, *38*(3), pp. 213–244.

Gould, W.A. and Heine, S.J., 2012. Implicit essentialism: genetic concepts are implicitly associated with fate concepts. *PLoS One*, 7(6), p. e38176.

Hamer, D.H., Hu, S., Magnuson, V.L., Hu, N. and Pattatucci, A.M., 1993. A linkage between DNA markers on the X chromosome and male sexual orientation. *Science*, *261*(5119), pp. 321–327.

Harvey, W. and Willis, R., 1847. *The works of William Harvey*. London: Sydenham Society. (original work published 1651).

Heine, S.J., 2017. *DNA is not destiny: The remarkable, completely misunderstood relationship between you and your genes*. New York: Norton.

Heine, S.J., Dar-Nimrod, I., Cheung, B.Y. and Proulx, T., 2017. Essentially biased: why people are fatalistic about genes. *Advances in Experimental Social Psychology, 55*, pp. 137–192.

Herrmann, P.A., French, J.A., DeHart, G.B. and Rosengren, K.S., 2013. Essentialist reasoning and knowledge effects on biological reasoning in young children. *Merrill-Palmer Quarterly*, *59*(2), pp. 198–220.

Jamieson, A. and Radick, G., 2013. Putting Mendel in his place: how curriculum reform in genetics and counterfactual history of science can work together. In Kampourakis, K. (ed.), *The philosophy of biology, history, philosophy and theory of the life sciences*, Vol. 1.Dordrecht: Springer, pp. 577–595.

Kampourakis, K. (2014). *Understanding evolution*. Cambridge: Cambridge University Press.

Kampourakis, K. (2017). *Making sense of genes*. Cambridge: Cambridge University Press.

Keller, J., 2005. In genes we trust: the biological component of psychological essentialism and its relationship to mechanisms of motivated social cognition. *Journal of Personality and Social Psychology*, *88*(4), pp. 686–702.

Kendler, K.S., 2006. "A gene for…": the nature of gene action in psychiatric disorders. *Focus*, *4*(3), pp. 391–400.

Kuhn, D., 1996. Is good thinking scientific thinking. In Olson, D.R., and Torrance, N. (eds.), *Modes of thought: Explorations in culture and cognition*. New York: Cambridge University Press, pp. 261–281.

Leslie, S.J., Cimpian, A., Meyer, M. and Freeland, E., 2015. Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), pp. 262–265.

Lowe, T., Garwood, R.J., Simonsen, T.J., Bradley, R.S. and Withers, P.J., 2013. Metamorphosis revealed: time-lapse three-dimensional imaging inside a living chrysalis. *Journal of the Royal Society Interface*, *10*(84), p.20130304.

Marchak, K.A., 2017. Children's and adults' understanding of the persistence of individual artifacts. Doctoral dissertation. University of British Columbia, Vancouver, Canada.

Meyer, M., Gelman, S.A., Roberts, S.O. and Leslie, S.J., 2017. My heart made me do it: Children's essentialist beliefs about heart transplants. *Cognitive Science*, *41*(6), pp. 1694–1712.

Meyer, M., Leslie, S.J., Gelman, S.A. and Stilwell, S.M., 2013. Essentialist beliefs about bodily transplants in the United States and India. *Cognitive Science*, *37*(4), pp. 668–710.

Miller, D.I., Nolla, K.M., Eagly, A.H., and Uttal, D.H., 2018. The development of children's gender-science stereotypes: a meta-analysis of 5 decades of US Draw-A-Scientist Studies. *Child Development, 89*(6), pp. 1943–1955.

Moore, D.S., 2013. Current thinking about nature and nurture. In K. Kampourakis (ed.), *The philosophy of biology: A companion for educators*. New York: Springer, pp. 629–652.

Moravcsik, J., 1994. Essences, powers, and generic propositions. In T. Scaltsas,, D. Charles, and M.L. Gill (eds.) *Unity, identity and explanation in Aristotle's metaphysics*. Oxford: Oxford University Press, pp. 229–244.

Moya, C., Boyd, R. and Henrich, J., 2015. Reasoning about cultural and genetic transmission: developmental and cross-cultural evidence from Peru, Fiji, and the United States on how people make inferences about trait transmission. *Topics in Cognitive Science*, 7(4), pp. 595–610.

Murphy, G.L., 2002. *The big book of concepts*. Cambridge, MA: MIT Press.

Nelkin, D. and Lindee, M.S., 1995. The media-ted gene: stories of gender and race. In J. Terry and J.L. Urla (eds.), *Deviant bodies: Critical perspectives on difference in science and popular culture*. Indianapolis: Indiana University Press, pp. 387–402.

Opfer, J.E., Nehm, R.H. and Ha, M., 2012. Cognitive foundations for science assessment design: knowing what students know about evolution. *Journal of Research in Science Teaching*, *49*(6), pp.744–777.

Oyama, S., 1985. *The ontogeny of information: Developmental systems and evolution*. Cambridge: Cambridge University Press.

Panchin, A.Y. and Tuzhikov, A.I., 2017. Published GMO studies find no evidence of harm when corrected for multiple comparisons. *Critical Reviews in Biotechnology*, *37*(2), pp. 213–217.

Pauker, K., Xu, Y., Williams, A. and Biddle, A.M., 2016. Race essentialism and social contextual differences in children's racial stereotyping. *Child Development*, *87*(5), 1409–1422.

Plaks, J.E., Malahy, L.W., Sedlins, M. and Shoda, Y., 2012. Folk beliefs about human genetic variation predict discrete versus continuous racial categorization and evaluative bias. *Social Psychological and Personality Science*, *3*(1), pp. 31–39.

Rhodes, M. and Brickman, D., 2010. The role of within-category variability in category-based induction: a developmental study. *Cognitive Science*, *34*(8), pp.1561–1573.

Rhodes, M. and Gelman, S.A., 2009a. Five-year-olds' beliefs about the discreteness of category boundaries for animals and artifacts. *Psychonomic Bulletin and Review*, *16*(5), pp.920–924.

Rhodes, M. and Gelman, S.A., 2009b. A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*(3), pp. 244–274.

Rhodes, M. and Mandalaywala, T.M., 2017. The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4), pp. 1–18.

Richard, M., Coley, J.D. and Tanner, K.D., 2017. Investigating undergraduate students' use of intuitive reasoning and evolutionary knowledge in explanations of antibiotic resistance. *CBE Life Sciences Education*, *16*(3), pp. 1–16.

Roberts, S.O., Ho, A.K., Rhodes, M. and Gelman, S.A., 2017. Making boundaries great again: essentialism and support for boundary-enhancing initiatives. *Personality and Social Psychology Bulletin*, *43*(12), pp. 1643–1658.

Rosengren, K.S., Brem, S.K., Evans, E.M. and Sinatra, G.M. (eds.), 2012. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*. Oxford: Oxford University Press.

Rosengren, K.S., Gelman, S.A., Kalish, C.W. and McCormick, M., 1991. As time goes by: children's early understanding of growth in animals. *Child Development*, *62*(6), pp. 1302–1320.

Saffran, J.R. and Kirkham, N.Z., 2018. Infant statistical learning. *Annual Review of Psychology*, *69*, pp. 181–203.

Samarapungavan, A. and Wiers, R.W., 1997. Children's thoughts on the origin of species: a study of explanatory coherence. *Cognitive Science*, *21*(2), pp. 147–177.

Shostak, S., Freese, J., Link, B.G. and Phelan, J.C., 2009. The politics of the gene: social status and beliefs about genetics for individual outcomes. *Social Psychology Quarterly*, *72*(1), pp. 77–93.

Shtulman, A., 2006. Qualitative differences between naïve and scientific theories of evolution. *Cognitive Psychology*, *52*(2), pp. 170–194.

Shtulman, A. and Harrington, K., 2016. Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, *8*(1), pp. 118–137.

Shtulman, A. and Schulz, L., 2008. The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science*, *32*(6), pp. 1049–1062.

Siegler, R.S., 1996. *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Speth E.B., Shaw N., Momsen J., Reinagel A., Le P., Taqieddin R. and Long T., 2014. Introductory biology students' conceptual models and explanations of the origin of variation. *CBE Life Sciences Education, 13*, pp. 529–539.

Suhay, E., Brandt, M.J. and Proulx, T., 2017. Lay belief in biopolitics and political prejudice. *Social Psychological and Personality Science*, *8*(2), pp. 173–182.

Taylor, M.G., Rhodes, M. and Gelman, S.A., 2009. Boys will be boys; cows will be cows: children's essentialist reasoning about gender categories and animal species. *Child Development*, *80*(2), pp. 461–481.

Templeton, A.R., 1998. Human races: a genetic and evolutionary perspective. *American Anthropologist*, *100*(3), pp.632–650.

Thomas, J., 2000. Learning about genes and evolution through formal and informal education. *Studies in Science Education, 35*(1), pp. 59–92.

Ware, E.A. and Gelman, S.A., 2014. You get what you need: an examination of purpose-based inheritance reasoning in undergraduates, preschoolers, and biological experts. *Cognitive Science*, *38*(2), pp.197–243.

Wilkins, J.S., 2013. Essentialism in biology. In Kampourakis, K. (ed.), *The philosophy of biology: A companion for educators*, Dordrecht: Springer, pp. 395–419.

13

CAN SCIENTIFIC KNOWLEDGE SIFT THE WHEAT FROM THE TARES?

A Brief History of Bias (and Fears about Bias) in Science

Erik L. Peterson

> Another parable put he forth unto them, saying, the Kingdom of Heaven is likened unto a man which sowed good seed in his field. But while men slept, his enemy came and sowed tares among the wheat, and went his way. But when the blade was sprung up, and brought forth fruit, then appeared the tares also. So the servants of the householder came and said unto him, "Sir, didst not thou sow good seed in thy field? From whence then hath it tares?"
>
> He said unto them, "An enemy hath done this."
>
> The servants said unto him, "Wilt thou then that we go and gather them up?"
>
> But he said, "Nay; lest while ye gather up the tares, ye root up also the wheat with them."
>
> (Matthew 13.24–28, King James Version)

Introduction

Bias, half-truths, deception, spin, #alternativefacts – "lies, damned lies, and statistics."[1] No matter the label, concerns about the unreliability of reports by official media splash across televisions, newspapers, and websites all too regularly these days (Kampourakis, this volume; Peregrine 2017). Scientists, of course, worry about bias. Some are concerned that the pressure to increase research production is inevitably leading to overstatements about the significance of their findings and finding correlations in data where none truly exist, phenomena nicknamed p-hacking, data mining, or dredging (N. Young et al. 2008; S. Young & Karr 2011). Together, these behaviors endanger not only the pursuit of truth but the cultural and political cache granted to scientific studies. Concern about bias has sharpened recently, especially as the non-specialist public has become more skeptical about a

range of scientific claims: anthropogenic climate change, the safety of genetically modified organisms (GMOs), and the shared ancestry of humans and other primates, to list only a few of the most prominent.

Bias and the doubts about science it stirs up among non-scientists might be a contemporary worry, but it is hardly new. Early in the 1600s, Francis Bacon famously outlined four "idols" – barriers to secure knowledge – apparent in the natural philosophers of his Elizabethan era. A few decades later, René Descartes employed a method of hyper-doubt – doubt everything except that I am the thing doing the doubting – to ground unbiased knowledge. Throughout the eighteenth century, European "Encyclopaedists," such as Nicolas de Condorcet and Denis Diderot, promoted the pursuit of knowledge in natural philosophy as a model of how society as a whole could become more rational. But in the nineteenth century, it became clear that the methods employed by those very confident natural philosophers needed to be scrutinized more carefully. Major figures, such as the polymath William Whewell in the middle of the century and physicist-philosopher Pierre Duhem toward its end, took stock of the relationship between theory and evidence and the degree to which scientists' preexisting beliefs, theories, and values could alter that relationship. Hoping to identify and preserve the unvarnished core of scientific knowledge, a variety of logical empiricists in the early twentieth century wrangled with the precise methods of scientific observation. But through the middle of the last century, historians, philosophers, and sociologists of science grew increasingly skeptical that science free of individual biases and idiosyncratic perspectives could exist at all.

Among the greatest fears about bias in science is that, like tares among wheat, it is hard to detect and remove.[2] Honest-to-goodness hoaxes like Piltdown Man and "cold fusion" grab headlines (Gardener 1957). But what about the ordinary "value-ladenness" of scientific claims and observations? As we will see below, sociologists and psychologists claim our observations get interpreted through the observer's personal perspective, and no observer is completely free of their own preconceived notions, their values. How can we be sure that the biased perspectives of individual scientists are not acting as weeds, choking out good, objective science?

Even before our contemporary era of worries over anthropogenic climate change, the safety of vaccinations, GMOs, and so on, natural philosophers (who we would now call scientists) attempted to demonstrate that their method of pursuing knowledge was trustworthy and produced unbiased conclusions about the world. Starting with William Whewell in the mid-1800s, this essay traces both the assurances that science is trustworthy and the discovery of biases that undercut those assurances. Given this history, I argue that fears about the biases of individual scientists – frauds, quacks, and the like – are largely overblown. Historically, biases belonging to individuals have not been threatening to the claims of the scientific profession. Perhaps, as Whewell hinted, biases may turn out to be a regular part of science, a feature rather than a bug. Provided that scientists conform to certain broad socio-cultural norms, science tends to self-correct. This means that

individual biases will not damage the trustworthiness of the majority of our scientific pursuits. To tie this notion to the ancient parable: the tares will grow up with the scientific wheat to be later sorted and discarded.

However, this history also shows that scientists do not always follow these social norms and thus cannot put to rest every variety of bias. This is especially problematic as scientists resist historical and philosophical reflection regarding their disciplines, leading to a widening "two cultures" problem. A lack of reflective or "slow" thinking feeds into norm-breaking and a growing tendency to overlook potential bias (Kahneman 2011). Furthermore, the influence of vast sums of money in pharmaceutical research, natural gas and petroleum research, agricultural chemical research, and so on, has had a biasing effect on sciences involved in that work.

What follows is a rough chronology of the history of ideas about how science is supposed to work, concerns about bias undermining science, and the philosophical and sociological attempts to shore up those concerns. In the end, I offer reasons to be optimistic about the ability of certain fields to deal with bias tempered with caution, given some problematic, lingering aspects revealed by past episodes in science.

What Does Scientific Knowledge Look Like? 1840s–1940s

The work of the Reverend William Whewell blazed a trail in the study of knowledge about the natural world that continues to influence the way we understand how science and scientists work. Though his official title in 1837 was President of the Geological Society of London (following Sir Charles Lyell's first term as president), Whewell had already set his sights on higher things than rocks. That year, he published the three volume *History of the Inductive Sciences*, the foundational document for the entire field of History of Science. In it, Whewell tried to capture the painstaking growth of knowledge regarding the natural world from what he called an "early twilight among primeval wilds" to the "lofty and commanding position" of the nineteenth century (Whewell 1984 [1857], pp. 3–4). In that *History* and later revisions, expansions, and clarifications, such as *The Philosophy of the Inductive Sciences, Founded Upon Their History* (1840) and *On the Philosophy of Discovery* (1860), Whewell attributed the progress of knowledge to a two-step process of making broader and more general accumulations of observations about the natural world and then weaving these into the existing web of natural philosophical meaning. Almost anyone with basic skills could make observations themselves. But the weaving of those observations into the web of scientific meaning required something more. "Men of great sagacity" – for those were the types of gentlemen who Whewell believed solved the lingering problems of the natural world – wove "Perceptions" and "Ideas" into patterns using a back-and-forth process he dubbed "colligation" (Whewell 1860, p. 307). Though observations and experiments might form those perceptions, they were essentially meaningless until

the scientist superimposed a "New Element" on the collected observations, an "act of thought" that would bring these perceptions into a new arrangement with each other and with other facts already known to the scientist (Whewell 1847, v.2, p. 48). Of course, each "act of thought" was essentially a creative one, a new way of seeing or pulling together existing facts.

If Whewell was correct, that means that scientists do not approach problems like blank slates, ready to be shaped by data. Instead, scientists draw on their prior knowledge and expectations about the world to discern which experimental results are legitimate and which are unimportant. When Johannes Kepler witnessed individual points in the orbit of Mars, for instance, he was seeing an ellipse – the particular elliptical path that Mars travels around the sun. Whewell, however, insisted that the individual facts of observation – including the idea of 'this data pattern might inscribe an ellipse' – were *already infused* with Kepler's particular worldview and then explicated outward in a way that aligned his observations and the pattern that those observations described (Buchdahl 1971; Forster & Wolfe 1998).

Nevertheless, natural philosophers like Kepler didn't merely *guess*, insisted Whewell. Nor were they explicitly forcing data to fit their expectations. They carefully moved from possible explanations, to patterns that they saw through long work in science, to the observations themselves, and back again. New perceptions needed to be sewn into the web of already existing explanation; then the web would have to be re-woven to account for the new perceptions. Much like artists, scientists painted a version of a piece of the world. Unlike artists, scientists drew from phenomena located out there in the world, not feelings or impressions unique to the scientist, to create a whole quilt of scientific understanding (Laudan 1981).

Whewell believed good science possessed three features that elevated observations and explanations by these men of great sagacity above lesser explanatory schemes that remained open to bias. First, real scientific explanations made novel predictions by collecting old data and anticipating future discoveries made by individuals other than the original formulator of the concept. Secondly, true scientific ideas must be chains; they have *history*, in other words. The history of astronomy again provides a classic example. Tycho Brahe's observations of Mars led to Johannes Kepler's elliptical orbits and three laws accounting for those orbits, which later led to Robert Hooke's and Isaac Newton's concepts of universal gravitation, which, in turn led Edmund Halley and Alexis Clairaut to formulate the strange, yet regular, 76-year-long path of a previously mysterious comet. Thirdly, authentic scientific understanding expands our ability to account for different classes of information about the world – Mars to our Moon to Halley's Comet, in this example. Whewell called this explanatory expansion "consilience" (Whewell 1847, v.2, p. 469). True facts in one scientific domain become true facts in other ones, in ideal cases explaining problems in other fields unrelated to the original ones. Though not everyone agreed with him, predictiveness, fit within a longer chain of evidence, and consilience became the hallmarks of proper science after

Whewell. By implication, then, the work of biased scientists would fail to exhibit one or more of these three traits.

French physicist, mathematician, historian, and philosopher, Pierre M. M. Duhem (1861–1916) offered one of the most insightful revisions to Whewell's model and one that changed the conversation about bias in science. In his 1906 milestone, *The Aim and Structure of Physical Theory*, Duhem agreed with Whewell that theories do not just emerge from data but are constructed from a combination of observations and the scientist's assumptions about the world. But Duhem went further. In physics at least, asserted Duhem, facts are fully mixed with theoretical interpretations. One might say that it is "impossible to express fact in isolation from theory" (Ariew 2014). In fact, in a direction that Whewell didn't take, Duhem denied that when physicists received disconfirming evidence, they would reject hypotheses. Hypotheses, Duhem thought, are a mélange of assumptions, measurements, and beliefs, none that can be tested in isolation from others, none that determine the correct theory, since theory is underdetermined by evidence. If Duhem was correct about the interconnectedness of physical theory, then biases might lurk throughout different parts of a hypothesis. It would be difficult to know how to disinfect scientific concepts of all biases. They could appear all throughout the scientific work (Agassi 1983).

Nevertheless, bias need not be all that disconcerting, given Duhem's model of science. He argued that theories in physics don't actually explain reality – a philosophical stance that does not demote the importance of physics (Maiocchi 2000). Theories simply account for observations, for how the world appears. Laws are merely abstract, though very precise, conventions and mathematical representations. We judge them true or false based not on some metaphysical notion of reality but based on how the equations conform to appearances – whether they "save the phenomena," in other words. The best scientific processes will lead to highly accurate theories and laws, yet, even in this case, we don't have the logical mandate to assume our concepts represent underlying metaphysical reality (Duhem 1991 [1906]). By this account, bias ceases to be a threat, so long as physics and metaphysics keep to their proper places (Agassi 1983).

Between World War I and II, philosophers and mathematicians argued that, by employing logically rigorous statements built from simple and complete observational data, scientists and philosophers could avoid bias. In the Ernst Mach Society at the University of Vienna – better known as the legendary Vienna Circle – luminaries such as Moritz Schlick, Otto Neurath, Rudolf Carnap, Kurt Gödel, and Hans Hahn debated issues of epistemology and metaphysics similar to those raised by Duhem. The manifesto of these logical positivists/empiricists, *Wissenschaftliche Weltauffassung: Des Wiener Kreiss* (*Scientific World-Conception: The Vienna Circle*, 1929), advocated the strongest possible relationship between empirical observation and logical statements as the only reliable path to knowledge.

Perception Bias Complicates the Picture of Science, 1930s–1970s

Perhaps it is only a coincidence. But at nearly the same moment that the Vienna Circle called for knowledge built strictly from empirical observations, psychologists began uncovering perceptual biases that would put the pure observations they sought out of reach.

Though Francis Bacon, David Hume, and others harped on them centuries earlier, Anglo-American psychologists began intensive studies into the biases of perception itself only around World War II. E. G. Boring at Harvard University's Psychological Laboratory conducted comprehensive early studies examining bias in the abstract visual perception of objects. Faintly echoing Whewell, Boring concluded that perception was indeed shaped both by "basic sensory excitation" and by the context, or the observer's past experiences. In the case of scientists, this might include years of controlled experimental observations. These past experiences influenced the explicit attention devoted by the perceiver to the act of observing in a process Boring labeled "perceptual selection." Driven to some degree by hard-wired biological tendencies, the perceiver – even an expert scientist – "picks out and establishes what is permanent and therefore important…" (Boring 1946: 107). In other words, past experiences allow the observer to change even the basic process of perceiving. A team led by Leo Postman of Indiana University soon amended Boring's account. Through a series of experiments, Postman, Bruner, and McGinnies (1948) seemed to demonstrate that explicit perceptual selection alone could not account for everything going on even in the bare perception of objects. Instead, a whole packet of largely implicit "interests, needs, and values" hidden even from the perceiver shaped perception. What's more, they found that the closely-held values of the perceiver might influence not only *selection* of certain desired perceptions but *resistance* to other, undesired ones. The observer might unwittingly place "barriers against precepts and hypotheses incongruent with or threatening to the individual's values" in a kind of "avoidance of meaning" (Postman et al. 1948, p. 154). Unconsciously, even experts perceived some phenomena and screened out others simply because of their implicit values.

Study after study through the middle of the twentieth century showed that something as seemingly straightforward as perception was anything but. Biases popped up everywhere. Confirmation bias always posed a threat in science. But others were as hard to spot as tares among wheat. Anchoring, focus, or sequencing biases privilege kinds of information that appear close together in time, space, or thematic content. Availability biases privilege information that is simpler to recall (whether or not it comports with Ockham's Razor). When combined with the Hedwig von Restorff effect, which highlights dramatic first-hand experiences over mundane or secondhand accounts, availability biases lead observers to grasp case studies or anecdotal evidence more tightly than ordinary but statistically more likely evidence. Framing biases skew information depending on its presentation

rather than its content. Conjunction biases, regression biases, the gambler's fallacy, the sunk-cost fallacy, and other misunderstandings of probability deceive even experts into making incorrect predictions, mishandling data, mistakenly attributing findings to favored sources instead of to chance – and to hold firmly onto all of these misconceptions even when shown to be wrong (Friedman 2017; Kahneman 2011; Kahneman & Tversky 1982; Tversky & Kahneman 1983).

By the middle of the twentieth century, psychologists had shown that perception was not straightforward, but *constructed*. Usually this process of construction worked very well for us. We needed heuristics to make sense of a very complicated world. But these heuristics occasionally backfired – that is essentially what they meant by bias: a backfiring heuristic (Kahneman & Tversky 1973). In these psychologists' depictions of perception, observers danced between witnessing the favorable, guarding themselves against the unfavorable, and somehow actually observing something out there in the world. However, additional studies showed that the boundary between the thing being observed out there in the world and a simulacrum constructed by the observer's hopes and fears was very thin. Even expertise in a subject area did not shield the observer from making biased judgements, from having their heuristic views lead them astray (Tversky & Kahneman 1983).

Does the Social Structure of Science Protect It from Bias? 1940s–2000s

The philosophy of Karl R. Popper offered a way out of the problem of individual bias. In his *Logic of Scientific Discovery* [*Logik der Forschung*, 1934], Popper seemed to sidestep the issue of biased perception entirely and, in so doing, revealed one of the blind spots in the models of scientific epistemology going back to Whewell. Whewell's 'men of great sagacity' depiction placed the rest of the scientific community conveniently in the background. For Whewell and many philosophers who followed him, science meant the focused effort of a singular honed mind. Epistemology was individual. Today, the collection of competing laboratories and professors, lab managers, post-docs, graduate students, and massive funding apparatuses required to do science cannot be ignored so easily. Science has always been a collective and hierarchical enterprise. Epistemology is social (see de Ridder, this volume).

Perceptual bias in science became an issue because earlier philosophers and scientists, even the logical empiricists to a degree, implicitly founded their models of scientific knowing on Whewell's sagacious-man history of science. The possibilities of perversion through various perceptual biases (such as those listed earlier) would indeed threaten the authority of science – that is, if science was really about Kepler's perceptions or Galileo's perceptions in isolation.

Reflecting a process quite similar to Whewell's colligations, Popper argued instead that scientific knowledge advanced by conjectures made by one

scientist – an individual Newton or Einstein – and then attempted refutations of those conjectures by other scientists – Robert Hooke or Niels Bohr perhaps. The manner by which a scientist arrived at their conjecture did not have to be free of bias. In fact, Popper compared this stage of science – the individual *context of discovery* – to thinking up a musical theme or work of literature. The context of discovery needn't be open to logical analysis at all (Popper 2002 [1934/1959], pp. 7–8). Instead, it was the *context of justification* that mattered. Any individual conjecture stood only until it was refuted or falsified by the community of scientists. And any conjecture that could not be refuted even in principle (Popper asserted the entire field of astrology fit this characterization), he rejected as *unfalsifiable* and, therefore, not truly scientific in the first place. While Popper would only later explore the inherently social nature of this process, sociologists would reveal and elaborate on it, pushing worries about individual scientists and the potential for biases in their observations gradually to the background.

American sociologist Robert K. Merton provided the most well-known and long-lasting attempt to define the social structure of properly functioning science in a series of essays including, "Science, Technology and Society in Seventeenth Century England" (1938) and "The Normative Structure of Science" (1942; see Merton 1973). Like Popper's conjectures-and-refutations, Merton's model of the social structure of science could save the objectivity of science from the panoply of perceptual biases by segregating what scientists do during the discovery portion of their work from what scientists do during the justification phase (Mitroff 1974).

Merton listed four social norms that he saw governing the pursuit of scientific knowledge and arranged these norms in a memorable acronym: "CUDOS." First, science is a "Communal" (originally "communistic") endeavor; meaning scientific knowledge is open to all. Secondly, claims to accuracy or truth are evaluated without respect to the status of the discoverer or researcher. Race, class, gender, religion, institutional affiliation, funding – none of these features matter when it comes to scientific claims. Merton called this second norm "Universalism." Thirdly, Merton said scientists were "Disinterested," meaning they did not self-aggrandize but willingly submitted their work up for evaluation by the community of scientific judges. There could be no skirting around scrutiny – attempts to falsify, in Popper's conceptualization. The final two letters of Merton's acronym stood for "Organized Skepticism." By this Merton meant that scientists hold all claims to the same critical standards, no matter their roots in deeply held religious, cultural, political, or economic belief systems. The heat of scientific inquiry melted away all false claims to knowledge no matter their source. So long as they were honored, CUDOS norms ensured unbiased scientific knowledge in the long run. Taking Merton's model seriously, we might grant that individual scientists could be biased in their initial perceptions. But the whole social enterprise of science put pressure on scientists to conform. Ultimately, the community rejected aberrant work.

Merton and others amended and clarified these norms several times throughout the 1950s and 1960s, especially as scientists like Michael Polanyi attacked the

de-personalized view of science CUDOS promoted, insisting that scientific discovery and justification were both deeply personal (Polanyi 1958). Still, by the 1960s and 1970s, some sociologists grew concerned that Merton's norms were only honored in the breach. These sociologists pointed out that scientists disregarded all four aspects of CUDOS to chase power, money, and prestige. To the dismay of many science observers as well as scientists themselves – biologists Barry Commoner and Rachel Carson became some of the most outspoken critics – scientists poured their talents into nuclear, biological, and chemical weaponry, pharmaceuticals of questionable value, toxic fertilizers, pesticides, herbicides, food additives, and other products of commerce rather than into scientific insights – to say nothing of the scientists enabling the tobacco industry's fight against the science that smoking caused cancer. In the face of these developments, Merton-style norms appeared naïve (Feyerabend 2010 [1975]).

At a handful of UK universities in the 1970s, scholars clamored for a more robust critique of a scientific worldview that they felt had run off the rails. Barry Barnes, David Bloor, Harry Collins, John Henry, Donald MacKenzie, and Trevor Pinch together advocated a more robust sociology of scientific knowledge (SSK) that they called the "Strong Programme." Their SSK work promoted the perspective that scientists were neither privileged observers nor men of any particular sagacity. As no set of norms seemed to be restraining bias, the same social, political, and economic concerns shaped scientists as any others. A similar skepticism regarding the insulated view of scientists and their insights motivated one of the era's groundbreaking studies: *Laboratory Life*, a mid-1970s ethnography of Roger Guillemin's Salk Institute laboratory by Bruno Latour and Steve Woolgar. SSK studies like these cast doubt on the notion that scientists could get free enough from bias to make science more trustworthy than other kinds of knowing.

In the last two decades of the twentieth century, philosophers of science finally joined their colleagues in history, psychology, and sociology, to admit scientific knowledge *just is* socially constructed knowledge from discovery through justification to application. Unlike their SSK compatriots, however, philosophers such as Helen Longino and Philip Kitcher took the social nature of scientific knowledge to be an advantage rather than a liability. It transcends individual experiments or the viewpoints of a subset of experimentalists or methodologies, 'schools,' working groups, or research programs. Cooperative and competitive, collaborative and antagonistic, scientific knowledge passes the muster of peer review to publication and then the secondary gatekeeping process that follows publication: citation and incorporation into the discovery processes of other scientists (Longino 1990, p. 69). Real science doesn't exist without uptake in the broader scientific community – an insight akin to Whewell's concept of idea chains. Therefore, the most effective way to combat bias is not to pretend that scientists are free from bias but to insist that the scientific community allows for an unconstrained marketplace of ideas. In an arena of free exchange – a true democratization of science – competing biases of individual scientists or laboratories will mitigate each other. The most successful

experiments, predictions, methodologies, theories, etc., will rise to the top not merely because some elite sagacious man considers them so, but because they promote "projects to which specific groups would subscribe after reflective deliberation" (Kitcher 2001, p. 180). In other words, just as Francis Bacon hoped long ago, science would be judged successful when it best promoted human flourishing. The tug-of-war between competing camps within what Kitcher considers a well-ordered science ensures that, in the long run, no single bias can detract from that pursuit.

In other words, however numerous, perception biases and violations of Mertonian norms by *individual* scientists trouble ordinary science little. As long as the by-turns antagonistic and collegial tug-of-war between labs and perspectives, authors and reviewers continues, the gears of the ordinary scientific machine will turn, unperturbed by something so ephemeral as the bias of a single human observer. Those who doubt the efficacy of the scientific process at removing bias misunderstand or undervalue the balance between cooperation and competition in science and the long commitment to get science right over generations (Godfrey-Smith 2003, pp. 224–26).

How Contemporary Science Deals with Bias: The Case of p-hacking

Current worries about bias center around "p-hacking," especially in the social sciences. P-hacking is data manipulation – often unintentional, despite the name – so that a given finding meets the twin criteria of novelty and significance necessary to warrant publication in a respected scientific journal. It is by nature a violation of the Mertonian norm of disinterestedness. Given the history I outlined above, how are social scientists dealing with p-hacking?

P-hacking perverts the p-value (probability-value) statistical convention employed in many sciences. Probability-values address the question, "Am I observing a real and persistent effect or just some event that would occur randomly anyway?" A researcher wants to be able to say that the likelihood that the effect is random, and *not* due to the phenomenon being investigated, is quite low. Put another way: in order to show that there *is* some effect, we first assume that there *is not* an effect. Next, we settle on statistical rules that let us know whether the data we collect are consistent with the assumption that there *is not* an effect. Finally, we analyze our data; if our data is inconsistent with the assumption that there *is not* an effect – our null-hypothesis – then there must be an effect (Dallel 2012, n.1; Pearson 1904, p. 6; Sterne & Smith 2001). For historically contingent reasons, the p-value should be lower than 5 percent: meaning 5 percent or less of the time the phenomenon under study would occur even if the entity, state of affairs, or process presumed to cause the phenomenon was absent (Ghaemi 2009, pp. 35–36).

Researchers "p-hack" to render their data more significant by eliminating unfavorable data, changing parameters to make new correlations not actually

tested in the study, or merely by adjusting the reporting of data to make it look significant. P-hacking persists because of incentive structures that practically require a researcher to violate the Mertonian norm of disinterestedness. Scientific disciplines and university administrations pressure young researchers to publish their studies early and often in prominent peer-reviewed journals. Faced with overwhelming numbers of manuscripts to review – and with little incentive for researchers to suspend their own research in order to review the work of others for publication – journals tend to screen out work judged insufficiently surprising or with p-values greater than five percent.

The pressure to publish by any means necessary can be intense. Merton and Barber (1963) noted that scientists' ambition (the counter-norm to disinterestness) could further scientific knowledge so long as it was held in equilibrium with the norm. Today's pressure is not self-aggrandizement but desperation. Publishing too infrequently or in insufficiently prominent journals leads to permanent loss of employment. Yet, there is too little high-ranking journal supply to fulfill all the submitted paper demand. In the absence of properly vetted journals or journals with the freedom to publish exploratory work that does not achieve obvious novelty or significance, the socio-economic structure of the modern research university actually incentivizes a higher level of norm-breaking bias (Young et al. 2008). And it has spread across a much wider swath of the social sciences than the allowances Merton and others made for such counter-normative behavior.

Thankfully, scientific disciplines are beginning to address the problem of data manipulation (Sterne & Smith 2001). We now recognize that certain traits of scientific studies, signal greater vulnerability to p-hacking. Meta-analyses have shown ways to shore up that vulnerability (Fanelli et al. 2017; Ghaemi 2009). Journal editors in numerous fields are aware of the problem. Even scientific journalists have caught on (Gutting 2013; Lehrer 2010). Some social scientists are even suggesting new styles of scientific publication space to change the incentive structure (Tullett 2015). So, though the "tares" of p-hacking and other forms of data manipulation are real and disconcerting, they will not choke out the "wheat" of science over the long run. Greater scrutiny will weed out the problematic practices, find qualitative and quantitative metrics other than p-values to determine significance, increase avenues for publication, dampen bias, and restore trustworthiness.

Lingering Issues: Identity Biases, Meta-theoretical Disputes, and For-Profit Bias

Despite these promising preliminary fixes for the acute problem of p-hacking, historians of science caution against a too-sunny reading of science's built-in ability to root out all tares from the wheat field of knowledge. Three kinds of biases resist detection and removal. These include identity or in-group biases, meta-theoretical disputes – which can sometimes devolve into tribal biases – and broad distortions of scientific work for financial gain. While the history of science

gives us confidence that many individual biases will be corrected through ordinary internal social norms of well-ordered science, these other biases have tended to persist sometimes for whole generations.

There is little reason to suggest that scientists are immune to social biases such as in-group, tribal, or identity biases. They may, in fact, be more susceptible to this type of bias than non-scholars (West et al. 2012). Psychologists recognize in-group biases, the tendency to favor those that are part of one's group versus another group, as particularly pernicious since the in-group acts to reinforce the bias (Tajfel 1969). And these biases can do things other biases can't, such as skew seemingly bias-immune statistics generated by carefully conducted surveys using very large data sets (Stephan 1939). Whole corners of scientific thought may be twisted by unconscious biases of group identity, including class, race, ideology, nationality, or gender identifications. Whewell, for instance, located a chauvinistic identity bias operating among prominent French physicists in the 18th century (Whewell 1984 [1857], p. 385). Anthropologists, geneticists, psychologists, geologists, physicians, and mathematicians in Britain, France, Germany, and the USA aligned in providing vigorous scientific backing for white supremacy well into the twentieth century (Graves 2001; Peterson 2017). Through the 1980s, hydrologists, physicists, and engineers deployed biased mathematics and empirical data to back so-called Scientific Creationism in service of closely-held religious convictions. Intelligent Design advocates exploit systems biology, network theory, and pet concepts in physics to do the same today. Still more recently, social psychologists may have identified ideological biases operating unchecked within the field of social psychology itself (Haidt & Joseph 2007; Tierney 2011).

Meta-theoretical disputes resist resolution because they are arguments over *theories of* theories. Additional evidence presented in piecemeal fashion does not resolve these disputes, since what counts as a valid evidence in one explanatory scheme is seen as nonsense by another (Hein 1968). For these debates to move forward, both sides need solve what amounts to "experimenter's regress." Experimenter's regress occurs when there is both an unexplained phenomenon and an untested experimental procedure. The experimenter might test their procedure, if there was a known phenomenon to test it against, or might test the phenomenon of interest if the experiment itself was already trusted (Collins & Pinch 1998 [1993]). When neither side can agree to an appropriate metric by which to run the test and cannot decide on an appropriately well understood phenomenon to test a metric against, then both sides are stuck arguing over otherwise intractable problems that probably require historians, philosophers, and sociologists of science to resolve. Unfortunately, these meta-theoretical conflicts remain in limbo for multiple generations as disagreements become balkanized. As an example, the preeminent cancer researcher Robert Weinberg confessed that his preconceptions about the strength of his own meta-theoretical approach to biology – genetic determinism – may have led cancer research on a sort of wild goose chase after a small number of oncogenes for decades (Weinberg 2014).

Finally, and most seriously, financial interests have in the past, and may be currently incentivizing scientists to violate Mertonian sociological norms in ways that even well-regulated science cannot handle (Blumenthal et al. 1986; Ioannidis 2005; Krimsky et al. 1998). This was one of the main critiques brought by the SSK movement in the 1970s, and it still holds true today. Even relatively small groups of scientists who choose to promote biased science out of self-interest can do a great deal of harm both to science and to society when they connect with powerful political and commercial forces, including multinational biochemical giants such as Monsanto, Bayer, Corteva Agriscience, and Syngenta (e.g., Kelland 2018a & 2018b). Sometimes the financial bias is obvious and long-lasting. Notorious "merchants of doubt," including most prominently once-respected physicists Bill Nierenberg, Fred Seitz, and Fred Singer, worked for decades to downplay or undercut studies linking smoking and pesticides to lung cancer. More recently the think-tanks and political lobbying organizations they supported, such as the CO_2 Coalition, have taken aim against the science of climate change (Oreskes & Conway 2010). Also, in another example of for-profit bias, actress Gwyneth Paltrow deploys biased scientific studies and the testimonies of medical professionals who doubt established studies to sell quack medical products (Caulfield 2015; Porter 2003). But sometimes, the bias is not so obvious. The influence of vast sums of money from the pharmaceutical industry, agricultural chemistry and biology, coal, natural gas, and petroleum industries, and so on bias their associated sciences (Krimsky 2003, Prinz et al. 2011). It is hard to say just how much money in science biases results. Occasionally, accusations of such bias grow insistent enough to crest into news media headlines (Seife 2012; Whoriseky 2012). The American National Institute of Health, for instance, recently shut down a $100 million trial after it discovered interference in scientific studies by alcoholic beverage giants (Faust 2018). However, since few comprehensive studies of this kind of bias exist, there is little way to know for sure beyond hunches supported by anecdotes (Bekelman et al. 2003; Falk Delgado & Falk Delgado 2017). Perhaps only when the outcomes of applied science get bad enough or government regulators intervene do scientists cease denials of disconfirming results and instead closely reexamine the for-profit bias in their research (Colpo 2005). This reexamination is especially difficult – but especially necessary – when issues of human flourishing are on the line.

Conclusion

To conclude, the history of bias in science offers several lessons, including:

(1) Scientists colligate between newly discovered "facts" and concepts that they have already learned though past experience, education, and training. This process is conservative and does not privilege flights of fancy. But psychological and sociological studies of scientists suggest that perspectival biases afflict individual scientists nonetheless. More problematically, scientists are

often less willing to admit to their own bias, given cultural assumptions regarding the objectivity of scientists.

(2) Perspectival biases, though possibly pervasive, do not greatly threaten science. The social process of knowledge making baked into science largely mitigates perspectival biases. Even when science seems to be in the middle of a watershed bias moment, such as the contemporary p-hacking crisis, the cooperative-competitive structure of science rectifies the effects of bias.

(3) Nevertheless, history shows in-group or tribal biases continue to exist within even well-ordered science. Some meta-theoretical disputes devolve into tribalism with attendant biases. Finally, as the SSK community and others proclaimed decades ago, the influence of money and political power might be the most tare-like bias – almost impossible to assess and root out until fully grown next to the good science. But by then, especially after the biased science leads to outcomes that endanger human health and well-being, it might be too late.

Notes

1 "Statistics" was the interpretation of this phrase attributed to American cultural critic Mark Twain in 1906. Twain credited British Prime Minister Benjamin Disraeli for coining the phrase. But something akin to it must have been a commonplace saying, at least in Britain. In a meeting of the "X-club" in February 1886, Thomas H. Huxley reported that he and his comrades discussed the three kinds of unreliable sources: "liars, d----d liars, and experts." No source is quite certain when "statistics" replaced "experts" (Huxley 1901, p. 278).

2 In the biblical parable, "tares" probably refers to darnel (*Lolium temulentum*), a noxious ryegrass that looks like wheat in its early development. Analyzed for centuries, the parable is thought to refer to the inability of humans to distinguish morally good individuals or groups from bad ones, with the injunction to leave that level of judgement up to spiritual beings in some eschatological future (Keener 2009, pp. 386–387).

References

Agassi, Joseph (1983) "Theoretical Bias in Evidence: A Historical Sketch," *Philosophica* 31(1): 7–24.

Ariew, Roger (2014) "Pierre Duhem," *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2014/entries/duhem/.

Bekelman, Justin E., Yan Li, and Cary P. Gross (2003) "Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review," *Journal of the American Medical Association* 289: 454–65.

Blumenthal, G., Gluck M., Louis K.S., Stoto M.A., Wise D. (1986) "University–Industry Research Relationships in Biotechnology: Implications for the University," *Science* 232:1361–1366.

Boring, Edwin G. (1946) "The Perception of Objects," *American Journal of Physics* 14: 99–107.

Buchdahl, Gerd (1971) "Inductivist *versus* Deductivist Approaches in the Philosophy of Science as Illustrated by Some Controversies Between Whewell and Mill," *Monist* 55: 343–67.

Caulfield, Timothy (2015) I*s Gwyneth Paltrow Wrong About Everything? How the Famous Sell Us Elixirs of Health, Beauty & Happiness*. New York: Beacon Press.

Collins, Harry M. and Trevor Pinch (1998 [1993]) *The Golem: What You Should Know About Science*, 2nd edition. Cambridge: Cambridge University Press.

Colpo, Anthony (2005) "LDL Cholesterol: "Bad" Cholesterol, or Bad Science?" *Journal of American Physicians and Surgeons* 10(3): 83–89.

Dallel, Gerard E. (2012) *The Little Handbook of Statistical Practice*. Self-published, Kindle Edition.

Duhem, Pierre Maurice Marie (1991) *The Aim and Structure of Physical Theory*, Philip P. Wiener (trans.). Princeton: Princeton University Press. [Originally, Duhem, Pierre. (1906). *La Théorie Physique, son objet et sa structure*. Paris: Chevalier et Rivière.]

Falk Delgado, Alberto F. and Anna Falk Delgado (2017) "The Association of Funding Source on Effect Size in Randomized Controlled Trials: 2013–2015 – A Cross-Sectional Survey and Meta-Analysis," *Trials* 18(1): 125–134.

Fanelli, Daniele, Rodrigo Costas, and John P. A. Ioannidis (2017) "Meta-Assessment of Bias in Science," *Proceedings of the Academy of Sciences* 114(14): 3714–3719.

Faust, Jeremy Samuel (2018) "A Major Industry-Funded Alcohol Study Was Compromised. How Many Others Are Out There?" Undark (July 13), https://undark.org/article/mach15-alcohol-nih-industry-funding/.

Feyerabend, Paul (2010 [1975]) *Against Method: Outline of an Anarchist Theory of Knowledge*, 4th edition. New York: Verso Books.

Forster, Malcolm R. and Ann Wolfe (1998) "The Whewell-Mill Debate in a Nutshell," http://philosophy.wisc.edu/forster/220/whewell.html.

Friedman, Hershey H. (2017). "Cognitive Biases that Interfere with Critical Thinking and Scientific Reasoning: A Course Module," SSRN, http://dx.doi.org/10.2139/ssrn.2958800.

Gardener, Martin (1957) *Fads and Fallacies in the Name of Science*, 2nd edition. London: Dover.

Ghaemi, S. Nassir (2009) *A Clinician's Guide to Statistics and Epidemiology in Mental Health: Measuring Truth and Uncertainty*. New York: Cambridge University Press.

Godfrey-Smith, Peter (2003) *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago: University of Chicago Press.

Graves, Jr., Joseph L. (2001) *The Emperor's New Clothes: Biological Theories of Race at the Millennium*. New Brunswick, NJ: Rutgers University Press.

Gutting, Gary (2013) "What Do Scientific Studies Show?" New York Times (Apr. 25), https://opinionator.blogs.nytimes.com/2013/04/25/what-do-scientific-studies-show/.

Haidt, Jonathan, and C. Joseph (2007) "The Moral Mind: How 5 Sets of Innate Moral Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules," in: P. Carruthers, S. Laurence, and S. Stich (eds), *The Innate Mind*, Vol. 3. New York: Oxford, pp. 367–391.

Hein, Hilde (1968) "Mechanism and Vitalism as Meta-Theoretical Commitments," *Philosophical Forum* 1(2): 185–205.

Huxley, Leonard. (1901). *Life and Letters of Thomas Henry Huxley*, Vol. 1. New York: D. Appleton.

Ioannidis, John P. A. (2005) "Why Most Published Research Findings Are False," *PLOS Medicine* 2(8): e124, https://doi.org/10.1371/journal.pmed.0020124.

Kahneman, Daniel (2011) *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, Daniel and Amos Tversky (1973) "On the Psychology of Prediction," *Psychological Review* 80(4): 237–251.

Kahneman, Daniel and Amos Tversky (1982) "The Psychology of Preference," *Scientific American* 246: 160–173.

Keener, Craig S. (2009) *The Gospel of Matthew: A Socio-Rhetorical Commentary*. New York: Eerdmans.

Kelland, Kate (2018a) "Pesticides Put Bees at Risk, European Watchdog Confirms," Reuters (Feb. 28), www.reuters.com/article/us-science-europe-pesticides/pesticides-put-bees-at-risk-european-watchdog-confirms-idUSKCN1GC18G

Kelland, Kate (2018b) "Special Report: WHO Cancer Agency 'Left Out Key Findings' in Benzene Review," Reuters (Feb. 28), www.reuters.com/article/us-who-iarc-benzene-specialreport/special-report-who-cancer-agency-left-out-key-findings-in-benzene-review-idUSKCN1GC1ND

Kitcher, Philip (2001) Science, *Truth, and Democracy*. Oxford: Oxford University Press.

Krimsky, Sheldon (2003) *Science in the Private Interest: Has the Lure of Profits Corrupted Biomedical Research?* New York: Rowman & Littlefield.

Krimsky, Sheldon, L. S. Rothenberg, P. Stott, G. Kyle (1998) "Scientific Journals and their Authors' Financial Interests: A Pilot Study," *Psychotherapy and Psychosomatics* 67(4/5):194–201.

Laudan, Larry (1981) "William Whewell on the Consilience of Inductions," in: *Science and Hypothesis*. The University of Western Ontario Series in Philosophy of Science, Vol 19. Springer, Dordrecht.

Lehrer, Jonah (2010) "The Truth Wears Off: Is There Something Wrong with the Scientific Method?" The New Yorker (Dec. 13), www.newyorker.com/magazine/2010/12/13/the-truth-wears-off.

Longino, Helen E. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Maiocchi, Roberto (2000) "Pierre Duhem's the Aim and Structure of Physical Theory: A Book Against Conventionalism," *Synthese* 83(3): 385–400.

Merton, Robert K. (1973) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

Merton, Robert K, and Elinor Barber (1963) "Sociological Ambivalence," in: E.A. Tiryakian (ed.), *Sociological Theory, Values, and Sociocultural Change*. New York: Free Press, pp. 91–120.

Mitroff, Ian I. (1974) "Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists," *American Sociological Review* 39(4): 579–595.

Oreskes, Naomi, and Erik M. Conway (2010) *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury Press.

Pearson, Karl (1904) *On the Theory of Contingency and Its Relation to Association and Normal Correlation*. London: Dulau.

Peregrine, Peter Neal (2017) "Seeking Truth Among 'Alternative Facts'," The Conversation (Feb. 23), https://theconversation.com/seeking-truth-among-alternative-facts-72733.

Peterson, Erik L. (2017) "Race and Evolution in Antebellum Alabama: The Polygenist Prehistory We'd Rather Ignore," in: Lynn C., Glaze A., Evans W., Reed L. (eds), *Evolution Education in the American South*. New York: Palgrave Macmillan.

Polanyi, Michael (1958) *Personal Knowledge: Towards a Post-Critical Philosophy*. New York: Harper & Row.

Popper, Karl (2002 [1934/1959]) *The Logic of Scientific Discovery*. New York: Routledge. [Published as *Logik der Forschung*, 1934; English translation, 1959].

Porter, Roy (2003) *Quacks: Fakers & Charlatans in Medicine*. London: Tempus.
Postman, Leo, Jerome S. Bruner, and Elliott McGinnies (1948) "Personal Values as Selective Factors in Perception," *Journal of Abnormal Social Psychology* 43(2): 142–154.
Prinz, Florian, Thomas Schlange, and Khusru Asadullah (2011) "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets," *Nature Reviews – Drug Discovery* 10(712), www.nature.com/articles/nrd3439-c1.
Seife, Charles (2012) "How Drug Company Money Is Undermining Science," Scientific American (Dec.), www.scientificamerican.com/article/how-drug-company-money-undermining-science/.
Stephan, Frederick F. (1939) "Representative Sampling in Large-Scale Surveys," *Journal of the American Statistical Association* 206: 343–52.
Sterne, Jonathan A. C., and George D. Smith (2001) "Sifting the Evidence – What's Wrong with Significance Tests?" *BMJ* 322(7280): 226–231.
Tajfel, Henri (1969) "Cognitive Aspects of Prejudice," *Journal of Social Issues* 25: 79–97.
Tierney, John (2011) "Social Scientist Sees Bias Within," New York Times (Feb. 7), www.nytimes.com/2011/02/08/science/08tier.html.
Tullett, Alexa M. (2015) "In Search of True Things Worth Knowing: Considerations for a New Article Prototype," *Social and Personality Psychology Compass* 9(4): 188–201
Tversky, Amos, and Daniel Kahneman (1983) "Extension Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review* 90(4): 293–315.
Weinberg, Robert A. (2014) "Coming Full Circle – From Endless Complexity to Simplicity and Back Again," *Cell* 157(1): 267–271.
West, Richard F., R. J. Meserve, and K. E. Stanovich (2012) "Cognitive Sophistication Does Not Attenuate the Bias Blind Spot," *Journal of Personality and Social Psychology* 103(3): 506–19.
Whewell, William (1847) *The Philosophy of the Inductive Sciences, Founded Upon Their History*, 2 vols. 2nd ed., London: John W. Parker.
Whewell, William (1984 [1857]) "Selections from History of the Inductive Sciences from the Earliest to the Present Time," in: Y. Elkhana (ed), *William Whewell, Selected Writings on the History of Science*, 3rd edition. Chicago: University of Chicago Press.
Whewell, William (1860) *On the Philosophy of Discovery: Chapters Historical and Critical*, London: John W. Parker.
Whoriskey, Peter (2012) "As Drug Industry's Influence Over Research Grows, so Does the Potential for Bias," Washington Post (Nov. 24), www.washingtonpost.com/business/economy/as-drug-industrys-influence-over-research-grows-so-does-the-potential-for-bias/2012/11/24/bb64d596-1264-11e2-be82-c3411b7680a9_story.html.
Young, Neal S., John P. A. Ioannidis, and Omar Al-Ubaydli (2008) "Why Current Publication Practices May Distort Science," *PLoS Medicine* 5(10): 1418–22.
Young, S. Stanley, and Allen Karr (2011) "Deming, Data, and Observational Studies," *Significance* 8(3): 116–20.

14

WHAT GROUNDS DO WE HAVE FOR THE VALIDITY OF SCIENTIFIC FINDINGS?

The New Worries about Science*

Janet A. Kourany

Introduction

Science is based on facts, not wishful thinking or revelation or speculation, facts that are systematically gathered by a community of enquirers through detailed observation and experiment. These facts are used to support the rest of science, the laws and theories and models and so on; and it is this grounding in facts that has made science the most trusted source of knowledge we have, distinguishing it from all other enterprises that claim to produce knowledge. Of course, the success of science has involved other factors besides its grounding in facts: a multitude of highly dedicated, imaginative contributors; a heady dose of genius now and then; a willingness to break with the ideas of the past; generous financial and social support; the availability of mathematics and other technological tools; and other factors as well. But, its grounding in facts is universally considered the most important – the absolutely crucial and indispensable – ingredient of science's success. This is why reasons for questioning science's factual grounding are cause for alarm.

And there are such reasons. For example, in order for the facts to offer a proper grounding for science, the facts must be independent of that science. And yet, since at least the 1960s some of the most prominent philosophers of science have charged that what science offers as its factual grounding is *not* independent of that science. Indeed, these philosophers of science have claimed that the facts of science are "theory-laden," that is, shaped by the theories that scientists accept – in fact, that scientists' own observations are shaped by those theories. Some philosophers have even claimed that what scientists offer as "the facts" is simply a convention, not justified by scientists' observations at all. As a result, what scientists take to be the facts has changed over time. Just as the theories of the past (even the

most spectacular ones such as Newton's theory of mechanics and Darwin's theory of evolution) have been overthrown or revised over time, so have the factual claims on which they were presumably based, giving way to new facts. The facts, rather than the cause of the changes, have thus appeared to be simply their effects. Needless to say, all this has seemed to threaten the validity of scientific findings. Call this the "old worries about science."

By contrast, what seemed to philosophers of science not to threaten that validity were various other apparently humdrum ways in which what science offers as the facts is shaped by science: how it depends on the areas of research the scientific community or its funders consider acceptable and important; how it depends on the particular questions scientists pursue in those areas, their methods and tools of observation and analysis; how it depends on the publishing choices of journal editors and book publishers; and so on. Such factors lead to the uncovering and showcasing of certain facts rather than others but do not preclude the uncovering of those other facts at other times. So, no threat to the validity of science seems in question. Call this the "old non-worries about science."

What I shall suggest is that the old non-worries about science are turning out to be far more worrisome than the old worries – are turning out, in fact, to be the *new worries about science*. With few exceptions, however, they have been the new worries of scientists, not philosophers of science. And this raises interesting questions about the role philosophers of science have played and might still play in dealing with these worries.

The Old Worries about Science

Start with the worries about the factual basis of science that occurred in the middle of the twentieth century. Voiced most prominently by American philosophers of science Thomas Kuhn, Paul Feyerabend, and Norwood Russell Hanson, they had had their groundwork laid much earlier by British philosopher of science Karl Popper. The assumption, at the time – among epistemologists as well as philosophers of science – was that, in order for the facts offered by scientists to form a proper foundation for science, two conditions would have to be met:

1. The reports of the facts would have to express only what the scientists directly observe as they pursue their research. That is, the reports cannot go beyond what these scientists directly observe, otherwise the reports will say more than what can be justified by the scientists' observations.
2. What the scientists directly observe would have to correspond to what is out there in the world, as measured by what others also observe. It has to contain nothing that is personal or subjective or idiosyncratic.

If these conditions are met, it was thought, then science will have a secure foundation.

But Popper and Hanson argued that these conditions cannot be met. To begin with, Popper ([1935] 1959, p. 76) argued that statements that purport to describe "immediate experience" are inherently more general than the experiences that call them forth. As he put it, "an 'immediate experience' is *only once* 'immediately given'; it is unique." Statements, on the other hand, even statements as pedestrian as "There is a glass of water on the table," are associated with an indefinite number and variety of such unique immediate experiences. Furthermore, such experiences can never actually *justify* the statements but can only motivate a decision to accept them. For, Popper argued, only statements can enter into justificatory relations with other statements. As a result, the empirical basis of scientific knowledge is ultimately constituted by decisions to accept unjustified statements of fact. These decisions are motivated by experiences but not justified by them.

So, Popper challenged condition 1. And Hanson (1958) challenged condition 2 (and in the process Popper's critique of condition 1 as well). For Hanson argued that when we accept a statement like "There is a glass of water on the table," we don't first have some kind of ineffable "immediate experience" that we then hypothesize to be a glass of water on the table. We simply see the glass of water on the table. The interpretation, if there is one, is simply there in the seeing from the outset. Of course, it takes knowledge to see all this. An infant cannot see what we see when we see the glass of water on the table. The infant must first learn the language that talks about glasses and water and tables. Only then will her visual field be organized in ways that reflect that linguistic knowledge. But this means that people who have learned very different languages will see very different things. It follows, Hanson argued, that scientists who have been trained within different theoretical traditions will simply see different things and thus report what they see in different ways. The facts they glean from their observations, in other words, will be different. But this means that condition 1 is satisfied but condition 2 is not.

Kuhn (1962) applied all this to the controversies he studied in great detail in the history of science. The reason scientific advocates of older theories, even the most brilliant ones, frequently took years to accept new alternative theories and sometimes were never able to accept them, Kuhn said, was that those scientists' training shaped their observations and hence what they took to be the facts in ways that the new theories simply could not accommodate. Indeed, when advocates of competing theories tried to compare their respective theories and reach a reasoned decision regarding the theory they should all accept, they inevitably ended up "talking through each other" (1962, p. 109). So theory choice in the history of science, Kuhn concluded, was a psychological affair of persuasion and gestalt switches and the like, not the reasoned comparison of fact and theory philosophers of science had always supposed. A careful look at the history of science, Kuhn argued, could support no other view.

Finally, Feyerabend (1965) maintained that nothing better could be hoped for, because there is no theory-neutral fact-stating language, and hence no theory-neutral facts, to do better with. Even our ordinary language is theory-laden with,

Feyerabend argued, very outdated scientific theories (think of our ordinary talk of sunrises rather than earthfalls, or colors and shapes that inhere in objects independently of the frames of reference from which they are observed). So, neither ordinary language nor any scientific language at our disposal can provide us with the theory-neutral facts with which to compare alternative theories. The upshot, said Kuhn, is that science is not progressing closer and closer to any truth fixed by nature, but is simply progressing away from "primitive beginnings."

Thus, in the hands of Popper, Hanson, Kuhn, and Feyerabend – four of the most gifted and historically informed philosophers of science of the twentieth century – the validity of science seemed completely undermined.

The New Worries about Science

Needless to say, philosophers of science were not happy with this result and spent the next decades of the twentieth century trying to respond. Their aim, of course, was to somehow salvage the validity of science. But scientists remained unconcerned. Instead, they forged ahead with their various research programs, and the results of those programs made possible a spectacular array of our modern conveniences – cell phones, computers, and the internet; oral contraceptives and vaccines; satellites and GPS navigation; and much, much more. The intellectual breakthroughs scientists achieved were stunning. If the mission of philosophers of science was to capture scientific rationality, philosophers were doing a poor job of it. For, scientific rationality seemed robustly healthy at precisely the moment that philosophers were struggling with its incurable ills.

But a different worry about science and its factual grounding *was* beginning to occupy the minds of scientists even while Kuhn and the other philosophers of science were writing about theirs. This different worry pertained, as before, to science's shaping of what it presents as the facts, but this time the shaping was being accomplished, not by using the language of some theories rather than others to report the facts, but simply by carrying out some kinds of factual investigations rather than others. Women scientists were concerned about this mode of shaping the facts, especially the women who entered the sciences in increasing numbers during the time of second wave feminism. What these women found reported in the outcomes of social and natural science investigations was a torrent of facts relating to men together with a dearth of facts relating to women. They found, for example, facts in archaeology about men's contributions to the great turning points of human evolution but no facts about women's contributions; facts in medicine about men's problems with heart disease and stroke and, later, AIDS as well as other diseases but few facts about women's problems with these diseases; facts in economics and political science and sociology about men's rationality and agency and leadership styles and abilities but no facts about such characteristics in women; and so on. Even the titles of the works these women scientists eventually produced – such as biologist Ruth Hubbard's "Have Only Men Evolved?" (1979),

or archaeologists Joan Gero and Margaret Conkey's *Engendering Archaeology: Women and Prehistory* (1991), or economists Marianne Ferber and Julie Nelson's *Beyond Economic Man: Feminist Theory and Economics* (1993), or health researcher Sue Rosser's *Women's Health – Missing from U.S. Medicine* (1994) – bespoke the low visibility, indeed near invisibility, of women and the high visibility of men in the accumulated facts of their disciplines. True, these women scientists did find facts relating to women in their disciplines – for example, a heady dose of facts regarding women and reproduction in medicine – but all too frequently the facts they did find, especially in psychology, regarded women's inferiority to men, where men, of course, were taken as the standard of comparison.

Twentieth century women scientists were not the first women to worry about this way of science's shaping of the facts however. Nineteenth century women, not permitted, like their twentieth century sisters, to enter the sciences, could still diagnose its shortcomings. And some, such as Caroline Kennard and Eliza Burt Gamble, did, and with gusto. As they saw it, women had been thought inferior to men – intellectually, socially, physically, and even morally inferior – ever since ancient times, and so women were never expected to play any significant roles in the great exploits and achievements of humankind. Hence, no serious attention to them was ever considered warranted. Still, as Eliza Burt Gamble pointed out in her *The Evolution of Woman, an Inquiry into the Dogma of Her Inferiority to Man* (1894, pp. vii–viii):

> With the dawn of scientific investigation it might have been hoped that the prejudices resulting from lower conditions of human society would disappear, and that in their stead would be set forth not only facts, but deductions from facts, better suited to the dawn of an intellectual age. … The ability, however, to collect facts, and the power to generalize and draw conclusions from them, avail little, when brought into direct opposition to deeply rooted prejudices.

So modern science simply followed the ancient tradition. The upshot was that the facts unearthed by modern scientific investigations, rather than undermining and displacing the old prejudicial picture of women, reinforced it instead.

By the beginning of the twenty-first century a second group of scientists had joined the feminists in worrying about the kinds of factual investigations pursued and not pursued in science, and how these shape science's representation of the facts and the conclusions drawn from the facts. Doubtless some of the most memorable of these scientists were the 2,000 from all over Canada who marched in white lab coats through Ottawa in July of 2012 carrying a coffin and tombstones. They then staged a mock funeral on Parliament Hill "to commemorate," as one speaker (then biology doctoral student Katie Gibbs) put it, "the untimely death of evidence in Canada" (Pedwell 2012, Smith 2012). Especially memorable, also, were the more than 800 scientists from 32 countries who wrote an open letter to

Canada's Prime Minister Stephen Harper thereafter in support of the marchers (Chung 2014). Among the actions that precipitated their protest:

- The Harper Administration had instituted sharp cutbacks in basic research and the overall funding of important research areas such as climate, energy, and environmental research. It had even tried to shut down world-class government research programs engaged with groundbreaking industrial pollution research and climate research, such as the Experimental Lakes Area research station and the Polar Environment Atmospheric Research Laboratory (*Nature* Editorial 2012).
- In place of all this government-run, environmentally relevant basic and applied research, the Harper Administration had pushed for government research partnered with industry and aimed at economic development (Hoag 2011).
- The Harper government had also placed new restrictions on government scientists that impeded the free flow of information both among these scientists and between these scientists and the public, especially when that information highlighted the undesirable consequences of industrial development (Linnitt 2013).
- The Harper government had also eliminated non-partisan sources of scientific information that in the past had provided expert advice to the government regarding sustainable economic growth and other issues of science policy (Hoag 2012).

The concrete results of these actions were jarring. Thousands of government research scientists were put out of work, and many of Canada's top scientists left the country. Two hundred scientific research institutions and more than a dozen federal science libraries were closed due to cutbacks in funding. Scientific books and journals were literally thrown in dumpsters, invaluable data archives dating back a century were destroyed, and reams of publicly funded data and reports from government websites were deleted. In consequence, Canada dropped out of the world's top ten research and development performers, and it was said that Canada's basic climate and environmental science, in particular, had been set back for decades (Kingston 2015, Munro 2015).

As far as the protesting scientists were concerned, then, the actions of the Harper Administration meant that present and future evidence that could be used to support a strong environmental and climate policy was simply being rubbed out by the Harper government, killed off. Hence the terminology that galvanized the protesters' movement: the "death of evidence." Indeed, according to the protesters, all these actions of the Harper Administration represented a political takeover of Canadian research, even a "war on Canadian science," impeding its ability to continue to make important applied as well as basic contributions to science. Not surprisingly, therefore, the ten-year reign of Prime Minister Stephen Harper and his Conservative Party ended with the Canadian election of October

2015, when the Liberal Party's Justin Trudeau was voted in as Canada's new Prime Minister.

Just a year later, however – in November 2016 – the Republican Party's Donald Trump was voted in as President of the United States, and the kinds of actions that kill off scientific facts, such as major funding cuts for specific kinds of research and restrictions on communications from government science agencies, began yet again, though now in a different country. The protest *this* time, the so-called "March for Science" held on Earth Day in April 2017, was the largest science demonstration in history, taking place not only in Washington DC (where 100,000 people gathered) but also in more than 600 other cities all over the world (Smith-Spark and Hanna 2017, March for Science 2017). And many more protests were planned to take place (Kaplan 2017).

At least one more group of scientists should be added to the two groups already mentioned, the two groups engaged with what I have called the new worries about science. This third group numbers among its members nearly all of today's scientists, and their concern with fact-shaping in science relates to science's internal workings rather than the cultural biases that find their way into science or the governmental interference that sometimes constrains science. More particularly, their concern relates to science's current reward structures and the effect these reward structures have on scientific replication. Replication, of course, is the successful reproducing of experimental results. Called the cornerstone of scientific method, it is an absolute requirement for the proper grounding of science. Yet, in recent years even attempts at replication in science have been relatively rare.

The reasons are many. For one thing, replication studies are not normally viewed as major contributions to their fields; hence they receive less funding and less attention from both scientists and the media. What's more, they are harder to publish since journals prefer original research to replications of previous research. And they take time and resources away from other projects that reflect scientists' own original research ideas. So there has been little incentive to attempt replications. And when they *are* attempted, and especially when the results are negative, there is little incentive to even try to publish them since journals have a strong disinclination to publish any kind of negative or failed experiments. Moreover, in some fields, such as biomedicine, just gathering the materials for a replication experiment can be daunting, since many such experiments require working out special agreements before the research group that did the original research can share the necessary items (such as cell lines or specially created laboratory animals or bits of DNA) with the group attempting the replication. And that's when these materials are still available and still in a useable form. Finally, having to track down the exact procedures used in the original experiment, especially if it was done years earlier with the assistance of graduate students or postdocs and described in notebooks no longer available, can be a deal-breaker for scientists in any field, especially given publish-or-perish pressures (Price 2011, *Economist* 2013, Sheldrake 2015, Engber 2016, Hastings 2017).

The upshot: replication has been judged to be crucially important to science, but at the same time it has been treated as insignificant within the reward structures of science – an alarming situation, to say the least. As a result, serious efforts are now underway to motivate the doing and publishing of replication studies regardless of their outcomes. Funds have been allocated, a few large-scale replication studies have followed, and the results have been depressing. In every case, a surprisingly *low* percentage of the studies previously thought to be replicable (including studies done by the best scientists using the best methods and published in the best journals) *were* replicated. For example, in 2012 it was reported in *Nature* that scientists at the biotechnology company Amgen had attempted to replicate 53 "landmark" cancer studies, but only 6 of the 53 attempts were successful (11%) (Begley and Ellis 2012). In 2015 it was reported in *Science* that a collaboration of 270 researchers from all over the world had attempted to replicate 100 psychological studies that had been published in three top-tier psychology journals in 2008, but only 39 of those attempts were successful (Open Science Collaboration 2015). In 2016 a survey reported in *Nature* of 1,576 scientists from a variety of fields – chemistry, biology, physics and engineering, medicine, and earth and environmental science – found that more than 70% of those scientists had tried and failed to reproduce at least one other scientist's experiment and more than 50% had even failed to reproduce one of their own experiments (Baker 2016). In 2018 it was reported in *Nature Human Behaviour* that an attempt to replicate 21 social science experiments published between 2010 and 2015 in *Science* and *Nature* yielded only 13 successes, though even in the successful 13 the observed effect was on average only about 75% as large as in the original experiments (Nosek et al. 2018). And the list goes on.

All this has precipitated a "replication crisis" across science, but especially in psychology and biomedical research. As scientists see it, they and their colleagues, under pressure to pursue, at an uncomfortably rapid pace, ever new and different – read "novel" and "original" – investigations rather than the more lackluster replication investigations science requires, end up shaping science's inventory of facts in intolerable ways. Indeed, they end up introducing into science legions of interesting new "facts," many of which, it now appears, are not facts at all.

A Comparison of the Old and New Worries about Science

Philosophers of science have been nearly as unimpressed by all these ("new") worries of scientists as scientists had been by philosophers' ("old") worries. True, there are exceptions. Feminist philosophers of science, for example, have been as dismayed as feminist scientists by the ways women and females in general have been ignored or in other ways short-changed by science; a few philosophers of science living in Canada at the time of the Death of Evidence marches have written papers on the situation (see the papers by philosophers of science Stathis Psillos,

Maya Goldenberg, and Ingo Brigandt, as well as legal theorist Helena Likwornik in the 2015 *Canadian Journal of Philosophy* Symposium "Science, Values and the 'Death of Evidence' in Canada" and also philosopher of science Heather Douglas's 2015 paper in the *Bulletin of the Atomic Scientists*); and there is now serious interest among some philosophers of science about the replication crisis, as shown by two sessions and a lunch gathering devoted to it at the November 2018 meeting of the Philosophy of Science Association (see also Bird 2018 and Guttinger 2018). Still, the groundswell of attention that philosophers of science had devoted to the old worries is nowhere to be seen. Why is that?

At least one reason may be that the fact-shaping at the heart of the new worries has appeared to philosophers of science quite pedestrian. After all, the old worries concerned science's shaping of what it presents as the facts via the theories scientists accept and thus via the fact-stating languages scientists employ. So, the old worries had to do in a very real sense with fact *construction*. And this construction ran rather deep, as can be seen, for example, by the circumstance that the facts expressed in the language of any theory preclude acknowledgment of the facts expressed in the languages of all its competitors. By contrast, the new worries concern science's shaping of the facts via the research projects scientists pursue. So, the new worries have merely to do with fact *discovery* or *selection* from the array that are available rather than fact construction. What's more, such discovery or selection occurs via quite humdrum mechanisms: the areas of research the scientific community or its funders consider acceptable and important; the particular questions scientists pursue in those areas, their methods and tools of observation and analysis; the publishing choices of journal editors and book publishers; and so on. Hence, none of these discoveries of fact appears to preclude any other discoveries of fact at either the same or other times. In short, the new worries appear to concern a far more trivial shaping of the facts of science than the old worries.

Add to this the sheer novelty of the old worries and the shopworn nature of the new. The talk of theory-laden facts, theory-laden observations, the nonexistence of a theory-neutral language, and the other ingredients of the old worries constituted in their day fascinating new contributions to the philosophical scene, far more provocative and exciting and even, in some ways, more insightful and historically informed than what went before; and these contributions came with supportive backing from promising new ideas in the psychology of perception and the philosophy of language. By contrast, what I have called the new worries about science are only new relative to these old worries. That is to say, they simply came after the old worries. In other respects, these so-called new worries are actually quite old. I already pointed out, for example, how twentieth century women scientists echoed the critiques of their nineteenth century would-be-scientist sisters regarding the privileging of men in the facts of science. And I could have added scientists' responses to the political suppressions of science that antedated the recent Death of Evidence and March for Science movements – the protests directed at the treatment of climate science and other areas of science by the George W. Bush

Administration in the United States, for example, or, most famously, the protests directed at the treatment of genetics under Stalin or physics under Hitler as well as the protests through the centuries directed at the treatment of Galileo by the Catholic Church. Only the replication crisis might actually be new.

Such considerations as these must be the reasons the new worries about science have not enjoyed more attention from philosophers of science – at least have not enjoyed the attention that the old worries had. But are these considerations ultimately persuasive, and hence are philosophers of science in the right to continue largely ignoring the new worries of scientists? When we reflect again on the cases that currently motivate the new worries of scientists, it appears that the fact-shaping at their heart is not quite so pedestrian as the above suggests. To begin with, the discovery of facts featured in these cases *does* preclude the discovery of other facts, just as the acknowledgment of facts in the old worries precluded the acknowledgment of other facts. For example, the hundreds of years of fact gathering regarding men and males in general *did* preclude all sorts of other research projects that would have made sense if only various kinds of facts about females had been gathered, and the years and sometimes decades of cancer and other sorts of biomedical research that assumed and built upon non-replicated, non-replicable results did preclude, because additional funding or additional researchers were not available, other more promising lines of research yielding more useful information.

Moreover, the mechanisms by which all this fact gathering and preclusion of fact gathering occurred were far from humdrum. Of course they included choices of research areas and methods and publication venues and the like, but what underlay all of this was the commitment to certain values over others – for example, the commitment to androcentric values over egalitarian values or pro-industry values over environmental values. The new worries, in short, concern science's shaping of the facts via the *values* scientists (or their funders) accept just as the old worries concerned science's shaping of the facts via the *theories* scientists accept. Might these new worries, nonetheless, still fail to be as fascinating and provocative now in their day as the old worries were in theirs? But what if they do? The old worries, though fascinating and provocative, still failed to apply to actual science, given that scientists were continually doing what the old worries claimed they could not do – for example, compare alternative theories against a set of facts that did not presuppose any of the theories. And this is a crucial failure for a contribution to philosophy of science, a field whose central aim is to be relevant and helpful to science. By contrast, the new worries exactly express what goes on in science as attested to by the scientists themselves.

A Role for Philosophers of Science

Should the new worries about science receive serious attention from philosophers of science – at least as serious a level of attention as the old worries received? I have

suggested that there are no viable arguments against it. But there are also strong arguments for it. After all, as was explained at the outset, science's grounding in facts is the absolutely crucial, the absolutely indispensable, ingredient of science's success, the ultimate source of science's validity. But the new worries of scientists as well as the old worries of philosophers call into question, though in different ways, just this grounding. It makes eminent sense, then, for both scientists and philosophers of science to carefully assess the doubts that have been raised by both camps and to deal with them constructively. It makes especially good sense for philosophers of science to do this since the new worries of scientists raise issues that scientists have not even attempted to resolve, issues that are distinctly philosophical. Thus:

1. The foregoing has compared the old worries of philosophers of science with the new worries of scientists, and one thing should have been clear from the start. The scope of the old worries of philosophers was very broad: talk of theory-laden facts, theory-laden observations, the non-existence of a theory-neutral language, and the rest was intended to apply to all of science. But what is the scope of the new worries of scientists? Unlike the old worries, the new worries were illustrated by three specific cases, though it was also suggested, or at least taken for granted, that other cases could be provided, such as racial analogues of the first case and the mentioned Hitler and Stalin analogues of the second. It was clear, however, that scientists did not intend these cases to be representative of all of science. On the contrary, the suggestion was that these three cases were considered by scientists to be aberrations of what science is in general, or what science is at its best, or at least what science is supposed to be. But is there a very general problem here, comparable to the old worries of philosophers?
2. How should we respond to the three cases (and their analogues) illustrating the new worries about science? Should we, for example, adopt some kind of epistemic affirmative action program, privileging from now on, or for some specified period of time, the variety of fact gathering that was shortchanged in the past? And if so, what kind of epistemic affirmative action program should this be, given that different affirmative action programs have been elaborated on by political philosophers? Or would any such epistemic affirmative action program fail to provide a meaningful rectification for the past?
3. What can we do to prevent the kinds of cases that now illustrate the new worries about science? Can we do anything at all, given that all research inevitably shortchanges some aspect of its subject matter? For example, if we take precautions so that someone like former Canadian Prime Minister Stephen Harper is unable to "put to death" some of Canada's basic research regarding the environment, climate, and energy – so that the wellbeing of Canadians is protected – might we thereby put to death some of Canada's applied research regarding its economic development – which will undercut the wellbeing of Canadians? And how can we resolve such issues?

These are only three of the interesting new issues that we philosophers of science can fruitfully explore if we pursue the new worries about science.

Note

* An earlier version of this paper was presented at the Universidad de los Andes. I would like to thank the audience there, and especially Manuela Fernández Pinto and Santiago Amaya, for very helpful questions and suggestions and a wonderfully lively discussion.

References

Baker, Monya. 2016. "1,500 Scientists Lift the Lid on Reproducibility: Survey Sheds Light on the 'Crisis' Rocking Research." *Nature* 533(May 26): 452–454.

Begley, C. Glenn, and Lee M. Ellis. 2012. "Drug Development: Raise Standards for Preclinical Cancer Research." *Nature* 483(7391): 531–533.

Bird, Alexander. 2018. "Understanding the Replication Crisis as a Base Rate Fallacy." British Journal for the Philosophy of Science (August 13), axy051, https://doi.org/10.1093/bjps/axy051

Brigandt, Ingo. 2015. "Social Values Influence the Adequacy Conditions of Scientific Theories: Beyond Inductive Risk." *Canadian Journal of Philosophy* 45(3): 326–356.

Chung, Emily. 2014. "Foreign Scientists Call on Stephen Harper to Restore Science Funding, Freedom." *CBC News* (October 20). At www.cbc.ca/news/technology/foreign-scientists-call-on-stephen-harper-to-restore-science-funding-freedom-1.2806571. Retrieved 22 April 2018.

Douglas, Heather. 2015. "Reshaping Science: The Trouble with the Corporate Model in Canadian Government." *Bulletin of the Atomic Scientists* 71(2): 88–97.

Economist. 2013. "How Science Goes Wrong." *The Economist* (October 19). *Expanded Academic ASAP*, At http://link.galegroup.com/apps/doc/A345848248/EAIM?u=nd_ref&sid=EAIM&xid=5b5cb667. Accessed 24 December 2018.

Engber, Daniel. 2016. "Cancer Research Is Broken: There's a Replication Crisis in Biomedicine — and No One Even Knows How Deep It Runs." Slate (Future Tense) (April 19). At www.slate.com/articles/health_and_science/future_tense/2016/04/biomedicine_facing_a_worse_replication_crisis_than_the_one_plaguing_psychology.html

Ferber, Marianne A., and Julie A. Nelson, eds. 1993. *Beyond Economic Man: Feminist Theory and Economics*. Chicago: University of Chicago Press.

Feyerabend, Paul. 1965. "Problems of Empiricism." In Robert G. Colodny, ed., *Beyond the Edge of Certainty: Essays in Contemporary Science and Philosophy*. Englewood Cliffs, NJ: Prentice Hall.

Gamble, Eliza Burt. 1894. *The Evolution of Woman, an Inquiry into the Dogma of Her Inferiority to Man*. London: G.P. Putnam's Sons.

Gero, Joan and Margaret Conkey. 1991. *Engendering Archaeology: Women and Prehistory*. Oxford: Blackwell.

Goldenberg, Maya. 2015. "Whose Social Values? Evaluating Canada's 'Death of Evidence' Controversy." *Canadian Journal of Philosophy* 45(3): 404–424.

Guttinger, Stephan. 2018. "A New Account of Replication in the Experimental Life Sciences." *PhilSci-Archive.* At philsci-archive.pitt.edu/14410/1/Guttinger_Microreplications_2018.pdf

Hanson, Norwood Russell. 1958. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.

Hastings, Conn. 2017. "Are Replication Studies Unwelcome?" *Frontiers* (May 1). At https://blog.frontiersin.org/2017/05/01/are-replication-studies-unwelcome/

Hoag, Hannah. 2011. "Canadian Research Shift Makes Waves." *Nature* 472: 269 (April 19). At www.nature.com/news/2011/110419/full/472269a.html. Retrieved 22 December 2018.

Hoag, Hannah. 2012. "Canadian Budget Hits Basic Science." Nature (March 30). At www.nature.com/news/canadian-budget-hits-basic-science-1.10366. Retrieved 22 December 2018.

Hubbard, Ruth. 1979. "Have Only Men Evolved?" In Ruth Hubbard, Mary Sue Henifin, Barbara Fried, eds., *Women Look at Biology Looking at Women: A Collection of Feminist Critiques*. Cambridge, MA: Schenkman Publishing.

Kaplan, Sarah. 2017. "Six Months Later, the March for Science Tries to Build a Lasting Movement." The Washington Post (October 23). At www.washingtonpost.com/news/speaking-of-science/wp/2017/10/23/six-months-later-the-march-for-science-tries-to-build-a-lasting-movement/?utm_term=.7f9d9fb22413. Retrieved 22 December 2018.

Kingston, Anne. 2015. "Vanishing Canada: Why We're All Losers in Ottawa's War on Data." Maclean's (September 18). At www.macleans.ca/news/canada/vanishing-canada-why-were-all-losers-in-ottawas-war-on-data/. Retrieved 22 December 2018.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Likwornik, Helena. 2015. "Who's Afraid of the Big Bad Wolf? The Interweaving of Values and Science." *Canadian Journal of Philosophy* 45(3): 382–403.

Linnitt, Carol. 2013. "Harper's Attack on Science: No Science, No Evidence, No Truth, No Democracy." *Academic Matters* (May 30). At https://academicmatters.ca/harpers-attack-on-science-no-science-no-evidence-no-truth-no-democracy/. Retrieved 22 December 2018.

March for Science. 2017. "The Science Behind the March for Science Crowd Estimates." Science Not Silence. At https://medium.com/marchforscience-blog/the-science-behind-the-march-for-science-crowd-estimates-f337adf2d665. Retrieved 22 December 2018.

Munro, Margaret. 2015. "Canadian Budget Pushes Applied Research." Nature 520(7549) (April 22). At www.nature.com/news/canadian-budget-pushes-applied-research-1.17305. Retrieved 22 December 2018.

Nature Editorial. 2012. "Death of Evidence: Changes to Canadian Science Raise Questions That the Government Must Answer." *Nature* 487: 271–272 (July 19). At www.nature.com/articles/487271b. Retrieved 22 December 2018.

Nosek, Brian A. et al. 2018. "Evaluating the Replicability of Social Science Experiments in *Nature* and *Science* Between 2010 and 2015." *Nature Human Behaviour* 2 (August 27): 637–644.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349(6251) (August 28).

Pedwell, Terry. 2012. "Scientists Take Aim at Harper Cuts with 'Death of Evidence' Protest on Parliament Hill." The Globe and Mail (July 10). At www.theglobeandmail.com/news/politics/scientists-take-aim-at-harper-cuts-with-death-of-evidence-protest-on-parliament-hill/article4403233/ Retrieved 22 December 2018.

Popper, Karl. [1935] 1959. *The Logic of Scientific Discovery*. London: Hutchinson. Originally published as *Logik der Forschung* (Vienna:Verlag von Julius Springer, 1935).

Price, Michael. 2011. "To Replicate or Not to Replicate?" Science (December 2). At www.sciencemag.org/careers/2011/12/replicate-or-not-replicate

Psillos, Stathis. 2015. "Evidence: Wanted, Alive or Dead." *Canadian Journal of Philosophy* 45(3): 357–381.

Rosser, Sue. 1994. *Women's Health-Missing from U.S. Medicine*. Bloomington: Indiana University Press.

Sheldrake, Rupert. 2015. "The Replicability Crisis in Science." Nature (September 1). At www.sheldrake.org/essays/the-replicability-crisis-in-science

Smith, Teresa. 2012. "Scientists Stage Mock Funeral to Protest Cuts to Research." Canada.com (July 11). At www.canada.com/business/Scientists+stage+mock+funeral+protest+cuts+research/6913396/story.html. Retrieved 22 December 2018.

Smith-Spark, Laura, and Jason Hanna. 2017. "March for Science: Protesters Gather Worldwide to Support 'Evidence'." *CNN*. At www.cnn.com/2017/04/22/health/global-march-for-science/index.html. Retrieved 22 December 2018.

15

IS SCIENCE REALLY VALUE FREE AND OBJECTIVE?

From Objectivity to Scientific Integrity

Matthew J. Brown

Objectivity and the Value-Free Ideal

Particle physicists must decide how much evidence to collect before announcing the discovery of a new particle like the Higgs Boson, balancing reasonable caution about premature or erroneous discovery claims against the value of a successful discovery claim (Staley 2017). Regulatory scientists assessing the potential toxicity of a chemical must determine thresholds of evidence in ways that balance the risk of falsely certifying a chemical as safe (thus increasing health risks) against the risk of falsely attributing toxicity (thus encouraging unnecessary regulation) (Douglas 2000, 2009). Social scientists must determine how to define value-laden terms like "rape" or "violence" (Dupré 2007) or "well-being" (Alexandrova 2017). Such decisions are at the heart of scientific inquiry, and yet they each require carefully weighing values. Is it possible for scientific knowledge to be objective, if scientists must make value judgments in the course of scientific inquiry?

Many have held that scientific objectivity requires that the parts of scientific inquiry concerned with weighing evidence and making empirical claims be value free. Of course, values ought to guide us in protecting human research subjects, and they might motivate scientists to work on certain problems over others, or even inspire scientists to suggest certain hypotheses or theories. But further into the core of scientific inquiry, where data is collected, evidence analyzed, hypotheses evaluated, and empirical claims judged and asserted, values can only lead science into bias, subjectivity, wishful thinking, and politicization.

On such a view, values are understood as intrinsically subjective and biasing factors. What sort of world we wish to live in is relevant to determining how we should treat each other, or what kinds of things it would be useful or interesting to know about, but is irrelevant to how the world really is. Anything more would

be an unacceptable sort of wishful thinking, claiming that something is the case because one wishes it were the case. For instance, feminist science has been accused of sliding from "women should be equal to men" as a political value to "women and men are equal" as a descriptive claim about, say, intelligence or capability (Haack 1993; Anderson 1995; Hicks and Elliott 2018). Call this "the problem of wishful thinking" (Brown 2013). Objectivity is taken to be the opposite of wishful thinking.

In this chapter I argue that this way of thinking is wrongheaded. Science is necessarily value-laden, and scientists must make value judgments in order to do science responsibly, with integrity. Whether value-laden science is objective is a vexed question, because there are so many different things we might mean by "objectivity." There are some accounts of objectivity which are compatible with value-ladenness, or by which we might even accord value judgments a kind of objectivity. Values are not inherently biasing in the way the views mentioned presuppose. However, one legitimate sense of "objective" is simply being value free. Objectivity is, in any case, too vexed and problematic a notion to be of any use in guiding science or philosophy of science. We should abandon it in favor of providing an account of scientific integrity, which involves both epistemic and ethical responsibilities, and answers concerns about trustworthiness which capture the important concern behind calls for objectivity.

The Need for Values in Science

The value-ladenness of science is unavoidable. The closest scientists can get to doing work that is value free is to either ignore the consequences of their work, or to do work that has few consequences for things that we care about. Far from realizing a scientific ideal, both of these approaches amount to massive irresponsibility on the part of the scientists. The first approach amounts to a kind of serious recklessness or negligence (Douglas 2009). The second approach intentionally turns science into an abstruse private pursuit, shirking the significant responsibilities that scientists have to produce knowledge useful to society.

To see why values cannot be avoided in science, we must consider the role of contingency in science in concert with the significant social and ethical consequences of science.[1] From the point of view of the scientific inquirer in the midst of inquiry, there are a number of contingent moments, places where reasonable inquirers could proceed in different ways. They must decide what to investigate and how to investigate it. They must choose concepts to use, hypotheses to pose, techniques for characterizing data. They must decide how much evidence would be sufficient to accept and publish their results.

I describe these contingencies as decisions, but this is something of an idealization. In fact, these contingencies might be settled by habit, custom, or convention. Only one option may occur to an inquirer, in which case it may not seem that there is a decision to be made. These are contingent moments in a normative and

counterfactual sense – other inquirers faced with the same decision *could reasonably* go in a different direction. While they are not necessarily actual decisions, they are *decision-points*.

The pervasiveness of contingency can be seen in the significant role that controversy plays in scientific progress. Science studies have made much of controversy in science (Collins 1981; Latour 1987; Pinch 2015). Science studies scholars have sometimes gone too far in the conclusion that they draw from such controversies, including arguments against the rationality of science. A more modest conclusion is that science is difficult, and there is rarely one obviously right choice in significant moments of scientific inquiry. The main lesson I take from the contingency of science is that inquirers are decision makers, that is, that they have options for how to proceed. This lesson must be considered in light of the significant social consequences of science in order to see the need for values in science.

That science has significant social consequences should not be particularly controversial. Scientific knowledge affects education, policymaking, court cases, individual decisions about things like diet and health care, as well as our conception of ourselves and of our place in the universe. Science can reinforce or undermine the most contemptible social stereotypes and prejudices as well as the highest human ideals. The decisions made in the course of scientific inquiry are thus actions with social, ethical, or political implications and consequences for what we value.

One might hope to deny these consequences by drawing some distinction, such as the distinction between scientific research and expert advising, or between science and technology. All such distinctions fail to reflect the reality of science as a social institution. First, the consequential sides of these dichotomies (advising, technology) are shaped by decisions on the "pure" side. Second, scientific research itself has a direct impact – scientific results are published where anyone can read them if they have the right access, through libraries or purchasing of articles or journal subscriptions. The results are frequently reported on in the popular press, blogs, and social media, making them even more widely available.[2] Third, the advisor or educator is often also the researcher, and these roles are blurred in their own lives. These distinctions do not and could not amount to practical divisions, and thus they cannot block the concern about consequences.

Everyone has the responsibility to consider the consequences of their actions. This is not a special responsibility that scientists have in their role as scientists, but one of their general responsibilities as moral agents. What's more, there are no special role responsibilities that scientists have that could screen them from this general responsibility. Science does not have professional exceptions to general responsibilities in the way that lawyers (attorney-client privilege) or doctors (patient confidentiality) do. Nor would we want them to (Douglas 2009, pp. 71–79).

Call the argument I have laid out here "the contingency argument." It can be summarized in this way:

1. Scientific inquiry has many contingent moments.
2. Each contingent moment is a decision point, a potential decision among multiple options.[3]
3. These decisions often have ethical and social consequences, or consequences for values generally.
4. Value judgments should settle choices that affect values.
5. Thus, scientists should make value judgments in settling scientific contingencies.

Identifying contingencies, alternative options, stakeholders, and values takes a significant amount of sensitivity and moral imagination. It may also require research, consultation, and epistemic humility.

Consider some of the examples mentioned at the opening of this chapter. The physicists looking for the Higgs boson had to decide when their evidence merited announcing the discovery of the particle, and they used a standard of 5-sigma. "5-sigma" means that the data taken to indicate the existence of the Higgs boon is five standard deviations above the mean of a normal distribution of data given the null hypothesis, that is, assuming that the Higgs boson does not really exist. In standard null-hypothesis statistical testing terms, this amounts to a p-value of 3×10^{-7} or 1 in 3.5 million (Lamb 2012). The p-value is the probability that, given some statistical assumptions, if the null hypothesis were true, we might observe data at least as extreme as the data in fact observed. A low p-value means that probability is low, which gives us some very conditional reasons to think that the null hypothesis should rejected. Physicists could have used a less extreme standard, such as 3- or 4-sigma, which would have increased their likelihood of mistakenly announcing a discovery, but would have also decreased the time and expense required before announcing the claim. 3-sigma is a rather high standard of evidence from the point of view of many fields of research (close to a p-value of 0.001 or 0.1%). On the other hand, the scientists could have raised the bar to 6- or 7-sigma, incurring much greater expense, keeping the relevant scientific communities waiting longer for this much-anticipated knowledge, and even decreasing the chance that a discovery would ever be announced. On the other hand, this standard would also decrease the chance that a false discovery claim would be made.

Value-laden scientific concepts present another example of the contingency argument at work. For instance, John Dupré (2007, pp. 28–30) briefly described social science work on *violence*. That "violence" has a (negative) evaluative connotation is obvious. But violence is also the sort of thing that sociologists may wish to construct a measure of, perhaps combining statistics on things like murder rates, frequency of crimes involving assault or deadly weapons, reports of domestic violence, and so on. A claim like "The United States is a violent country" or "Sam is a violent child" might reflect both an evaluation and a report of a measurement. It is not that these claims are ambiguous between a descriptive and an evaluative claim that should be clearly disambiguated. Rather, the connection between the

evaluative and descriptive elements of the concept are what connect scientific work to our goals and reasons for action, and it permits us to adequately evaluate competing ways of operationalizing the concept (Dupré 2007, pp. 30–31). In a similar vein, Anna Alexandrova (2017) considered scientific claims about *well-being*, and argued that such claims are "mixed claims" (descriptive and evaluative). Mixed claims should be retained, and not disambiguated, because the normative element of a concept like "well-being" is crucial to the normative decisions that must be made throughout the scientific process (Alexandrova 2017, p. 91).

As may already be apparent from the examples, "contingency" doesn't mean that anything goes. There are genuine contingencies where there is room for reasonable disagreement among experts about the options at hand. This is a normative matter; whether the experts agree or disagree can be a reason to think that the matter is a contingent, but cannot decide the case – for example, there might be closed-mindedness about certain options, or certain unconceived alternatives, generating hasty consensus. Values should not, for example, replace evidence wholesale. Values should guide the decision between reasonable interpretations of that evidence, or should help evaluate the reliability and relevance of evidence. Values should not short-circuit inquiry, and they have no role to play where there is no alternative courses of action open to inquirers.

More generally, the role of values in scientific inquiry is to guide decision-making about genuine contingencies. What counts as a *genuine* contingency is determined by what is reasonable given the state of scientific practice at a time, the available and relevant evidence, the track record of theoretical explanations and experimental techniques, the course of the specific inquiry up to this point, and so on. The guiding role for values takes two major forms: first, values determine and promote the *aims* of the particular inquiry (Elliott 2013; Hicks 2014; Intemann 2015). Some inquiries may have more epistemic or cognitive aims, like providing a simple, comprehensive explanation of a body of phenomena that generates novel predictions.[4] Biomedical inquiries aim at health, while environmental risk assessments might aim at both human safety and ecosystem integrity. Second, values may act as side constraints, even when they might tend to frustrate inquiry. Protections of the rights and welfare of human research subjects must always be a constraint on inquiry. Avoiding other kinds of social harms, as might be caused by assertions of racial or gender difference in abilities, might likewise serve as a general constraint on inquiry, limited by the genuine contingencies of the situation.

Objectivity without Epistemic Purity

The first point nearly every philosopher of science makes about the concept of "objectivity" is that it is complex, ill-defined, difficult to characterize, "essentially contested" (Harding 1995), and attributed to a great variety of different things – individuals, groups, knowledge claims, methods, processes, practices, observations, measurements, and so on. Objectivity is taken to mean true or real, based in

(observed) facts, done according to specified rules or criteria, unbiased, impartial, or value free, or a view from nowhere, independent of human perspectives. Those who hold that science is and ought to be value-laden have generally argued that out of this mess, a perfectly good sense of "objectivity" can be found that still applies. Value-laden science can still be objective; objectivity does not require epistemic purity.

A useful framework for understanding these appeals to objectivity can be found in the work of Heather Douglas (2004, 2009). First, Douglas divided up different sorts of *processes* whose objectivity are at issue. The *products* of science (knowledge claims) are objective insofar as they are produced by objective processes (Douglas 2004, 454, 2009, pp. 116–117). This makes sense, as it is not possible for us to read objectivity off of a knowledge claim directly. (But note, this already rules out the equation of "objective" with true or real.) Second, within each type of process, there are different senses in which that process could be objective.

The three types of process that Douglas distinguished are (1) human interactions with the world, as in experimental and observational processes, (2) individual reasoning and thought processes, and (3) social processes, such as peer review, criticism, and consensus-formation. The first operationalizes the idea that objectivity has to do with capturing or being guided by "facts," and includes experimental manipulation and robustness or concordance of different types of experiments. The second concerns individuals being unbiased and impartial. The third concerns whether a community of experts and its processes and structures are objective, and has been the type of objectivity feminist philosophers of science have often hoped to (re)claim.

The second type of objectivity has been the most problematic for the critics of the ideal of value-free science. At a first pass, for an individual's reasoning to be objective just seems to mean for it to be unbiased, neutral, impartial, or value-free. This seems to put us between the rock of shirking the responsibilities entailed by the contingency argument and the hard place of failing to be objective. Rather than give up on this form of objectivity, and focus solely on the other two, Douglas attempted to distinguish between difference senses of individual objectivity.

One sense of individual objectivity is what Douglas called "detached objectivity." Here, the prohibition is on taking values or preferences as a *reason* to make a knowledge claim in a way that is resistant to or in conflict with the evidence. When someone denies that climate change exists because they value a lack of regulations, they fail to be detached. Similarly, the inventor of a theory who continues to defend it in the face of mounting evidence to the contrary may have our sympathy, but we would not call them objective in this matter.

Detached objectivity is not the same as value-free objectivity. The latter forbids *any* role for values in science. The equation of *objective* and *value-free* depends on the idea that all values are biasing or subjective, and by playing a role in science, as Douglas put it, they "contaminate it" (2004, p. 459). Such arguments are doubly

mistaken. First, it is not true that values are themselves necessarily biasing or subjective. Second, the "contamination" claim is groundless, due to a simplistic, structureless notion of inquiry.

Values are not necessarily subjective in any meaningful sense, and they need not have a biasing effect on science. The position that values are wholly subjective is both a controversial opinion within ethics[5] and difficult to square with ordinary moral practice. We tend to treat disputes about some values as substantive disagreements rather than differences in taste. Furthermore, we readily distinguish between values stated unreflectively or habitually, and those that are the product of careful value judgment. The claim that values are necessarily biasing is similarly problematic. A common narrative about the influence of feminist values in late twentieth century science is that they tended to remove, rather than create, misleading biases in science (Harding, 1995). Inclusivity, fairness, and respect for marginalized persons are values that might decrease rather than increase bias.

The idea that values inherently "contaminate" inquiry is likewise a highly problematic view. We can see this from two directions. First, there are several uncontroversial restrictions on inquiry by ethical values, for example, protections for human subjects. The influence of values may slow or halt certain lines of inquiry when those lines of inquiry would require unethical treatment of human subjects. The results of inquiry that is undertaken instead are not therefore "contaminated" by the value of respect for persons or concern for human welfare. Second, where values are guiding decisions about *genuine* contingencies in the sense discussed above, there is no sense in which leaving value judgments out of those decisions makes them more reasonable or more "objective." If anything, the failure to consider relevant factors to the decision makes the process not only reckless but also irrational. So, value freedom is not a type of individual objectivity worth having.[6]

As mentioned earlier, feminist philosophers of science and others denying the value-free ideal have tended to focus on the social mode of objectivity. The most prominent such account is Helen Longino's critical contextual empiricism, according to which, "A method of inquiry is objective to the degree that it permits *transformative* criticism" (Longino 1990, p. 76). Douglas called this "interactive objectivity" (Douglas 2004, p. 463). That is, objectivity requires that the inquiry be subjected to critical discourse by the relevant scientific community that follows certain norms, including uptake of criticism and equality of intellectual authority among qualified practitioners. "Method of inquiry" refers neither to individual reasoning processes, nor to procedures followed in the laboratory, but rather to social processes of discourse, assessment, and criticism. According to Longino, a scientific community that follows her four norms for critical discourse, with sufficient diversity within the community to ensure that important assumptions are not so universally shared as to be free from scrutiny, will be objective.

Another influential social account of objectivity, not mentioned in Douglas's typology, is Sandra Harding's theory of "strong objectivity." Harding focused on diversity, taking it not in a liberal pluralist direction, as Longino did, but rather

in the direction of feminist standpoint epistemology. According to Harding's program, inquiry should begin from the position of socially marginalized people (e.g., women) in order to uncover biases and expose them to scrutiny. Because the values, interests, and assumptions of the dominant members of a hierarchically structured society tend to become naturalized and implicit, starting from the position of the marginalized tends to increase, rather than reduce, scrutiny of bias. This approach, which she labeled "strong objectivity," strengthens objectivity more so than any form of impartiality or attempt to transcend perspectives or subject-positions.

The various notions canvassed in this section have good call to be regarded as virtues of scientific inquiries, inquirers, and communities. What may seem questionable is whether they really capture what is meant by "scientific objectivity," which seems to many essentially linked to inquiry that is value-free. What's more, there is cause to ask whether in all this diversity of norms and criteria there is sufficient unity justifying the use of the single term, "objectivity." The next section considers arguments against retaining the focus on objectivity in science.

Against Objectivity

One might be tempted to think that "objectivity" is a merely honorific term, an "empty compliment" paid to good ideas or procedures.[7] Another way to put it is that "objectivity" serves the rhetorical purpose of lumping together a variety of virtues for scientific theories, ideas, methods, or techniques. "Objective" here is just a highfalutin way of saying that something is epistemically good. The things called "objective" are good in very different ways: they are empirically grounded, reliable, trustworthy, detached, open-minded, rigorous, or critically engaged. These specific terms better capture the relevant scientific or epistemic virtues than the general lumping term "objectivity." In the context of the arguments for the ideal of value-free science that depend on the unprincipled lumping of value-freedom with these other virtues, the usage becomes positively vicious.

Ian Hacking (2015) has a related set of concerns. According to Hacking, there are two main concerns about talk of "objectivity." First, it is an abstraction from a variety of "ground-level" concerns that have little if anything to do with each other. Trying to figure out what objectivity is, or providing a theory of objectivity, distracts from these ground-level concerns (Hacking 2015, p. 20). Second, to call something "objective" is to say that it lacks one or more epistemic vices, rather than to attribute some epistemic virtue to it (Hacking 2015, pp. 24–26). So objectivity doubly lacks content: it is abstracted from the details that really matter, and it has no positive content of its own.

Jack Wright (2018) has responded to Hacking in two ways. First, he argued that despite being an abstraction, the concept of "objectivity" can nonetheless help address "ground-level" concerns. It does so because it is a "relational category," i.e., because it serves to relate diverse practices to one another and to various goals

and ideals. In bringing them into relation, practitioners can compare, assess, refine, and justify practices in ways that help deal with difficult questions. This proposal bears significant resemblance to a point made by Douglas: "Even with eight senses, objectivity is conceptually coherent … there are conceptual links across the senses, but no one sense fully captures the meaning of objectivity" (Douglas 2004, p. 467). The ways in which the different senses of objectivity connect and "evoke each other" (p. 468) is one of the more suggestive features of Douglas's account.

These moves do not seem to me, however, to save the concept of objectivity from the charge of incoherence. If one sees "objectivity" as a word covering for a broad collection of virtues (or absence of vices), then even if there is no coherent core to the collection, it would not be surprising to find relations between them. We cannot reduce honesty to kindness or vice versa, but it is not much of a surprise to find the two traits often going together. This does not mean they are two different aspects or species of the same abstract virtue. Likewise, that following impersonal rules and being detached may often go together, or convergence of multiple lines of evidence might frequently go along with increasing consensus on some conclusion does not require that these all be instances of some abstract category of objectivity.

Pluralism is not really a solution here, either. Wright attempted to compare his defense of the concept of objectivity to Ingo Brigandt's (2003) defense of the species concept in the face of calls for species eliminativism. Brigandt rightfully pointed out that "species" occupies a place in general theoretical accounts, and that each version of the species concept adequately fills that role. "Objectivity" is different, however. There seems to be no such unified account, no such functional role for the different concepts of objectivity to play.[8] All that the different forms of objectivity have in common is that they are good things for some element of science to have (or the lack of something it is bad for them to have).

Wright went beyond a pluralist approach and attributed a core concept that provides unity to the category of objectivity. The core idea Wright adopted is that objectivity involves a "stepping back" from some aspect of the context of inquiry or assertion, an idea Wright attributed to Thomas Nagel (Nagel 1986). This stepping back is goal-directed. Stated more precisely: "A knowledge claim is objective to the extent that it is produced in a way that steps back from features of the context in which it was produced relevant to meet a goal" (Wright 2018). Objectivity-ascriptions, then, involve a relation between two different contexts: the context of use, which sets the goal, and the context of production, from which the knowledge claim "steps back."

Consider the case of regulatory science mentioned above. One goal of such research is to protect the health of citizens and ecosystems. According to this goal, we might call regulatory research "objective" if it steps back from the interests of the companies that produce the relevant chemicals. Those are features of the context of inquiry that might influence the research in a way that that hampers the goal.

This account of objectivity has much in common with the so-called "aims approach" to values in science, according to which the use of values in science is legitimate insofar as that use contributes to the aims of the research (Elliott 2013; Hicks 2014; Intemann 2015; Steel 2017). The comparison raises two concerns about Wright's account, however. First, the aims approach typically does not worry about "objectivity," and focuses instead on what contributes to or detracts from the aim of the research. The addition of the word "objectivity" does not seem to add much to those accounts. Second, a concern has been raised that the aims approach focuses only on issues of instrumental rationality but gives us no tools to evaluate the aims of inquiry. If the goal of regulatory science is reconceived by the chemical companies as freedom from burdensome regulation, an inquiry that "steps back" from concerns about health and safety may be regarded as legitimate (by the aims approach) or objective (by Wright). More generally, it is not clear that what Wright identified as objective is generally a good thing. Sometimes stepping back from features of a context that help one meet a goal is still undesirable, as when it causes us to lose track of the harms done by the research.

Once we acknowledge that science is and must be value-laden, and we question the assumption that values are inherently subjective or biasing, it becomes difficult to pinpoint what the contrast class for "objectivity" is, such that objectivity is generally a good thing and the opposite is generally to be avoided. If that's so, this reinforces the idea that "objectivity" is an empty honorific paid to various ways of doing science regarded as good.

Here is what I mean in saying that objectivity has no meaningful contrast. Two candidate contrast terms come to mind: *subjectivity* and *bias*. What could be "subjective" in the context of scientific knowledge? Even if it makes sense to talk about certain perceptions or beliefs as subjective, the stock and trade of science is not belief but public knowledge claims. Even two different interpretations of the same data, supporting competing claims, are typically based on articulable and often articulated methodological, modeling, or theoretical assumptions. Scientific knowledge claims are found in published articles, in discourse, at conferences, in textbooks. As they are publicly accessible, they are publicly assessable. They might be poorly supported, or controversial, but those aren't the same as being subjective.

Two cases in which we might want to call knowledge claims "subjective" are, on the one hand, cases of mere opinion and, on the other hand, claims based on tacit knowledge. First, unsurprisingly, someone will occasionally try to pass off mere opinion as scientific knowledge. But such moves are easily spotted, even by non-experts, and more precisely called "ungrounded," "wrong," or "propaganda posing as science" than "subjective." The second case, claims based on tacit or implicit knowledge, are trickier. We might point to skills learned in the laboratory, or long experience in clinical practice, as examples of tacit knowledge relevant to scientific (or medical) knowledge claims. But note that claims are never based *entirely* on tacit knowledge – the laboratory scientist also provides evidence,

measurements, descriptions of methods. Your physician provides not only their judgment, but information about, for example, possible treatments, their success rates and side effects, based on published research. What's more, there are ways of publicly assessing tacit knowledge, even if they are indirect, for example, by examining credentials, by appealing to the reliability or success rate of the practitioner, through observation by another skilled expert, and so on. Tacit knowledge is thus not genuinely subjective, since it is publicly assessable.

Our second candidate for the opposite of "objective" is being biased; one fails to be objective if one is biased in favor of one "side" over the other. This sort of concern is at work when we worry about having an objective trial judge or dispute mediator. There are several problems with the notion of "bias" in science. First, in science, there are not often "sides" the same way as in a court case, where the goals of the parties are diametrically opposed. Scientists are primarily engaged in inquiry in order to solve problems about their subject of research. Sometimes they collaborate, and sometimes they engage in a bit of competition to see who can solve the problem first or best. Of course, they sometimes act as partisans for or against their favored theories or approaches, but even then, their competition takes place within a background of shared goals. More often, scientists work on different problems or aspects of problems.

Second, even in a court case, it is widely recognized that complete impartiality is not always appropriate. In criminal cases, the burden of proof is very different for the defendant and the prosecution. While we may want judge and jury to be unbiased in the sense of not having any preconceptions about the case, we do not want them to apply the same standards to both sides. A more general concern with the topic of "bias" is that it is often equated with value-ladenness; according to this common view, to be objective merely is to be value free, which is the view we're trying to avoid.

As we have seen, being value-free is not generally a virtue, and indeed, it can amount to being irresponsible. Of course, if one reaches conclusions entirely on the basis of values instead of doing inquiry (a failure of detachment, in Douglas's terms), one is doing something illegitimate. But the sin here is greater than "bias," it is to cloak propaganda in the vestments of science. When values are used to manage genuine contingencies, however, this can be a virtuous thing. As such, "objectivity" and "bias" in these senses are poor tools for guiding the interaction of values and science.

While most of the things called "objective" in the previous section are virtuous in one way or another, there seems little that is useful in lumping them together under one philosophically fraught term. What's more, it is not always the case that the absence of these virtues is necessarily vicious. Communities structured differently from Longino's ideal still seem able to produce some scientific knowledge. Tenacious defense of a favored hypothesis has a role to play in the scientific process, even though it is a failure of detachment. Sciences where manipulation or convergence are unachievable are not somehow defective.

The concept of "objectivity" seems not to get us what we want in a normative account of scientific knowledge. Nothing holds the different meanings of "objectivity together." The concept, such as it is, has no clear contrast class. And it continues to carry the normative baggage of the untenable value-free ideal. In the next section, I argue that what we need instead is a good account of scientific integrity.

From Objectivity to Scientific Integrity

We want to know which theories, which results, which cases of scientific consensus, which expert advice we can trust. Hacking (2015) referenced Theodore Porter (1995) and Naomi Scheman (2001), both of whom closely connected trust to objectivity; indeed, on Scheman's account, objectivity *is* trustworthiness. I see the move from objectivity as discussed previously to trust as a positive shift; but when can we trust an expert, a result, a theory?

According to Scheman, the need for trust arises from what she calls our "epistemic dependency," the fact that it is not possible in practice (perhaps not in principle) to assess every knowledge claim for ourselves (Scheman 2001, p. 30). We rely on the testimony of others and in particular on the judgments and claims of experts. Scheman sees trustworthiness as having two components – competence and integrity (Scheman 2001, p. 33). The competence of an expert, a method, or a study can be evaluated in familiar epistemic terms. Integrity, on the other hand, is a partly social and partly ethical notion. Given that science is value-laden, what we really want to know (beyond whether it is done competently) is whether it is done with integrity. This question captures what is valuable about objectivity.

What are some familiar moments when scientists act without integrity? One example is when scientists speak with authority well outside of the area of their expertise, as when scientists distant from the field of climate science challenge the expert consensus on anthropogenic climate change. Another example is when scientists present claims as more certain or less controversial than they really are. Other failures include close-mindedness, a failure to consider all aspects of a problem, failing to question problematic assumptions, or shutting down inquiry prematurely.

In positive terms, what does scientific integrity involve? I posit three core components: critical sensitivity, responsibility, and humility. Each of these components involves elements that are typically classified as *epistemic* and *social*, though those elements are not necessarily extricable from one another.[9]

Critical sensitivity is an awareness of the potential issues that arise in inquiry, a sensitivity to the contingencies that arise in the scientific process, and a recognition that value judgments must be made as part of settling those contingencies. Critical sensitivity involves being relatively less likely to rely on habit and convention, when doing so could have harmful consequences. It is a protection against negligence and recklessness in scientific inquiry. It can be cultivated by periodic

questioning of decisions in the scientific process.[10] Critical sensitivity sometimes requires creativity and imagination, in identifying or creating alternatives and empathizing with potential stakeholders in order to make the relevant value judgments.

The responsible scientist is careful, open-minded, methodical; they do not rush to judgment or make hasty assumptions. They are sensitive to both the epistemic and social consequences of their decisions, and they consider the relevant reasons and the interests of the relevant parties carefully. They make value judgments where needed, and they take care to make those value judgments well.

Scientific humility requires recognizing one's limitations as an inquirer. This requires knowing that the scope of one's expertise is relatively limited, and therefore limiting the way one presents oneself. Scientific humility means not presenting one's claims as more certain than they are, not making grandiose claims about what a limited or initial result means or what a research program can do. Scientists drawing deep philosophical claims about, for example, the nature of free will, the existence of god, or the nature of morality based on a limited collection of specific results, are typically overreaching. Finally, humility also requires recognizing our limitations as trustees of public interests or the welfare of stakeholders and taking steps to engage or consult with others to be more socially responsible.[11]

Conclusion

Science is necessarily value-laden, as a result of the endemic contingencies of science coupled with its significant social consequences. The attempt to be value free cannot succeed; it can only amount to irresponsible carelessness about the consequences of the decisions that are made in the course of inquiry. Accounts of objectivity tend to be tied to this mistaken notion that there is a virtuous way of doing value-free science. As I have shown, despite the interesting ideas that have been posed in the attempt to save "objectivity" in the face of the demise of the ideal of value-free science, the concept is not worth saving. The work we wanted to do by appealing to objectivity was to ensure the trustworthiness of science. This should lead us to focus on scientific integrity rather than objectivity. Future work should focus on further (or better) articulating the requirements of scientific integrity, and the conditions that scaffold or inhibit its development, rather than trying to determine the nature of objectivity.

Notes

1 Biddle and Kukla (2017) argue much the same point using the language of "epistemic risk" where I refer to contingencies with significant social and ethical consequences.
2 The quality of this reporting often leaves something to be desired and is subject to a variety of common problems. See Kampourakis, *this volume*.
3 Even if the second option is merely not to proceed with the first option.

4 Though even such aims may be swayed by non-epistemic values. See Rooney (1992) and Longino (1996).
5 Value subjectivism is denied by moral realists, some moral naturalists, divine command theorists, cultural relativists, moral universalists, those who believe in intrinsic values, and many others.
6 A third sense of individual objectivity is what Douglas called "value neutrality" (Douglas 2004, 460, 2009, pp. 123–124). Being value neutral requires taking a middle or compromise position where values are controversial, being fair and balanced among competing positions. In some cases, this is a desirable approach, as when we hope to find an "objective" judge or mediator for a dispute. In other cases, the result is a centrist position that may be far from desirable.
7 Compare Richard Rorty on "accurate representation" as "empty compliment" (Rorty 1979, p. 10)
8 Wright pointed to "methodological generalizations" as an analogue to Brigandt's "theoretical generalizations" to answer this point. This argument seems to backfire to me, however. The account of methodology he points to uses "objectivity" in an unhelpfully vague and indeterminate way. It also contrasts "objectivity" with "interpretive judgment" in a way that makes clear that "objectivity" is not generally a good thing (because interpretive judgment is sometimes a good thing).
9 This account of scientific integrity is thus a form of coupled ethical-epistemic analysis as described by Nancy Tuana (2013).
10 Erik Fisher's Socio-Technical Integration Research program shows that through the intervention of a humanities scholar or social scientist embedded in the laboratory, scientists and engineers can improve their critical sensitivity (though this is not his term). See Fisher (2007); Fisher, Mahajan, and Mitcham (2006); Fisher and Schuurbiers (2013)
11 Sharyn Clough has been emphasizing the importance of "epistemic humility," along with empathy, as crucial to a peace-literacy approach to values in science; for example, in her talk at Southern Methodist University on "Science, Politics, and Peace Literacy" on March 2, 2018.

References

Alexandrova, Anna. 2017. *A Philosophy for the Science of Well-Being*. Oxford: Oxford University Press.

Anderson, Elizabeth. 1995. "Knowledge, Human Interests, and Objectivity in Feminist Epistemology." *Philosophical Topics* 23 (2): 27–58.

Biddle, Justin B, and Rebecca Kukla. 2017. "The Geography of Epistemic Risk." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards, 215–238. New York: Oxford University Press.

Brigandt, Ingo. 2003. "Species Pluralism Does Not Imply Species Eliminativism." *Philosophy of Science* 70 (5): 1305–1316.

Brown, Matthew J. 2013. "Values in Science Beyond Underdetermination and Inductive Risk." *Philosophy of Science* 80 (5): 829–839.

Collins, H. M. 1981. "Introduction: Stages in the Empirical Programme of Relativism." Social Studies of Science 11 (1): 3–10. www.jstor.org/stable/284733.

Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–579.

Douglas, Heather. 2004. "The Irreducible Complexity of Objectivity." *Synthese* 138 (3): 453–473.

Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.

Dupré, John. 2007. "Fact and Value." In *Value-Free Science?: Ideals and Illusions*, edited by Harold Kincaid, John Dupré, and Alison Wylie, 27–41. Oxford: Oxford University Press.

Elliott, Kevin C. 2013. "Douglas on Values: From Indirect Roles to Multiple Goals." *Studies in History and Philosophy of Science Part A* 44 (3): 375–383.

Fisher, Erik. 2007. "Ethnographic Invention: Probing the Capacity of Laboratory Decisions." *NanoEthics* 1 (2): 155–165.

Fisher, Erik, Roop L Mahajan, and Carl Mitcham. 2006. "Midstream Modulation of Technology: Governance from Within." *Bulletin of Science, Technology and Society* 26 (6): 485–496.

Fisher, Erik, and Daan Schuurbiers. 2013. "Socio-Technical Integration Research: Collaborative Inquiry at the Midstream of Research and Development." In *Early Engagement and New Technologies: Opening up the Laboratory*, edited by Neelke Doorn, Daan Schuurbiers, Ibo van de Poel, and Michael E. Gorman, 16:97–110. Philosophy of Engineering and Technology. New York: Springer.

Haack, Susan. 1993. "Epistemological Reflections of an Old Feminist." *Reason Papers* 18: 31–43.

Hacking, Ian. 2015. "Let's Not Talk About Objectivity." In Objectivity in Science, edited by Flavia Padovani, Alan Richardson and Jonathan Y. Tsou, 19–33. Cham: Springer.

Harding, Sandra. 1995. "'Strong Objectivity': A Response to the New Objectivity Question." *Synthese* 104 (3): 331–349.

Hicks, Daniel J. 2014. "A New Direction for Science and Values." *Synthese* 191 (14): 3271–3295.

Hicks, Daniel J, and Kevin C Elliott. 2018. "A Framework for Understanding Wishful Thinking." PhilSci Archive. http://philsci-archive.pitt.edu/14348/.

Intemann, Kristen. 2015. "Distinguishing Between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–232.

Lamb, Evelyn. 2012. "5 Sigma What's That?" Scientific American Observations Blog. https://blogs.scientificamerican.com/observations/five-sigmawhats-that/.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Longino, Helen E. 1996. "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, edited by Lynn Hankinson Nelson and Jack Nelson, 39–58. Dordrecht: Kluwer Academic Publishers.

Nagel, Thomas. 1986. *The View from Nowhere*. New York: Oxford University Press. www.loc.gov/catdir/enhancements/fy0638/85031002-d.html.

Pinch, Trevor. 2015. "Scientific Controversies." In *International Encyclopedia of the Social and Behavioral Sciences (Second Edition)*, edited by James D. Wright, 281–286. Oxford: Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.85043-6.

Porter, Theodore M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press. www.loc.gov/catdir/description/prin031/94021440.html.

Rooney, Phyllis. 1992. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1992: 13–22.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.

Scheman, Naomi. 2001. "Epistemology Resuscitated: Objectivity as Trustworthiness." In *Engendering Rationalities*, edited by Nancy Tuana and Sandra Morgen, pp. 23–52. Albany: State University of New York Press.

Staley, Kent W. 2017. "Decisions, Decisions: Inductive Risk and the Higgs Boson." In *Exploring Inductive Risk*, edited by Kevin C. Elliott and Ted Richards, 37–55. New York: Oxford University Press.

Steel, Daniel. 2017. "Qualified Epistemic Priority: Comparing Two Approaches to Values in Science." In *Current Controversies in Values and Science*, edited by Kevin Elliott and Daniel Steel, 49–63. New York: Routledge.

Tuana, Nancy. 2013. "Embedding Philosophers in the Practices of Science: Bringing Humanities to the Sciences." *Synthese* 190 (11): 1955–1973.

Wright, Jack. 2018. "Rescuing Objectivity: A Contextualist Proposal." *Philosophy of the Social Sciences*. https://doi.org/10.1177/ 0048393118767089.

PART IV

Is Scientific Knowledge Limited?

16

SHOULD WE TRUST WHAT OUR SCIENTIFIC THEORIES SAY?

Martin Curd and Dana Tulodziecki

Introduction

A major theme in the philosophy of science since the 1960s has been the dispute between scientific realists and antirealists.[1] Both sides take science seriously. And both sides think that scientific theories should be interpreted literally. Their disagreement is not *semantic*. It is not about what theoretical terms such as 'gene', 'electron, and 'quark' mean. Gene talk, if true, is about genes; electron talk, if true, is about electrons.

Realists and antirealists differ over the following, *epistemological* question: What do our best scientific theories – ones that are most successful – justify us in believing? Realists insist that we should accept our best, most successful, scientific theories as true (or, more cautiously, as probably true, or approximately true, or partially true, or close to the truth) including all the claims that such theories make about entities and processes that cannot be directly observed. Often called 'unobservables', these are things that either we cannot observe even in principle or, if we can detect them, we can do so only indirectly, using instruments such as microscopes, cloud chambers, and Geiger counters. When it comes to the unobservables postulated by even our best theories, antirealists advocate caution. Bas van Fraassen (1980) is a typical example. He thinks that when we accept a scientific theory we should limit our belief to what the theory says about *observable* objects and processes. We should withhold belief about the theory's claims about unobservables. According to van Fraassen, the aim of science is not to discover true theories but rather to come up with theories that are empirically adequate, theories that model observable phenomena in ever more accurate and precise ways.

In this chapter we explore the main argument for scientific realism, the No-Miracle Argument (NMA), and two antirealist arguments criticizing scientific realism, the Pessimistic Induction (PI) and the argument from Underdetermination (UDT).

The No-Miracle Argument (NMA)

The NMA takes its name from an oft-quoted sentence in a paper by Hilary Putnam (1975, p. 73): "The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle." Alan Musgrave (1988) aptly called it "The Ultimate Argument" for scientific realism. Putnam's talk of "a miracle" is his vivid way of saying that realism is the only account of science that does not make its success highly improbable. Realists who endorse the NMA (Boyd 1984, 1989; Musgrave 1988; Psillos 1999) usually understand it to be an inference to the best explanation (IBE).

IBE is common in everyday life and in the sciences. We observe some phenomenon, *P*, and then consider a number of competing hypotheses, each of which could explain *P*. If one of those hypotheses, *H*, gives the best explanation of *P*, then we conclude that *H* is probably true. The general form of an IBE is:

(1) *P*.
(2) Among the available, competing hypotheses that could explain *P*, the best explanation is *H*.

---

(3) [Probably] *H* is true.

Kevin McCain explores the role of IBEs in generating scientific knowledge in Chapter 4 of the present volume. In this chapter, we assume that IBEs rationally support their conclusions and focus on whether the same form of reasoning can be used to justify scientific realism. We shall ignore attacks on the NMA that deny that any argument to the best explanation is justified, or that limit the scope of justified IBEs solely to those whose conclusions can be secured through regular inductive reasoning.[2] Granted, just like regular inductive arguments, IBEs are not deductively valid: their conclusions can be false even when their premises are true. But that does not establish that IBEs cannot justify their conclusions. As we shall see, scientific realists regard IBEs as justified in the same general way that inductive reasoning is justified. In either case, the source of that justification is held to lie in a proven track record of reliability within the empirical sciences.

In its simplest form, the NMA runs as follows:

(1) Scientific theories (in the mature sciences) are successful.

(2) Among the available, competing hypotheses that could explain the success of scientific theories (in the mature sciences), the best explanation [by far] is their (approximate) truth.

---

(3) [Very probably] scientific theories (in the mature sciences) are (approximately) true.

The phrases in square brackets capture the realist's insistence that because nothing comes close to scientific realism in explaining the success of science, it should be accepted as *highly* probable. The phrases in parentheses are qualifications that realists make in order to avoid obvious objections. Not all scientific theories are referred to, for those would include theories from the early beginnings of chemistry, psychology, medicine, etc., when those sciences were what Kuhn calls "immature." Theories from the days of alchemy, astrology, and ancient medicine, for example, presumably were not successful in anything like the sense in which our current scientific theories are. The other qualification is the substitution of "approximate truth" for truth. This is a concession to the fact that science improves with age. Many past theories in the mature sciences that worked well, perhaps for a considerable time, were subsequently discovered to be false. The idea is to try and avoid this problem for the NMA by attributing to those theories a degree of approximate (or perhaps, partial) truth sufficient to explain their success in the past. (This problem, for the realist, of false but successful theories is discussed further, below, in the section on the Pessimistic Induction.)

An important virtue of the NMA in the eyes of realists who endorse it is that it is thought to reflect a "naturalist" attitude to philosophy of science (and to philosophy more generally) by using the same methods as scientists do when choosing among competing theories. Just as scientists accept the particular scientific theories that give the best explanation of the relevant data and phenomena, so too should philosophers of science accept the best philosophical explanation of why our best scientific theories are so successful, namely, they are successful because they are true (or approximately true, or close to the truth, etc.). More cautiously: it is because the electron theory, say, successfully refers to electrons (i.e., electrons really exist) and the theory more or less correctly describes the laws governing their behavior, that the electron theory is successful.

What does "success" mean in the NMA? The usual answer is that success for a scientific theory means making the right predictions. The more predictions a theory makes, of different kinds, and of impressive accuracy, the more successful the theory. Quantum electrodynamics (QED) predicts the magnetic moment of the electron to better than 1 part in a billion. How could QED do so exquisitely well unless electrons exist and behave very much as the theory describes? Despite this, some realists have argued that predictive accuracy by itself is insufficient. At the very least, we need some assurance that the prediction is one that the theory

would have gotten wrong if the theory were in fact false. Ptolemaic astronomy is a good example of a theory that had impressive predictive accuracy; but most of its correct predictions were secured by fixing adjustable parameters within the planetary mechanisms used by the theory. The values of those parameters came from previous observations. With sufficient tinkering, the theory was bound to predict close to the right answers even though it was false.[3] For this reason, many scientific realists insist that predictive success be confined to novel prediction, rather than predictions in general.[4] Others (McMullin 1993; Doppelt 2007) have argued that "success" should also be broadened in scope to include other theoretical virtues such as explanatory scope, fertility, and consistency with other well-accepted scientific theories.

The NMA's second premise asserts that the (approximate) truth of our scientific theories is the best explanation of their success. Are there no other explanations of success than truth? Van Fraassen argues that the success of science is no more surprising than the emergence of biological species adapted to their environments: both are the result of selection. Scientists value empirically adequate theories and discard those that fail to meet that standard. Truth in the full-blown realist sense has nothing to do with it. The realist reply is that van Fraassen's "selectionist" criticism has changed the issue. The relevant question for the realist is, "What is it about successful scientific theories that accounts for their success?" Selection does not explain why this or any other particular theory succeeds, no more than does membership in an exclusive club that admits only redheads explain why a person has red hair (Lipton 2005, p. 1267).

What does it mean for one thing to explain another? If we insist that explainers have to entail what they explain, then proponents of the NMA would have to defend this claim: if a theory is true, then it will be successful. But truth all by itself seems too meager to guarantee success. Without additional restrictions, a theory could be true but so limited in its scope and so insulated from the rest of science as to yield few if any interesting predictions. Genuine explanations (in science) must have other desirable features. Michael Levin (1984), Michel Ghins (2002), and Greg Frost-Arnold (2010) attack the notion that truth explains success given the way that working scientists understand explanation. Frost-Arnold, for example, argues that proponents of the NMA who regard themselves as naturalists will run afoul of the following requirement on genuine explanations: they must either generate novel predictions or unify apparently disparate established claims (Frost-Arnold 2010, p. 37). Anything that does neither of these two things is not a genuine (scientific) explanation.

Finally, there are doubts about whether realists are committed to endorsing the truth of a scientific theory in its entirety as an explanation of that theory's success. For, as Kitcher (2001), Psillos (1999), and other realists have argued, we should focus solely on those parts of a theory, its "working posits," that are relevant to its successful predictions. Superfluous parts, mere "idle wheels," should not receive credit that they have not earned. Thus, the conclusion of the NMA should be

construed more narrowly as inferring the (approximate) truth only of those theoretical assumptions that are indispensable to the successful predictions achieved by using them.[5]

Scientific realists have articulated many different versions of their doctrine in response to the acknowledged shortcomings of the original NMA. While most still rely on an IBE, they propose stricter notions of novel predictive success, richer notion of success in general (to include theoretical virtues such as explanatory power that are important in scientific practice), and more discriminating ways of identifying the parts of a theory responsible for that success. In addition to these strategies, there are also variants of realism that have tried to be more selective about which parts of our theories we can make the best epistemic case for. Entity realists, for example, insist that we ought to believe only in those entities we can causally manipulate (for example, Hacking 1982). According to structural realists, we ought to believe in theoretical structures, not entities (Worrall 1989). Chakravartty's semirealism combines elements from both, with Chakravartty arguing that we have epistemic justification only for believing in detection properties, properties that are "causally linked to the regular behavior of our detectors" (2007: 47).

The Pessimistic Induction (PI)[6]

When one looks at the history of science, especially the scientific revolutions in physics, astronomy, and chemistry chronicled in Kuhn (1962), the realist seems to be in a bind. For Kuhn, and later Larry Laudan (1981), gave many examples of theories that were successful in their day and yet we now recognize as based on radically false assumptions about the world. Ptolemaic astronomy, Newton's gravitational theory, the phlogiston theory of chemistry, the caloric theory of heat, the wave theory of light, Maxwell's electromagnetic theory, and so on, all posited entities and processes that, by our lights, do not exist. There are no epicycles, no instantaneous action at a distance, no phlogiston, no caloric, no mechanical aether, no optical aether, no substance-like electromagnetic aether at rest in absolute space. Hence, Laudan concludes, the theories that contained terms that purport to refer to such entities must have been false. And, just as past theories' successes did not guarantee their truth, we ought not view our current theories' successes as reliable indicators of theirs.

We expect the PI to be an inductive argument with a pessimistic conclusion. But it is important to distinguish between two different things that such "pessimists" might be arguing. It might be that the pessimist aims to draw a conclusion that directly contradicts the epistemic optimism of realists. As we have seen, realists typically deploy a version of the "No-Miracle" argument (NMA) to conclude that we have good reason to believe that our best, successful current theories are true. On one reading of the PI, the pessimist thinks that the many cases of theories that were successful in their day but turned out to be false is good

inductive grounds for thinking that our current successful theories are likely to be false, too. This is the "direct" version of the PI: an induction from past failures to present pessimism and the likelihood of replacement in the future. Peter Lipton (2000) has called this "the disaster argument."

There is a second, subtler version of the PI. On this reading, pessimists are not arguing that our current scientific theories are likely to be false; rather they aim to undermine a key premise in the realist's NMA. In other words, they are attacking the realist's argument for optimism, not arguing directly for pessimism. (An analogy would be the difference between criticizing a theist's argument for God's existence and offering a direct argument for atheism.) The premise under attack in this version of the PI is that success (of a scientific theory) is a reliable indicator of truth. This "metalevel" version of the PI has come to be known as the "pessimistic meta-induction" or PMI.

There have been many different realist replies to the PI and the PMI. Some realists argue that the evidence for our current scientific theories is much better than what supported theories in the past. Others (Kitcher 2001; Psillos 1999) have argued that past theories did succeed in referring to unobservable entities but past scientists had false beliefs about some of the properties of those entities (e.g., when Fresnel talked about the optical aether he was really referring to the electromagnetic field). Others see no problem in allowing that a theory can be approximately true even though one or more of its central terms do not refer. Hardin & Rosenberg (1982) gave the example of "gene." There is no single unitary entity that has all the functional, hereditary, and mutational properties that were ascribed to genes. Instead, those properties are now parceled out to different segments and combinations of DNA. In this way, we can see why classical genetics worked as well as it did: it was close to the truth even though strictly false.

According to Lewis (2001), Saatsi (2005), and others there is a flaw in Laudan's PMI, construed as a meta-level argument.[7] Laudan gives numerous examples from the history of science of theories that were once successful but are now regarded as false. Is this sufficient to show that success is not a reliable indicator of truth? Lewis says "no" because he argues that the argument commits *the base rate fallacy*.[8]

The base rate fallacy is familiar from medical contexts in which a diagnostic test, *S*, is relied on to indicate a disease, *T*. In just the same way, realists insist that success (*S*) is a reliable indicator of truth (*T*) because, *according to realists, most true theories will be successful*. In other words, realists claim that $P(S|T)$ – the probability of success given truth – is high, say 0.9. But it doesn't follow from this alone that the proportion of true theories among successful theories in the past will be high, no more than it follows that most people who show positive on a diagnostic test will have the disease in question. Two other things are relevant to the value of $P(T|S)$. First, how specific is the test? In other words, how likely is someone to show positive on the test even though they are in fact disease-free? The analogous question for the PMI is: How probable is it that a theory will be successful even though false? The realist assumes that $P(S|\sim T)$ is low because false theories

(unlike true ones) will make false predictions and thus be much more vulnerable to failure. A typical realist assumption is that $P(S|{\sim}T)$ is less than a half, say 0.25. So, in regarding success as a reliable indicator of truth, the scientific realist will typically assume that $P(S|T) = 0.9$, and $P(S|{\sim}T) = 0.25$.

The other relevant factor in determining the value of $P(T|S)$ is the base rate, $P(T)$. How prevalent is the disease in the population from which patients are being selected for testing? If the disease is very rare, then most of the positive tests will be false positives; they will come from people who are disease-free. Even though $P(S|T)$ is high and $P(S|{\sim}T)$ is low, arguing from S to T is warranted only when $P(T)$ is high. Analogously, in the scientific case, when true theories are rare and false theories common, most cases of success will come from false theories. In citing many cases of false but successful theories from the history of science as an argument against success as a reliable indicator of truth, Laudan has ignored the crucial importance of the base rate, $P(T)$.

Realists are optimists. They assume that science gets better as time goes on: false theories are weeded out, and a higher proportion of true theories are retained. On the realist picture, therefore, a much higher percentage of accepted theories are true now than in the past. Thus, we would expect that, in the past, success was a poor guide to truth, not because true theories are not more likely to be successful but because in the past there was a much higher fraction of false theories among those that were successful. Now, in the present, with the advantage of several hundred years of scientific progress, a much higher fraction of accepted theories is true. Hence, a realist might argue, a much higher proportion of successful theories will turn out to be true. Saatsi's diagram illustrates how the fraction of false theories that are successful (0.25) and the fraction of true theories that are successful (0.9) can remain constant over time while the proportion of successful theories that are true increases as true theories begin to outnumber false ones.

There is much here that can be contested. One thing to be clear about is that this is not intended to be an argument for the realist position that will convince

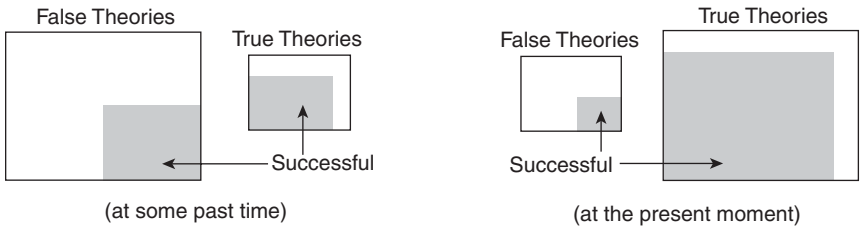

FIGURE 16.1 The ratio of successful (gray) to unsuccessful (white) theories is constant over time.

Source: From p. 1095 of Juha T. Saatsi, "On the Pessimistic Induction and Two Fallacies," *Philosophy of Science* 72 (December 2005): 1088–1098.

antirealists since antirealists will reject the assumption that the proportion of true to false theories in science grows over time. The point of the base-rate-fallacy response is to show that, in appealing to cases of successful but false theories in the past of science, the antirealist has not made a decisive case against the realist conviction that the success of theories is a reliable indicator of their truth.

In particular, Lewis denies the validity of the inference in Laudan's PMI from "Success is a reliable indicator of truth" to "Most successful theories (whether current or past) are true." This is because Lewis chooses to interpret "Success is a reliable indicator of truth" (by analogy with diagnostic tests in medicine) as "The probability of success given truth is high." The price of identifying reliability as $P(S|T)$ is that no conclusion about $P(T|S)$ can be drawn without knowing the base rate, $P(T)$. Thus, Lewis is able to resist Laudan's PMI only by relying on assumptions that no antirealist would be willing to accept. By the same token, realists are unwilling to accept the assumptions about base rates that, according to the same analysis, are required for the pessimistic induction of the antirealists. It is for this reason that Magnus and Callender (2004) conclude that the realism versus antirealism debate at this general, "wholesale" level is a standoff.

Some realists have tried to undermine the PMI by considering theories at the particular, "retail" level to motivate refinements of selective realism. Vickers (2013), for example, offers 20 new examples from the history of science of false theories that were successful in their day, but which, unlike Laudan's twelve examples, also appear to have made novel predictions. By considering several of these cases in detail, Vickers argues for stronger concepts of "novel prediction" and "working posit" that realists can then use to disqualify such cases from counting as genuine counterexamples (for a response, see Tulodziecki (2017)).

Structural realists also think they can avoid the PMI, since continuity of reference with respect to entities is not required for retention of structures. They claim to be immune to the examples on Laudan's list, since that list lacks cases of genuinely successful past theories whose structural elements turned out to be completely false. Worrall (1989), for example, gives Fresnel's theory of light as an example of a theory whose equations were preserved intact in Maxwell's theory of electromagnetism. Psillos (1995) disagrees, arguing that realists require more than the preservation of purely mathematical equations devoid of a physical interpretation. This debate, just like the discussion surrounding the NMA, has not been settled and has recently been subject to renewed interest through the introduction of new historical cases.

The Underdetermination Argument (UDT)[9]

Another avenue of attack against scientific realism exploits the gap between our scientific theories and the observational evidence: evidence alone, no matter how extensive or good, will never single out just one theory as the correct one. Since any set of evidence is always logically compatible with many different theories, it

can never guarantee that one particular theory is correct; thus, scientific theories are underdetermined by the evidence. Underdetermination is sometimes likened to plotting a finite set of points on a graph. Even if there are lots of points, we can draw many different curves that pass through them. Similarly, there are always in principle many different theories that will entail a given set of evidence.

Proponents of the underdetermination argument (UDT) often present it as a direct threat to scientific knowledge. Scientific knowledge typically comes in the form of scientific theories, but if the UDT is correct and we cannot pick out scientific theories solely on the basis of evidence, it follows that we cannot have scientific knowledge either. Scientific realists and other opponents of the UDT think we should not be that worried about underdetermination because there are other criteria besides the evidence alone that can help us decide what theories to select and that, as science progresses, we get better at this.

There are many questions about underdetermination: What kinds of underdetermination, if any, are pervasive in science? Does underdetermination really pose a threat to scientific knowledge? Are the many theories that are compatible with the evidence equally good? For example, are they all scientifically credible? Do they explain the evidence equally well? Does the evidence confirm them to the same extent? Does it matter how well they fare with respect to everything else we know? And so on. To answer these questions, philosophers of science have made the UDT more precise. It is usually characterized as follows:

(1) The available observational evidence (including possible future evidence) is always compatible with several theories, not more than one of which can be true.
(2) The observational evidence is the only epistemically legitimate reason for believing our scientific theories to be (approximately) true.

---

(3) We cannot have any justification for picking one scientific theory over its rivals.

The first premise is sometimes called the 'Empirical Equivalence Thesis' (EET) because it says that every theory has logically incompatible empirically equivalent rivals (Kukla 1998; Psillos 1999).[10] The second premise is sometimes referred to as the 'Entailment Thesis' (ET) because it asserts that the only epistemic constraint on theory-choice is entailment of the empirical evidence (Kukla 1998; Psillos 1999). From EET and ET, proponents of underdetermination claim that it follows that we can never be justified in picking one scientific theory over its competitors. And if we cannot do that, we cannot have the kind of scientific knowledge that scientific realists assume we have.

What is usually taken to be at stake in this debate is not merely a temporary version of the problem, according to which theories are underdetermined by the

currently available observational evidence – temporary underdetermination is a frequent and familiar scientific predicament, and often scientific experiments are designed precisely to settle such matters. Instead, the argument seeks to establish a permanent version of the problem, according to which even any potential future evidence could not help break the tie among competing theories. In the literature, a famous example of the latter has been that of Newtonian mechanics coupled with a variety of competing hypotheses about the absolute velocity of the universe's center of mass (although some philosophers deny that this case is interesting since the rival hypotheses involve the same ontology and concepts; Earman 1993). If Newtonian mechanics were correct, then such apparently different Newtonian universes would be indistinguishable regardless of any further evidence. So, in this case future evidence could never break the tie.

One argument in favor of the first premise – the premise that says that the observational evidence is always compatible with many theories – is based on the Duhem-Quine thesis. The Duhem-Quine thesis says that scientific theories do not entail observational consequences on their own, but only when conjoined with various auxiliary assumptions, initial conditions, and background theories (about how our instruments work, for example). Here is how this helps to establish premise (1): if a theory does not entail certain evidence, and so is not empirically equivalent to an existing scientific theory, that theory can be made to entail the right evidence by embedding it in a different auxiliary framework. If one changes one's auxiliary assumptions in just the right way, then that will give us a theory empirically equivalent to the first. And, since the auxiliaries can be changed in quite creative ways, it seems possible to do this for any theory whatsoever.

In a related vein, proponents of underdetermination have pointed to a number of algorithms that could be used in constructing rival theories. Some of these algorithms are quite fanciful, such as Kukla's hypothesis (Kukla 1998, pp. 74–77) that the universe was created by beings in such a way that it seems identical to ours, but depends on a machine that is occasionally turned off (for maintenance and repair), thus causing our world to pop out of existence and then back in again in a way that we cannot ever detect.

Realists have typically dismissed these sorts of cases since they consider them trivial and not much different from skeptical hypotheses. Antirealists need something more substantial than a mere skeptical scenario for the argument to work, since whatever they propose as a rival, has to work specifically against scientific theories, but not equally threaten our general knowledge of everyday observable things. They don't want to throw out the trees with the electrons. So, there is a debate here about what sorts of theories ought to count as proper competitors to our scientific theories (Laudan and Leplin 1993; Kukla 1993).

Further, Laudan and Leplin (1991) have argued that the mere existence of empirically equivalent theories is not good enough to establish a permanent version of premise (1). Even if two theories are observationally equivalent now, nothing guarantees that they will remain so in the future. First, what is observable

may change, and if what is unobservable now becomes observable, two theories that are now observationally equivalent may cease to be so in the future. Second, a given theory's auxiliaries may change, and so in the future it may have different observational consequences from its current rival(s).[11] Proponents of underdetermination have countered that we can still run the UDT at the level of total science, where the unit of underdetermination is not a specific scientific theory, but the state of total science at a given time (Okasha 2002).

According to premise (2), theories that are observationally equivalent are equally worthy of belief. Laudan and Leplin (1991) reject this. They argue that, even if the second premise were true and observational evidence the only legitimate factor in theory-choice, underdetermination still would not follow. First, they point out that merely being an observational consequence of a theory is insufficient for being evidence for that theory, since not every observational consequence confers confirmation on a theory. Second, theories can receive support from evidence they do not entail (see also Werndl 2013; for a response to Laudan and Leplin, see Okasha 1997). Lastly, one also ought to note that even if two theories have exactly the same observational consequences, this does not mean that these consequences support the theories equally well (for work on empirical support relations, see Earman 1993; Mayo 1997; Massimi 2004).

In another line of response to premise (2), realists have argued that there are many other criteria besides just the observational evidence that govern what theory to choose – for epistemically legitimate reasons. Popular candidates are various theoretical virtues such as coherence with other theories, unifying power, consilience, genuine explanatory power, a theory's capacity to generate novel predictions, its not being ad hoc, simplicity, and so on (see McMullin 1993, 2014). Most realists think that at least some of these virtues are epistemic properties and that theories possessing them are epistemically superior to theories lacking them. As a result, these virtues can help us pick out theories that are more likely to be (approximately) true than others. Anti-realists, of course, disagree, and instead have argued that these virtues are solely aesthetic or pragmatic (van Fraassen 1980, p. 87). A remaining problem for the realist response to premise (2) of the UDT is to specify how exactly theoretical virtues are supposed to link with truth. For relevant work in this area, see Kelly 2007; Psillos 1999; Tulodziecki 2013.

Conclusion

Although the classic arguments in the scientific realism debate are by now several decades old, the debate is still very much alive. One recent realist trend has been to go particularist or local by arguing on a case-by-case basis instead of relying on the general strategies of the NMA, PMI, and UDT, global arguments which are supposed to apply to all mature and successful scientific theories, regardless of discipline. Localists have criticized these arguments for being overly general and disregarding important features of actual science. In order to remedy this situation,

localists offer more extended and detailed case-studies of individual episodes in the history of science. A welcome outcome of this has been to bring into the debate many new cases from different fields. This diversity is important, because it is an open question to what extent we ought to regard the various scientific fields as governed by the same principles and methods. Saatsi, for example, asks:

> In the face of all the diversity, why think that one (or even a handful) of recipes uniformly and fairly captures – across the board – the way in which theories' empirical success is correlated with the way they latch onto reality? (2017, p.6)

Other realists have questioned whether, by engaging in localist strategies, realists are forced to sacrifice too much (see, for example, Henderson 2017). Virtually every aspect of the realism debate is still open; for a comprehensive overview, see Saatsi (2018). Wherever the participants in the debate stand, it is important to emphasize that none of them doubt the credibility of science.

Notes

1 The pivotal year for scientific realism was 1962, with the publication of papers by Hilary Putnam ("What Theories Are Not") and Grover Maxwell ("The Ontological Status of Theoretical Entities") arguing that there is no distinction in kind between observational and theoretical terms that can carry the epistemological weight assigned to it by logical empiricists. Once freed from the empiricists' bad semantic theory and the skepticism about theoretical entities to which it had led, most philosophers of science welcomed defenses of scientific realism by figures such as Wilfrid Sellars, J. J. C. Smart, Richard Boyd, W. H. Newton-Smith, and Ernan McMullin. The first major backlash against scientific realism's new popularity came from Bas van Fraassen.

2 What we have in mind here is Bas van Fraassen's "bad lot" objection to the NMA, which seems to depend, at least in part, on a general skepticism concerning IBEs. For a rebuttal to van Fraaseen, see Schupbach (2014). The epistemological debate about explanatory inference is reviewed in Lycan (2002). For some of the difficulties in reconciling explanatory reasoning with the Bayesian approach to confirmation, see Henderson (2014) and the exchange between Peter Lipton and Wesley Salmon in Hon and Rakover (2001).

3 Realists such as Ernan McMullin (1993) see an important contrast here with the genuine explanatory power of the essential elements of the Copernican theory. See Hall (1970) and Swerdlow (2004) for more on how Copernicus explains what Ptolemy merely predicts.

4 There is interesting work by Carrier (1991, 1993) and Carman and Díez (2015) arguing that important but false theories in the past – the phlogiston theory and Ptolemaic astronomy, for example – generated novel predictions.

5 The truth of a theory *in its entirety* is not something that realists should be concerned about, especially since one false conjunct, however trivial and irrelevant, suffices to render false an entire conjunction of claims. What matters are the components of a theory that are *essentially* involved in the theory's success.

6 This section is adapted from the Commentary to chapter 9 of Curd, Cover, and Pincock (2013).
7 Most responses to the historical objection have focused on the PMI. For a notable exception arguing specifically against the PI, see Lange (2002).
8 In an interesting twist, some philosophers of science have accused the NMA of committing the base rate fallacy. For details and a defense of the NMA against this charge, see Henderson (2017).
9 For a more detailed discussion of the issues in this section, see Tulodziecki (2018).
10 Two theories are empirically equivalent just in case they have the same observational consequence classes or share the same class of empirical models. Note that EET is a problem for scientific realism only if the theories in question are logically incompatible; if they were compatible, there would be no need to pick one over the other(s).
11 Of course, the Duhem-Quine thesis guarantees that the new theory will also have empirically equivalent rivals.

References

Boyd, Richard N. 1984, 'The Current Status of Scientific Realism', in *Scientific Realism*, J. Leplin (ed.), Berkeley: University of California Press, pp. 41–82.

Boyd, Richard N. 1989, 'What realism implies and what it does not', *Dialectica*, 43(1–2): 5–29.

Carman C., & Díez, J., 2015, 'Did Ptolemy make novel predictions? Launching Ptolemaic astronomy into the scientific realism debate', *Studies in History and Philosophy of Science* 52, 20–34.

Carrier, M., 1991, 'What is wrong with the miracle argument?' *Studies in History and Philosophy of Science* 22, 23–36.

Carrier, M., 1993, 'What is right with the miracle argument: Establishing a taxonomy of natural kinds', *Studies in History and Philosophy of Science* 24, 391–409.

Chakravartty, A., 2007, *A metaphysics for scientific realism: Knowing the unobservable*, Cambridge University Press, Cambridge.

Curd, M., Cover, J. A., & Pincock, C., 2013, (eds.), *Philosophy of science: The central issues*, 2nd edn, W. W. Norton, New York.

Doppelt, G., 2007, 'Reconstructing scientific realism to rebut the pessimistic meta-induction', *Philosophy of Science* 74, 96–118.

Earman, J., 1993, 'Underdetermination, realism, and reason', *Midwest Studies in Philosophy* 18, 19–38.

Frost-Arnold, G., 2010, 'The no-miracles argument for realism: Inference to an unacceptable explanation', *Philosophy of Science* 77, 35–58.

Ghins, M., 2002, 'Putnam's no-miracle argument: A critique', in S. Clarke and T. Lyons, (eds.), *Recent themes in the philosophy of science: Scientific realism and commonsense*, pp. 121–138, Kluwer, Dordrecht.

Hacking, I., 1982, 'Experimentation and scientific realism', *Philosophical Topics*, 13(1), 71–87; reprinted in Leplin 1984, pp. 154–172.

Hall, R. J., 1970, 'Kuhn and the Copernican Revolution', *British Journal for the Philosophy of Science* 21, 196–197.

Hardin, C. L., and Rosenberg, A., 1982, 'In defense of convergent realism', *Philosophy of Science* 49, 604–615.

Henderson, L., 2014, 'Bayesianism and inference to the best explanation', *British Journal for the Philosophy of Science* 65, 687–715.

Henderson, L., 2017, 'The no miracles argument and the base rate fallacy', *Synthese* 194, 1295–1302.

Hon, G. & Rakover, S. S., (eds.), 2001, *Explanation: Theoretical approaches and applications*, Kluwer, Dordrecht.

Kelly, K., 2007, 'A new solution to the puzzle of simplicity', *Philosophy of Science*, 74, 561–573.

Kitcher, P., 2001, 'Real realism: The Galilean strategy', *Philosophical Review* 110, 151–197.

Kuhn, T. S., 1962, *The structure of scientific revolutions*, University of Chicago Press, Chicago; 2nd edn, 1970; 3rd edn, 1996.

Kukla, A., 1993, 'Laudan, Leplin, empirical equivalence, and underdetermination', *Analysis* 53, 1–7.

Kukla, A. 1998, *Studies in scientific realism*. Oxford: Oxford University Press.

Lange, M., 2002, 'Baseball, pessimistic inductions and the turnover fallacy', *Analysis* 62 (4), 281–85.

Laudan, L., 1981, 'A confutation of convergent realism', *Philosophy of Science* 48, 19–49; reprinted in Leplin 1984, pp. 218–249.

Laudan, L. & Leplin, J., 1991, 'Empirical equivalence and underdetermination', *Journal of Philosophy* 88, 449–472.

Laudan, L., & Leplin, J., 1993, 'Determination undeterred: Reply to Kukla', *Analysis* 53, 8–16.

Levin, M., 1984, 'What kind of explanation is truth?' in J. Leplin (ed.), *Scientific realism*, pp. 124–139, University of California Press, Berkeley.

Lewis, P. J., 2001, 'Why the pessimistic induction is a fallacy', *Synthese* 129, 371–380.

Lipton, P., 2000, 'Tracking track records', *Aristotelian Society Supplementary Volume* 74 (1), 179–205.

Lipton, P., 2005, 'The truth about science', *Philosophical Transactions of the Royal Society B* 360, 1259–1269.

Lycan, W. G., 2002, 'Explanation and epistemology', in P. K. Moser (ed.), *The Oxford handbook of epistemology*, pp. 408–433, Oxford University Press, New York.

Magnus, P. D., & Callender, C., 2004, 'Realist ennui and the base rate fallacy', *Philosophy of Science* 71, 320–338.

Massimi, M., 2004, 'What demonstrative induction can do against the threat of underdetermination: Bohr, Heisenberg, and Pauli on spectroscopic anomalies (1921–24)', *Synthese* 140, 243–277.

Mayo, D. G., 1997, 'Severe tests, arguing from error, and methodological underdetermination', *Philosophical Studies* 86, 243–266.

McMullin, E., 1993, 'Rationality and paradigm change in science', in P. Horwich (ed.), *World changes: Thomas Kuhn and the nature of science*, pp. 55–78, MIT Press, Cambridge, Mass.

McMullin, E., 2014, 'The virtues of a good theory', in M. Curd and S. Psillos (eds.), *The Routledge companion to philosophy of science*, 2nd edn, pp. 561–571, Routledge, New York.

Musgrave, A., 1988, 'The ultimate argument for scientific realism', in R. Nola (ed.), *Relativism and realism in science*, pp. 229–252, Kluwer, Dordrecht.

Okasha, S., 1997, 'Laudan and Leplin on empirical equivalence', *British Journal for the Philosophy of Science* 48, 251–256.

Okasha, S., 2002, 'Underdetermination, holism and the theory/data distinction', *Philosophical Quarterly* 52, 303–319.

Psillos, S., 1995, 'Is structural realism the best of both worlds?' *Dialectica* 49, 15–46.

Psillos, S., 1999, *Scientific realism: How science tracks truth*, Routledge, London.

Putnam, H., 1975, *Philosophical papers*, *Vol. 1: Mathematics, matter and method*, Cambridge University Press, Cambridge.

Saatsi, J., 2005, 'On the pessimistic induction and two fallacies', *Philosophy of Science* 72, 1088–1098.
Saatsi, J., 2017, 'Replacing recipe realism', *Synthese* 194(9), 3233–3244.
Saatsi, J. (ed.), 2018, *The Routledge handbook of scientific realism*, Routledge, New York.
Schupbach, N., 2014, 'Is the bad lot objection just misguided?' *Erkenntnis* 79, 55–64.
Swerdlow, N. M., 2004, 'An essay on Thomas Kuhn's first scientific revolution, *The Copernican Revolution*', *Proceedings of the American Philosophical Society* 148, 64–120.
Tulodziecki, D., 2013, 'Underdetermination, methodological practices, and realism', *Synthese*, 190(17), 3731–3750.
Tulodziecki, D., 2017, 'Against selective realism(s)', *Philosophy of Science* 84, 996–1007.
Tulodziecki, D., 2018, 'Underdetermination', in J. Saatsi (ed.) *The Routledge handbook of scientific realism*, pp. 60–71, Routledge, London and New York.
Van Fraassen, B., 1980, *The scientific image*, Clarendon Press, Oxford.
Vickers, P., 2013, 'A confrontation of convergent realism', *Philosophy of Science* 80, 189–211.
Werndl, C., 2013, 'On choosing between deterministic and indeterministic models: Underdetermination and indirect evidence', *Synthese* 190, 2243–2265.
Worrall, J., 1989, 'Structural realism: The best of both worlds?' *Dialectica* 43(1–2), 99–124.

17

WHAT ARE THE LIMITS OF SCIENTIFIC EXPLANATION?

Sara Gottlieb and Tania Lombrozo

Introduction

Mary is a brilliant scientist who specializes in human vision. Her mind is so able, and her knowledge so comprehensive, that she knows all the physical facts there are to know about the perception of color. She knows exactly how light of different wavelengths travels through the environment, how it affects our retinae, and what happens in our brains when we see different colors. She could describe the firing of every neuron and how it relates to what people report that they see. And yet, poor Mary herself has never seen the color blue (or red, or yellow …). Due to unspecified forces, she has experienced the world from the confines of a black and white room, with her only access to the outside world provided through a black and white monitor.

One glorious day, Mary emerges from her black and white chamber to the outside world. For the first time, she sees a blue sky, a red flower, a yellow bird. As the world's expert on color vision, she already knows precisely how each surface affects the wavelengths that bounce from it, and how her brain responds to the corresponding stimulation on her retina. Yet she has never had the first-hand experience of observing these colorful entities with her very own eyes. In having this experience, does Mary learn something new about color? Or, as an expert with knowledge of all the scientific facts about color, did she already know all there is to know?

This famous thought experiment by the philosopher Frank Jackson (1986) motivates a compelling intuition: that some things can only be known through personal experience. While Mary knows everything there is to know about the science of color, there is something she doesn't know. Before emerging from her chamber, she doesn't know *what it is like* to see blue (or red, or yellow …).

Jackson's thought experiment is usually framed in terms of the "physical information" that Mary does and doesn't know, but the example also suggests that there may be some types of knowledge that fall beyond the scope of science. Perhaps this seems obvious – science cannot, after all, tell us what values we ought to have, or what sorts of behaviors are morally good. Science is a *descriptive* enterprise, not a *prescriptive* one. But Jackson's thought experiment is powerful because it suggests that even something like color vision – a descriptive matter that vision scientists like Mary are able to study empirically – might fall beyond the scope of science. For if the first-person perspective that comes from actually *experiencing* color teaches Mary something new – *what it is like* to see blue, or red, or yellow – then there must be some kinds of knowledge about human color vision that cannot be derived from physical (or scientific) facts alone.

Those who accept Jackson's argument (and not everyone does) face a difficult choice. One possibility is to radically change the way we think about scientific knowledge to broaden its scope. More precisely, our notions of science and scientific knowledge could be expanded to include the kind of first-personal knowledge that comes from first-hand experience: the *what-it's-like* to see blue. But it's not really clear how this would work. For Mary, it would mean a rejection of the premise that she can know all "scientific facts" about color from the confines of her black and white chamber. When she emerges from her chamber, she would gain new "scientific" knowledge. This challenges the way we normally think about science as an enterprise concerned with objective and verifiable knowledge – the sort of knowledge that can be captured in textbooks or formal models.

A second possibility is to accept that scientific knowledge is limited in an important respect. On this view, a complete scientific explanation for human color perception leaves something out: it doesn't supply Mary with what she needs to know *what it's like* to experience color. Correspondingly, first-person experience can supply something that falls beyond the scope of science – something that cannot be captured by a scientific explanation.

The Perceived Limits of Science

Philosophers and scientists have debated the correct response to Jackson's argument, with no clear consensus. But anecdotal evidence suggests that for many people, the view that first-person experience can supply something that falls beyond the scope of science is especially compelling. Some phenomena – especially some *psychological* phenomena – just seem to elude a scientific explanation. Could science ever really explain romantic love, altruistic behavior, or religious faith? *Should* science even be concerned with these seemingly ineffable aspects of the human condition? Perhaps scientific explanations are, in principle, limited in these domains, and perhaps this limitation is a good thing.

One source of evidence that (some) people find a view along these lines compelling comes from the experience of researchers studying romantic love (Hatfield

2006). In the mid-1970s, psychologists Elaine Hatfield, Mary Utne O'Brien, and Jane Traupmann Pillemer were awarded a small grant by the National Science Foundation for their research on passionate love and sexual desire. They were also awarded a "Golden Fleece Award" by U.S. Senator William Proxmire, who claimed that they were "fleecing" taxpayers with their research. A press release explained: "not even the National Science Foundation…can argue that falling in love is a science" (Hatfield 2006). He also opposed the research because he didn't *want* the answer: "I believe that 200 million other Americans want to leave some things in life a mystery" (Hatfield 2006). Proxmire urged the NSF to leave love to Elizabeth Barrett Browning and Irving Berlin.

Other examples of perceived limits to science come from the domains of religion and spiritual experience. In a 2012 commentary published in the journal *Nature*, for example, author Daniel Sarewitz compared the discovery of the Higgs boson to the experience of visiting the Angkor temples in Cambodia. After describing the powerful sense of mystery and transcendence elicited by the temples, alongside the sense of a universe that evades comprehension, he wrote that: "Science is supposed to challenge this type of quasi-mystical subjective experience, to provide an antidote to it." Religion can offer "an authentic personal encounter with the unknown," whereas the Higgs is "an incomprehensible abstraction, a partial solution to an extraordinarily rarified and perhaps always-incomplete intellectual puzzle." Sarewitz concludes by suggesting that "whereas the Higgs discovery gives me no access to insight about the mystery of existence, a walk through the magnificent temples of Angkor offers a glimpse of the unknowable and the inexplicable beyond the world of our experience" (Sarewitz 2012, p. 431). The upshot is that there are some things science cannot, and perhaps should not, aim to provide – a personal encounter with the unknown, or insight into the mystery of existence, chief among them.

From Anecdote to Science

Considering these examples of people's reactions to science shifts us from the realm of philosophy to the realm of human psychology. Why might (some) people have the intuition that (some) aspects of human experience, such as color perception, romantic love, or transcendent awe, fall beyond the scope of scientific knowledge? What governs which phenomena are seen as falling beyond this scope, and which within? And do these views have implications for people's attitudes towards science or scientific explanations?

The perspectives voiced by Proxmire and Sarewitz could reflect a deep truth about what scientific knowledge can and cannot do. If they are right, then scientific knowledge has real limits that both producers and consumers of science should acknowledge. Perhaps there are important aspects of color perception or romantic love or religious experience that science cannot and should not explain. Understanding such limits would be important in directing the enterprise of

science, and in recognizing the complementary contributions of other human endeavors, be they poetry or religion.

On the other hand, it could be that these intuitions about the scope of science are deeply misguided. Perhaps science *can* explain all facets of human experience, and perhaps there are important benefits that would arise from its success. On this view, perspectives like Proxmire's could impede the production or uptake of scientific knowledge. If people oppose research on sexual desire and romantic love, for example, that could interfere with the development of interventions to improve relationships or resolve sexual dysfunction.

To evaluate the basis for these intuitions, it's important to move beyond anecdote to evidence. Instead of considering potentially unrepresentative perspectives voiced in popular media, we can turn to science itself. Specifically, what does psychological science tell us about the nature and sources of people's intuitions about the scope of scientific knowledge and the limits of scientific explanation?

We decided to find out. In a series of empirical studies (Gottlieb & Lombrozo 2018), we investigated whether people in fact find some phenomena – such as love or spirituality – less amendable to scientific explanation, and we evaluated several hypotheses about why this might be the case. In the rest of this chapter, we explain what we found. But first, we consider some hypotheses that motivated our approach.

Motivating Hypotheses: Intuitive Dualism and Our Creaturely Selves

Why might people be inclined to regard some aspects of human experience as falling beyond the scope of science? One hypothesis is that people are "intuitive dualists," on some level committed to the Cartesian idea that we have minds or "souls" that are wholly different from material bodies. This view is called *intuitive* dualism because the claim isn't that people have explicitly worked-out ideas about the way the mind and the body relate, the way Descartes did, but rather that on a more intuitive or gut level, they act as if minds and bodies are fundamentally different sorts of things – the former accessible through introspection and reasoning; the latter extended in space such that we can measure and prod.

With a view like this, it makes sense that science can only offer adequate explanations for our material bodies – for our headaches, but not for our heartaches. Developmental psychologist Paul Bloom, a prominent advocate for this hypothesis about human cognition, has argued that dualist tendencies are often at odds with what science has to tell us about the physical and mechanistic substrates of the mind (Bloom 2004). At the same time, these tendencies can also help explain the allure of neuroscientific explanations for the mind. "We intuitively think of ourselves as non-physical," writes Bloom, "and so it is a shock, and endlessly interesting, to see our brains at work in the act of thinking" (Bloom 2006).

One piece of evidence for people's "intuitive dualism" comes from a clever study by Preston, Ritter, and Hepler (2013). They had people read about the psychology of love, but only some people were given additional information about the *neuroscience* of love. They read, for example, that "the ventral tegmental area and the medial caudate nucleus, associated with other forms of reward and motivation, are activated when thinking about a romantic partner." After reading this information, participants in this latter group, compared to those in the former group, reported decreased belief in a human soul or spirit. This suggests that the notion of a human soul is, at least to some extent, believed to be at odds with a reductive, scientific understanding of the human mind. In explaining the *brain*, we can't also be explaining the soul. If anything, we are explaining it away.

If intuitive dualism underlies people's resistance to the idea that science can explain human experience, then we might expect such resistance to be particularly apparent for aspects of the mind or behavior typically associated with a soul. For example, we might be more reluctant to accept a scientific explanation for why people act altruistically or fall in love than for face recognition or forgetfulness.

A related hypothesis is that for some people, the idea of being "merely" creatures – nothing more than a part of the biological world – threatens a sense of what it means to be human. According to a psychological proposal known as "Terror Management Theory," humans' awareness of their own mortality can provoke great anxiety. As a means of assuaging this anxiety, individuals tend to respond to reminders of their own mortality by employing tactics that inhibit or ease these mortality-related thoughts (Pyszczynski, Solomon, & Greenberg 2003). For example, death-related thoughts have been shown to increase the extent to which people report having religious beliefs (including belief in an afterlife, which offers a kind of immortality), even among people who do not identify as strongly religious.

Based on these ideas, we hypothesized that some scientific explanations could be "terror"-inducing. Specifically, research has shown that emphasizing our animal nature – or our own "creatureliness," as it is called within this body of research – can be perceived as threatening because it reminds us of our own mortality, triggering the terror management of Terror Management Theory (Goldenberg, Pyszczynski, Greenberg, Solomon, Kluck, & Cornwell 2001). In particular, scientific explanations that account for human traits in physical and reductionist terms, or in a way that holds equally for other species, could be rejected in an effort to manage the existential threat that they induce. This rejection could be especially robust for aspects of the mind that are perceived to make humans special. For instance, people might be less inclined to accept scientific explanations for religiosity or language, which are often perceived to be uniquely human, than for aspects of depth perception or motor control, which we share with other species to a greater extent. We refer to this hypothesis as *human exceptionalism*.

Initial Evidence that Intuitive Dualism Guides Intuitions about the Scope of Science

The two hypotheses we've introduced – that people are (to some extent) intuitive dualists, and that people are (to some extent) threatened by humans' "creatureliness" – are empirical hypotheses about the human mind. Accordingly, we can use the methods of psychological science to test them, and that's precisely what we did (you can read about these findings in more detail in Gottlieb and Lombrozo 2018).

In our first study, we presented over 300 participants with a variety of psychological phenomena, including perceiving color, experiencing "love at first sight," and having a spiritually transformative experience. But we also included phenomena that we typically associate with lower-level perceptual or cognitive processes, such as remembering somebody's name, recognizing another person's face, and reaching for and grabbing an object. There were 48 phenomena in total, and for each one, we asked: to what extent do you agree that science will one day provide a complete explanation for this phenomenon? As we expected, people were much more likely to say that science could explain phenomena that we typically associate with lower-level perceptual or cognitive processes than phenomena such as experiencing "love at first sight," or having a spiritually transformative experience.

Our next goal was to understand *why*. If people are, as psychologists have theorized, intuitive dualists, and the relevant demarcation is between bodies and minds or souls, then we would expect phenomena that involve minds or souls to be the ones most often considered beyond the scope of science. But which phenomena are these? We expected them to be those that people seem to direct or assess with their minds – that is, those over which we think we have conscious control, and those that involve an experiential quality that we can access through introspection.

To test these ideas, we had the participants in our studies again look at the same list of 48 phenomena that we presented them initially, but this time we asked them to make two novel judgments. The first was about the extent to which they considered each to involve conscious will, or the ability to deliberately influence how, when, or why the phenomenon happens. Participants tended to rate phenomena like decision-making high in conscious will, but phenomena like dreaming low in conscious will. For the second judgment, we asked a question reminiscent of Jackson's point about what Mary learned when she escaped from her chamber: "To what extent does this involve a subjective experience (a feeling of what it is like) that only the individual experiencing it can know?" For this question, participants tended to give phenomena like falling in love and believing in God high ratings, as well as acting altruistically, feeling love toward one's children, and having a sense of personal identity that persists over time. Phenomena like perceiving depth and identifying sounds received much lower ratings.

Overall, participants' judgments were consistent with the predictions of intuitive dualism: they were more likely to say that a phenomenon was beyond the scope of science if they rated that experience as requiring conscious control and as having a personal, experiential component. Phenomena that were rated high on both dimensions – such as experiencing love towards one's children or acting altruistically – were thus among the most likely to be deemed resistant to a complete scientific explanation.

These initial results echo the intuition that is so often elicited by Jackson's thought experiment about Mary. There are some mental processes or experiences – such as seeing color – that have a first-personal, experiential quality associated with them, such that a purely scientific description seems to fall short. Our results also resonate with some of the public responses to science that we quoted above. Elaine Hatfield and her colleagues received pushback when they used scientific tools to study romantic love, which rates particularly highly in having a personal, experiential component. Daniel Sarewitz argued that the Angkor temples offer "an authentic personal encounter with the unknown" (Sarewitz 2012, p. 431) – the kind of spiritual experience that our participants similarly viewed as personal and experiential, and as more likely to fall beyond the scope of science compared with many of the other phenomena that we tested.

Digging Deeper: What's So Special about Personal, Subjective Experience?

As a next step, we decided to dig deeper into what it is about having an experiential, personal, and introspectively-accessible experience that leads people to judge a phenomenon beyond the scope of science. Recall that we asked participants to answer the following question: "To what extent does this involve a subjective experience (a feeling of what it is like) that only the individual experiencing it can know?" For a philosopher of mind, this single question packs in several potentially distinct components. First, there's the matter of a *subjective experience* – the "what it is like" to experience some phenomenon. Second, there's the idea that the experience is somehow personal and *privileged* – that only the individual experiencing it can know. And finally, there's the implication that the form of access to this experience is through *introspection*, an examination of one's own thoughts and feelings. Which of these components was driving participants to judge some phenomena beyond the scope of science? Or was it all three?

In a follow-up study, we teased apart these three components of our original question. We again had participants tell us the degree to which they thought that science could ever fully explain each of the mental phenomena from our initial studies, but this time they rated those same experiences on three additional dimensions: subjective experience ("This has a subjective experience associated with it: a 'feeling' of what it is like"), privileged access ("Only an individual him or herself can know that he or she is experiencing this; an outside observer might be

able to guess but can't truly know"), and introspection ("An individual having this experience can know he or she experiences it through introspection: the examination of one's own internal feelings or reflection").

In this follow-up study, we found that the latter two components – privileged access and introspection – were driving the intuition that some phenomena cannot be captured by a scientific explanation. In other words, the phenomena that participants rated as highly privileged ("only I could know!") and as accessible through introspection were the ones rated least likely to be fully explained by science. To illustrate with an example, this suggests that people are dissatisfied with scientific explanations because falling in love has an experiential quality to it that is accessible through introspection and *only* to the experiencer herself.

It is important to note that the three dimensions we were interested in here – subjective experience, privileged access, and introspection – are all highly related to one another. However, subjective experience – the dimension related to phenomenology, or the degree to which people feel that an experience has a distinctive feeling of what it "is like" – did not have a statistically significant association with scientific explanation judgments when we statistically controlled for the other two dimensions. This could be surprising in light of Jackson's example about Mary. What she seems to learn upon first seeing color is precisely the phenomenological component of color perception, the "what it is like." On the other hand, it makes sense that this experience is inaccessible to science, and that this is so precisely because of the way in which we access phenomenology ourselves: through the private process of introspection. An individual can introspect about her own experience; a scientist cannot do the introspecting for her, and Mary cannot "introspect" her way to the experience of seeing color by reading scientific papers or conducting research on the color perception of others.

A Role for Human Exceptionalism, Too

The evidence presented in the preceding two sections provides some support for our hypothesis about intuitive dualism: people are more resistant to the idea that science can explain a psychological phenomenon when that phenomenon is something that we take ourselves to control with our minds – through conscious will – or access with our minds – through private introspection. These are aspects of our *mental* experience – our minds, rather than our bodies.

What about our other hypothesis, linking scientific explanation to "creatureliness" and human exceptionalism? We were especially curious to test the idea that people are more resistant to scientific explanations for traits that are believed to be uniquely human, relative to those we share with other species. Recall that this hypothesis was motivated by the idea that reductionist or cross-species explanations for uniquely human traits could be threatening because they liken us to our animal relatives and remind us of our own mortality. To test this, we asked people to rate the very same list of 48 mental phenomena from the studies

already described, but this time we had them indicate whether they thought they were uniquely human, or present in other species as well. Things like falling in love, making moral judgments, and religion were considered uniquely human, but so were a variety of complex cognitive tasks, such as engaging one's imagination and thinking creatively. Other phenomena, such as dreaming and integrating sensory information to figure out where a sound is coming from, received low ratings for human uniqueness.

Consistent with our prediction, we found that the phenomena rated high on human uniqueness were also more likely to be judged beyond the scope of science. That is, a phenomenon related to hearing or seeing – examples of perceptual processes – was more likely to be judged amenable to a "complete scientific explanation" than a phenomenon like making moral judgments.

In a follow-up study, we dug deeper into the idea of human uniqueness by unpacking two separate components that could have contributed to people's judgments. We ultimately found that people were particularly resistant to the idea that science could explain things that contribute to making humans *exceptional* relative to other species. So it isn't just that imagination and creativity are present *only* in human minds (or that people believe them to be so), but that the ability to exhibit these characteristics is perceived to be part of what makes humans special. If this is correct, then fully explaining imagination or creativity (versus motor control or depth perception) in scientific terms could seem implausible because it's taken to imply a fully physical or reductionist account of the capacity. It folds us into the biological realm – a mere creature among many – and fails to set humans apart from other species. And if Terror Management Theory is right, being a mere creature is an uncomfortable reminder of our own mortality.

Our studies thus provide some support for two initial hypotheses – that people are (to some extent) intuitive dualists, and that scientific explanations are (on average) judged to be less likely for our less-creaturely, more uniquely human characteristics. But these are just two of many possible hypotheses. We also tested a third hypothesis: that people might treat *complexity* as a fundamental constraint on scientific knowledge. That is, people might consider something like romantic love to be beyond the scope of scientific explanation because it is perceived to be too complex, and in particular more complex than basic cognitive or perceptual processes. The most interesting thing about this hypothesis is that we robustly failed to find any support for it. We found that romantic love *was* considered highly complex, but so were things like logical reasoning and memory. More crucially, these complexity ratings were *unrelated* to judgments about the possibility of obtaining a complete scientific explanation for the corresponding phenomenon.

What *Should* Science Explain?

Summarizing the findings we've just described, our studies revealed that some phenomena are typically judged to resist a complete scientific explanation, and

that science is not perceived to be limited by the complexity of its subject matter, but instead by its third-personal and potentially reductive methodology. These findings speak to public skepticism about the idea that science could one day fully explain romantic love or transcendent awe. But they don't yet speak to another aspect of our introductory examples: the sense that there would be something *bad* about achieving a complete scientific explanation; that when it comes to some things, science *should* be limited. Recall Proxmire's admonition that "Americans *want* to leave some things in life a mystery" (emphasis added), and Sarewitz's implication that science shouldn't purport to offer more than it does – to quote the title of his piece, "sometimes science must give way to religion." Do most people share this sense that some scientific explanations are not only impossible, but also unwelcome?

To find out, each of our studies also asked people to tell us how *uncomfortable* they would be if science could fully explain the phenomenon in question. Interestingly, people were most uncomfortable with the idea that science could explain things like love, morality, or religious belief – the very same things they said that science could never *possibly* explain. Also mirroring our initial results, we found that people were uncomfortable with science explaining things they felt they could consciously will, and things that made humans exceptional compared to other species. Moreover, these judgments related to ratings of privileged access and introspection in exactly the same way as the scientific possibility questions did: people were most uncomfortable with the idea that science could fully explain the phenomena that they deemed knowable only by the experiencer herself, and those that supported introspective access.

Why do scientific explanations for some phenomena generate discomfort? It could be that intuitive dualism and creatureliness are at work once again, but this time manifesting in a more visceral form, making us *uncomfortable* with the very idea that science could succeed when it comes to explaining our uniquely human minds. But these findings also raise interesting questions about the source of people's beliefs concerning what science *should* or *should not* seek to explain – the sorts of beliefs that affect the research scientists choose to pursue, the projects that funding agencies choose to support, and the public's response to their efforts. Beyond practical considerations, are these beliefs about what science should and shouldn't pursue largely governed by the suite of epistemic and affective responses that our studies reveal? If so, our findings have important implications, as they uncover subtle aspects of human psychology that shape the course of science.

Some Open Questions

So far we've been talking about averages – how people respond, on average, to questions about whether science could possibly provide a full explanation for a given phenomenon, and about whether such an explanation would be

uncomfortable. Our studies also raise some questions about differences across individuals that warrant further investigation.

In our original studies, we surveyed a rather diverse online sample of individuals who had a range of educational backgrounds. Surprisingly, our findings replicated fully in a sample of undergraduate students who had taken, on average, a handful of psychology courses, suggesting that commitments about the appropriate scope of science are fairly stable despite modest scientific training. It remains unclear, however, whether professionally-trained cognitive scientists would demonstrate the same pattern of results.

That said, there did exist some variation among individuals in the degree to which they thought science could or should explain the mind. And although these differences did not differ systematically with education, they did correlate with political ideology and religiosity: people who thought that science cannot and should not explain aspects of the human mind were more likely to be politically conservative and religious. Was it conservatism and religiosity that led to intuitive dualism and exceptionalism, or the other way around? This is an important question for another day.

Another open question concerns the possibility that despite resistance to the *idea* of a complete scientific explanation, such an explanation – once offered – might actually be accepted, and even welcome. Recall psychologist Paul Bloom's observation that despite finding brain-based explanations for the mind unintuitive, we find them "endlessly interesting" (Bloom 2006).

In fact, there's evidence that people *like* explanations for psychological phenomena that appeal to neuroscience. Specifically, research has found that people are susceptible to what is called the "reductive allure" effect: they prefer explanations at lower levels (e.g., that appeal to neuroscience) to explanations at higher levels (e.g., that appeal only to psychology), even when the lower-level content does not offer additional explanatory information (Hopkins, Weisberg, & Taylor 2016). Take, for example, the psychological phenomenon known as "the other-race effect," which shows that people have difficulty telling two faces apart when those faces come from a race other than their own. In the "reductive allure" studies, participants were asked to evaluate a psychological explanation for the effect. For half of those participants, the psychological explanation did not appeal to neuroscience:

> In communities where the majority of the people are white, white faces are seen more frequently than are those of other races. This greater experience with white faces tunes the perceptual system to recognize greater detail across those faces, making it easier to tell them apart.

For the other half, the psychological explanation was more reductive in that it included additional neuroscientific information: "This greater experience with white faces tunes the fusiform face area to recognize greater detail across those faces, making it easier to tell them apart." Participants without relevant expertise

generally thought that the latter explanation was a better one, even though expert participants did not.

So could it be that our participants were simply wrong about their anticipated discomfort, and that scientific explanations for all mental phenomena will in fact be welcome as soon as they're on offer? We suspect not. When it comes to evaluating explanations for the types of phenomena that tend to fall beyond the scope of science – such as romantic love or religious experience – the allure of reduction could be offset by the allure of intuitive dualism, and by the repulsion of our "creaturely" selves. Our dualism and exceptionalism hypotheses would both predict that, despite the reductive allure, people will be more uncomfortable with explanations of love, for example, as those explanations become increasingly reductive. That is, people would be more uncomfortable with a chemical explanation than a neuroscientific one, and more uncomfortable with a neuroscientific explanation than a psychological one. This is a question for future research.

Implications for the *Real* Limits of Scientific Explanation

It's important to emphasize that the body of research we've been discussing reveals *intuitions* about what science can or cannot explain, and that it does not speak directly to what science can, in fact, explain. How, then, should these results be interpreted? On the one hand, it could be that people's intuitions track some epistemic truth about the limits of science. If these intuitions are correct, then despite methodological advances, science will never fully explain something like romantic love because of its rich, first-personal and uniquely human experiential quality. On the other hand, if these intuitions are instead misguided, they could prove to be serious barriers to scientific advance, leading people to have intuitive biases against scientific explanations in certain domains. Should this be the case, there is the concern that scientists could avoid or fail to receive support for research in areas that many consider outside the scope of science, even when that research could lead to important theoretical and practical advances.

Going one step further, if people falsely believe that a scientific perspective is not only *insufficient,* but also misplaced or even harmful, we could miss out on potentially important truths about ourselves and the world. For example, many believe that gene editing technologies, such as CRISPR-Cas9, hold the promise of transforming medicine by eliminating previously incurable diseases and disorders. In 2015, the journal *Science* referred to it as the "breakthrough of the year." But many surrounding discussions, including those by scientists working on such technologies, have focused on whether we *should* be implementing these technologies in the first place, and for what purposes. George Daley, a stem cell researcher and the dean of Harvard Medical School, remarked on a 2017 success in using gene-editing technology to alter viable human embryos, saying, "The question now remains should we – and for what purposes and should there be certain applications that are allowed and others that are prohibited?" (Maron 2017).

As society moves forward in both debating and embracing advances in gene editing, it will be important to query public opinion: which types of applications do people generally consider acceptable, and which types of applications do people generally consider unacceptable? Are people more opposed to intervening on some traits than others? These are open empirical questions, but it might be that people are more uneasy with gene editing when it plays upon some of our commitments about the limits of science – for example, when it aims to target traits perceived as uniquely human, or ones typically associated with a human soul or essential spirit. These are important questions to be addressing – and it's important to get the answers right, even when they might violate initial intuitions.

Thus returning from the realm of human psychology to the realm of philosophy, we can ask with more urgency: When it comes to the scope of science, are people's intuitions getting things right, or getting things wrong? We think the answer is "both."

Regarding intuitive commitments about introspection and privileged first-person access, we think people's intuitions might be onto something important. Perhaps there really are, in principle, certain aspects of experience that cannot be captured by scientific knowledge alone. This is one of the points illustrated by the case of our black-and-white Mary who has the scientific knowledge, but not the experiential knowledge, of seeing color. Science benefits from its objective, third-person methodology, and this methodology will one day allow us to explain why Mary does or doesn't have the experiences that she does. But these explanations will supply scientific knowledge, not personal experience.

Regarding intuitive commitments about human exceptionalism and our desire to be more than mere creatures, we're more inclined to dismiss intuition. If resistance to scientific explanations for uniquely human traits is motivated by mortality-related anxiety and existential threat, it's not clear whether or why these judgments might also track some epistemic truth. However, it could well be that scientific explanations fall short of providing everything we want. It's not that they fail to fully explain a phenomenon, but rather that they don't put it in a personal and cultural context that reflects its human significance. For that we may well benefit from the arts and humanities, from poetry and music. We agree with Daniel Sarewitz that "[t]he Higgs boson, and its role in providing a rational explanation for the Universe, is only part of the story." (Sarewitz 2012, p. 431).

Coda

We can now imagine the subject of a new thought experiment, Mary-Lou. Mary-Lou is a college student taking her first philosophy course. Her professor has just assigned Jackson's piece about Mary the color scientist and asks the students to ponder whether there are limits to scientific knowledge. That night, she sits in her dorm room reading about Mary stepping out of her black and white chamber for the first time. Consistent with the data we've presented, Mary-Lou thinks back to

her professor's question and intuits that scientific knowledge is limited in its scope. Does the case of Mary-Lou the philosophy student demonstrate that there *are* true limits to scientific knowledge? On its own, the answer is surely no: intuitions are not always correct. But intuitions are often a first step; armed with data and arguments, they can sometimes show us the way.

References

Bloom, P., 2004. *Descartes' baby: How the science of child development explains what makes us human*, New York: Basic Books.

Bloom, P., 2006. Seduced by the flickering lights of the brain. *Seed Magazine*, *27*.

Goldenberg, J.L., Pyszczynski, T., Greenberg, J., Solomon, S., Kluck, B. and Cornwell, R., 2001. I am not an animal: Mortality salience, disgust, and the denial of human creatureliness. *Journal of Experimental Psychology: General*, *130*(3), p. 427.

Gottlieb, S. and Lombrozo, T., 2018. Can science explain the human mind? Intuitive judgments about the limits of science. *Psychological Science*, *29*(1), pp. 121–130

Hatfield, E., 2006. The Golden Fleece Award: Love's labours almost lost. *APS Observer*, *19*(6): 16–17.

Hopkins, E.J., Weisberg, D.S. and Taylor, J.C.V. 2016, The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition*, 15, 67–76.

Jackson, F., 1986. What Mary didn't know. *Journal of Philosophy*, *83*(5), pp. 291–295.

Maron, D. 2017. Embryo gene-editing experiment reignites ethical debate. Scientific American, [online] Available at: www.scientificamerican.com/article/embryo-gene-editing-experiment-reignites-ethical-debate/ [Accessed 7 April 2018].

Preston, J.L., Ritter, R.S. and Hepler, J., 2013. Neuroscience and the soul: Competing explanations for the human experience. *Cognition*, *127*(1), pp. 31–37.

Pyszczynski, T., Solomon, S. and Greenberg, J., 2003. *In the wake of 9/11: Rising above the terror*. Washington, DC: American Psychological Association.

Sarewitz, D., 2012. Sometimes science must give way to religion. *Nature News*, *488*(7412), p. 431.

18

SHOULD WE ACCEPT SCIENTISM?

The Argument from Self-Referential Incoherence

Rik Peels

Introduction

An influential idea in science, philosophy, and popular science writing these days is that science and the natural sciences in particular always reliably lead to rational belief and knowledge, whereas non-scientific sources of belief never do. This view is sometimes referred to as 'scientism'. The word has often been used pejoratively, but, nowadays, the word is frequently adopted as a badge of honour: Alex Rosenberg (2011), Don Ross, James Ladyman and David Spurrett (2007), and others call themselves adherents of scientism and defend it in detail.

In this chapter, I discuss a specific argument against scientism. I call it the 'argument from self-referential incoherence'. The point of the argument is that scientism itself is not – and, I will argue *cannot* be – sufficiently supported merely by natural science and, therefore, scientism cannot be rationally believed or known. I also argue that this counts against scientism. It might seem obvious that it does, but this is an important additional argumentative step. If scientism cannot be rationally believed or known, it is *epistemically improper* to believe scientism, but scientism might still be *true*. I, therefore, also defend the view that, even though scientism's self-referential incoherence does not imply that it is false, it provides us with good reason to reject it. What I will argue implies that scientism is self-refuting. To say that it is self-referentially incoherent is to be more specific, though, for it draws attention to the fact that scientism is self-refuting partly in virtue of the fact that it (implicitly or explicitly) refers to itself.

The chapter is structured as follows. First, I explain in some more detail what scientism amounts to. Subsequently, I spell out the argument from self-referential incoherence. One might think that scientism is so strong a position that it *obviously* defeats itself. Remarkably, however, this is not the case; fairly sophisticated

responses to the argument are available for the adherent of scientism. I, therefore, discuss three responses that one might give to the argument. First, it may be argued that scientism itself *is* sufficiently supported by scientific evidence. Second, one could suggest that we can embrace scientism and simultaneously make an exception for scientism itself – that is, rationally believe it, even though it does not meet its own criteria. Third, one might propose that we should think of scientism as a thesis that is *pragmatically* rather than *epistemically* justified. I argue that each of these responses fails. I conclude that scientism is hoist by its own petard.

What Is Scientism?

Before I spell out the argument from self-referential incoherence, let us first consider scientism in some more detail. I take scientism to be a thesis that refers to the natural sciences, such as biology, chemistry, geology, and physics, because paradigm cases of scientism are theses that put the natural sciences centre stage rather than, say, history or psychology. In fact, we find in the literature statements to the effect that academic disciplines such as psychology and economics should adopt the methods of natural science or even be reduced to natural science in order to deliver epistemically rational belief or knowledge.[1] I focus on scientism as an *epistemological* rather than an *ontological* claim, that is, as a claim to the effect that only science delivers rational belief or knowledge rather than as the claim that what exists is only what science tells us exists or only that which can in principle be investigated by science.[2] Let me point out two distinctions that can be used to further specify the variety of scientism in question.

First, scientism can be understood as the claim that only natural science, for instance, (a) delivers, produces, leads to, or issues in – I use these terms equivalently – *rational belief*, (b) produces *knowledge*, or (c) *reliably* leads to rational belief or knowledge. These theses are conceptually distinct. One may take it, for instance, that non-scientific beliefs can still be rational or reasonable, but that they cannot constitute knowledge. Or one might think that non-scientific sources of belief *incidentally* rather than *reliably* produce knowledge. It seems that (c) is the strongest variety, whereas (a) is the weakest. For, if non-scientific sources cannot even produce rational belief, then surely they cannot lead to knowledge or reliably deliver rational belief, since, on virtually all philosophical views,[3] knowledge presupposes rational belief. I confine myself mostly to those versions of scientism that say that only natural science delivers rational belief, but I will at some places in this chapter also take into account versions that say that only natural science delivers knowledge.

One might think that the latter is an implausibly strong view. Do the humanities, such as history, for instance, not deliver knowledge? Surprisingly, various adherents of scientism *do* indeed embrace such a strong view. Alex Rosenberg is quite explicit that the humanities do *not* deliver any knowledge:

> When it comes to real understanding, the humanities are nothing we have to take seriously, except as symptoms. But they are everything we need to take seriously when it comes to entertainment, enjoyment, and psychological satisfaction. Just don't treat them as knowledge or wisdom. (2011, p. 307)

Some others do not explicitly use the word 'knowledge' or the phrase 'rational belief', but make claims that are conceptually highly similar to this and that can easily be understood along these lines. According to Daniel Dennett, for instance, "when it comes to fact, and explanations of facts, science is the only game in town."[4] Some might be willing to count, say, philosophy among the sciences, but many adherents of scientism *expressis verbis* reject this option. The renowned physicist Stephen Hawking famously declared at the 2011 Google Zeitgeist Conference that "philosophy is dead" and that "scientists have become the bearers of the torch of discovery in our quest for knowledge."[5]

Of course, there are also academic disciplines that count neither as humanities nor as natural sciences, such as social science and economics. Some adherents of scientism are explicit that even those sciences do not deliver knowledge. According to E.O. Wilson (1975, p. 4), "[i]t may not be too much to say that sociology and the other social sciences, as well as the humanities, are the last branches of biology waiting to be included in the Modern Synthesis." His idea seems to be that *all* academic disciplines should be reduced to the natural sciences, especially to biology. Francis Crick (1966) claims that everything can be explained by physics and chemistry, and Alex Rosenberg (2011) defends the view that physics is the whole truth about reality.

Some other adherents of scientism *do* believe that other sciences than the natural sciences deliver knowledge. Therefore, towards the end of this chapter, I return to the question of whether broadening the notion of 'science', so that it includes, say, social science and economics, helps to refute the accusation that scientism is self-referentially incoherent.

The second dimension along which varieties of scientism could be distinguished concerns the non-scientific sources of belief that are discarded. There is, of course, a wide variety of such sources: vision, taste, smell, hearing, and touch (the five senses), memory, introspection, metaphysical intuition, logical intuition, mathematical intuition, linguistic intuition, and so forth. Stronger versions of scientism will discard *all* these non-scientific sources of belief, whereas weaker versions will discard only *some* of them.[6] Like Otto Neurath (1987), Don Ross, James Ladyman and David Spurrett (2007, p. vii, 65), adopt a *weaker* version of scientism when they claim that "analytic metaphysics (…) contributes nothing to human knowledge", whereas science does. Another weak version is embraced by Eric Schwitzgebel (2011), who claims that introspection is untrustworthy.[7] Still others make a much more general claim, though. According to Alex Rosenberg, for instance, scientism

> (…) is the conviction that the methods of science are the only reliable ways to secure knowledge of anything; that science's description of the world is correct in its fundamentals (…) Science provides all the significant truths about reality, and knowing such truths is what real understanding is all about. (…) Being scientistic just means treating science as our exclusive guide to reality, to nature – both our own nature and everything else's. (2011, pp. 6–8)[8]

Henceforth, I focus on the stronger version of scientism that says that *in any domain* of reality only natural science delivers rational belief or knowledge. In the final section of this chapter, I show what our discussion entails for scientistic claims about particular domains, such as metaphysical intuition and introspection.

The Argument from Self-Referential Incoherence

I take it that a thesis is *self-referentially incoherent* if and only if it somehow explicitly or implicitly refers to itself and the thesis – sometimes in conjunction with one or several plausible principles – is incoherent at least partly in virtue of the fact that it refers to itself.[9] I focus on what I call *epistemic* self-referential incoherence. Examples are the claim that no proposition can be known and the claim that any belief formed upon considering this proposition is irrational. These propositions are self-referentially incoherent, because they respectively implicitly and explicitly refer to themselves and – in conjunction with a plausible principle about knowledge or rationality – are incoherent. If no proposition can be known, then that proposition cannot be known either. And if any belief formed upon considering this proposition cannot be rationally believed, then it cannot be rationally believed. I say 'in conjunction with a plausible principle about knowledge or rationality' because these propositions are incoherent only if we add the premise that these propositions themselves can be respectively known and rationally believed – which is a premise that one seems committed to if one believes them. Below, I return to the issue of which epistemic principle makes scientism self-referentially incoherent and why we should think the adherent of scientism is committed to that epistemic principle.

It seems that the argument from self-referential incoherence against scientism would say that, since on scientism no proposition can be rationally believed unless it is based on natural scientific research, scientism itself cannot be rationally believed, because it is *not* based on scientific research. We find rough and sketchy versions of this argument in the literature. According to Jeroen de Ridder, for instance:

> scientism suffers from self-referential problems. Not being a scientific claim itself, it would seem scientism cannot be known by anyone. This raises the question of why anyone should assert or believe it in the first place. (2014, p. 27)

And according to Mikael Stenmark in his book on scientism:

> The most troublesome difficulty with T1 [a variety of epistemological scientism; RP], however, is that it appears to be *self-refuting*, that is, T1 seems to tell us not to accept T1. This is a very serious problem for the defenders of Scientism, because if T1 is self-refuting then it is not even possible for T1 to be true. (2001, p. 32)

Earlier on, he was slightly more detailed about this objection:

> (…) how do you set up a scientific experiment to demonstrate that science or a particular scientific method gives an exhaustive account of reality? I cannot see how this could be done in a non-question begging way. What we want to know is whether science sets the limits for reality. The problem is that since we can only obtain knowledge about reality by means of scientific methods (that is T1), we must use those methods whose scope is in question to determine the scope of these very same methods. If we used *non*-scientific methods we could never come to *know* the answer to our question, because there is according to scientistic faith no knowledge outside science. We are therefore forced to admit either that we cannot avoid arguing in a circle or that the acceptance of T1 is a matter of superstition or blind faith. (2001, pp. 22–23)

Now, before we try to spell out the argument more formally, let me make two preliminary remarks.

First, the argument can be cast in terms of knowledge, rationality, justification, warrant, understanding, or other epistemic desiderata, since scientism itself can be spelled out in each of these terms. In this chapter, I largely confine myself to the argument cashed out in terms of *rational belief*. The arguments can easily be revised in order to draw conclusions about, say, knowledge or understanding. As I wrote above, I take scientism to be the thesis that we can rationally believe a proposition *p* only if our belief that *p* is based merely on scientific research.

Second, there are different ways to structure the argument. I present one variety of the argument which is a *reductio ad absurdum*.

Argument

(1) Scientism is true. [Assumption for *reductio*]
(2) If scientism is true, we can, merely on the basis of scientific research, rationally believe that it is true. [Premise]
(3) We can, merely on the basis of scientific research, rationally believe that scientism is true. [from (1), (2)]
(4) It is impossible to rationally believe merely on the basis of scientific research that scientism is true. [Premise]

(5) It is possible *and* it is impossible to rationally believe merely on the basis of scientific research that scientism is true. [Conjunction of (3), (4); Reductio ad absurdum]
(6) Conclusion: (1) is false

Let me say a bit in defence of premises (2) and (4). As to (2), the adherent of scientism seems committed to this premise, because she claims (asserts) that scientism is true. Here is one reason to think that a person who asserts scientism is committed to (2). An idea that is widely advocated among philosophers these days is that knowledge is the norm of assertion: one should assert that *p* only if one knows that *p*.[10] And knowledge implies rational belief. Therefore, one should assert that *p* only if one rationally believes that *p*. Even if one does not accept the knowledge norm of assertion, though, there is good reason to embrace (2). This is because all (2) says is that if scientism is true, we *can* rationally believe that it is true, not that we actually *do* rationally believe that it is true. And it seems undeniable that one should *not* assert something if there is good reason to think that one *cannot* even rationally believe it.

Premise (4) says that we cannot rationally believe scientism on the basis of scientific research. The motivation for (4) is rather simple: scientism is not some empirical truth that we can find out by way of setting up an experiment. Nor does it seem to be an *a priori* truth that can be deduced by mathematical or logical methods from elementary truths that we know *a priori*. Rather, it seems to be an *epistemic* principle that needs to be backed up by philosophical argumentation. And whatever philosophy is, it is widely considered *not* to be one of the natural sciences. Peter Atkins (1995) boldly claims that there appear to be no boundaries to the competence of science. But if the above argument from self-referential incoherence against scientism is sound, there is at least one boundary to the competence of science: science is incompetent to provide sufficient scientific support for making belief in scientism rational.

If this argument is convincing, then this leaves the adherent of scientism with three options:

A. Premise (4) is false, because we *do* or at least *can* rationally believe scientism on the basis of scientific research.
B. We *can* rationally believe scientism, even though *not* on the basis of scientific research. Scientism itself is an exception to scientism. This would amount to a slight, albeit important revision of scientism.
C. We do *not* rationally believe scientism, but we should nevertheless accept it for pragmatic reasons. This amounts to rejecting premise (2).

Below, in Sections 4–6, I argue that each of these options is wanting. I conclude that we ought to reject scientism.

First Response: Should We Believe Scientism on the Basis of Scientific Inquiry?

A first response to the argument from self-referential incoherence is that we *do* or at least *can* have scientific evidence for scientism. It is undeniable that science has an impressive track record. We have discovered all sorts of things about the cosmos, about space and time, about life, about ourselves. One might think that this provides some kind of *inductive* argument for scientism in the sense that each discovery provides us with rational belief and knowledge and that all these cases together, therefore, provide us with good reason to think that only natural science provides rational belief or knowledge. It is not that scientism can be *deduced* from the results of natural science or that it is the *best explanation* for a series of phenomena that we encounter, but rather that even the comparatively short history of science with its impressive successes gives us good reason to think that scientism is true, and that even if the evidence is not yet sufficient, that at some point it may very well be if science continues to be as successful as it has been so far or if it becomes even more successful.

At least two comments on this response are in order. First, even if natural science's track record were impeccable and continued to be so indefinitely while the body of scientific knowledge continually expands, that would in no way give a good reason to think scientism is true. It would justify at most the claim that *if* something is the result of natural science, then we have good reason to think that it is rational to believe that particular result, *not* that it is rational to believe something *only if* it is based on science. For, when one derives from the fact that if something is an established result of natural science then it is rational to believe it that if one rationally believes something, one must do so on the basis of natural science, one commits the logical fallacy of affirming the consequent.

What we would need as well, of course, in order to make scientism plausible, is evidence for the unreliability of non-scientific sources of belief. Note that evidence for the thesis that non-scientific sources of belief are *less* reliable than scientific sources of belief will not do. For, even if they are less reliable, it does not follow that their deliverances do not amount to rational beliefs. Thus, we would need good empirical arguments to think that, say, metaphysical intuition, introspection, and memory are so unreliable that we cannot rationally embrace their deliverances and that beliefs from these sources do not count as rational beliefs. We can find such arguments in the literature, for example in the writings of Daniel Dennett (1991, 2003) and Eric Schwitzgebel (2011), but the arguments these authors adduce in favour of their radical theses are highly controversial.[11] Of course, natural science *could* in principle at some point come up with convincing arguments for the unreliability of, say, introspection or, at least, for the unreliability of the introspection of certain *kinds of* mental states. In order for scientism to be tenable, though, we would need good reason to discard *all* non-scientific sources of belief, and it is not at all clear that we could ever have good reason to do so.

Second, imagine that we had good reason to think that scientific research would always (or often enough) issue in rational belief and that non-scientific sources of belief always (or often enough) deliver irrational belief or at least not rational belief, for instance, because we have good scientific empirical evidence to think that non-scientific sources of belief are unreliable. We can imagine (possibly, per impossibile), for instance, that we have good empirical reasons to think that introspection, memory, and logical reasoning are unreliable. That would still leave us with the question of *how* we could *rationally believe* scientism itself. Presumably, in order to *rationally believe* scientism, it would have to be a scientific hypothesis that has been tested and confirmed sufficiently frequently.

Now, we should note that if scientism is a scientific hypothesis, the fact that it is self-referential is as such *not* a problem. The sentence "This sentence contains English words" also refers to itself, but it seems nonetheless true. Thus, even though scientism may implicitly refer to itself, this alone does not make it self-referentially incoherent.

The problem is rather that if scientism is a scientific hypothesis that has been empirically confirmed by testing cases of beliefs based on science and beliefs from non-scientific sources, we still need an answer to the question of *how* we know in each particular case that it is an instance of rational belief or that it is *not*. It seems that one's verdict in each case will depend on one's theory of rationality, such as whether or not it requires evidence that is accessible to the subject, whether a belief can be rational merely in virtue of being undefeated, and so forth. And, clearly, whether or not one takes each of these to be criteria of rational belief is *not* a matter that science can establish. What is relevant here is *epistemic* intuitions (or epistemic beliefs) and epistemological arguments on the basis of those intuitions. Thus, the inductive argument for the scientific hypothesis of scientism will get started only if from the very beginning we assume that certain beliefs from non-scientific sources that we hold are instances of rational belief.

One may reply that there is a large movement in epistemology that pleads for a *naturalization* of epistemological questions. And one might suggest that this implies that we can do epistemology without any epistemic intuitions or philosophical arguments. My reply is twofold. First, most adherents of naturalized epistemology, such as Robert Almeder (1998) and Richard Fumerton (1994), argue that epistemology needs to be *empirically informed* in order to answer epistemological questions, *not* that epistemic intuitions and epistemological arguments are superfluous. Second, those who embrace the more extreme versions of naturalized epistemology, such as W.V.O. Quine (1969), typically claim that natural science should take over answering questions about the causal connections between our sensory evidence and our beliefs about the world and that questions about what it is for something to be epistemically rational or to count as knowledge should be abandoned, *not* that natural science can give us answers to questions about epistemic rationality.[12] But if it cannot give such answers, then we have no reason to think that natural science can tell us when it is rational to adopt a belief and when it is not.

This means that the argument from self-referential incoherence against scientism stands unscathed: according to scientism only those propositions supported by natural science can be rationally believed, but scientism itself cannot be sufficiently supported by natural science and, therefore, cannot be rationally believed. The thesis of scientism, therefore, implies that it cannot be rationally believed.

Second Response: Should We Make an Exception for Scientism?

A second line of response is that we can rationally believe some proposition *p* only if *p* is the result of science *or* if *p* is the thesis of scientism itself. Scientism would, thus, be an exception among the propositions that can be rationally believed: it can be rationally believed, even though it is *not* the result of scientific research.

The main problem with this kind of reply is, of course, that it seems unduly *ad hoc*: what is so special about scientism that we can rationally believe a proposition only if it is the result of scientific research unless it is the thesis of scientism itself? Scientism is a claim about rational belief and if it is allowed in, why would other epistemological claims or, for that matter, metaphysical or ethical claims not count as rational? The restriction to all views except scientism itself seems arbitrary.

One might reply that it is not unreasonable to make an exception for scientism itself, since one *has to* make an exception for *any* epistemological theory in order to avoid a regress. Some beliefs will simply have to be accepted as rational, even if they are not based on arguments. For three reasons, however, this response is unconvincing as it stands.

First, it is controversial that there are properly basic beliefs, that is, that some beliefs are rational even if they are not in any way supported by one's other beliefs. Adherents of *coherentism* and *foundherentism* (rather than foundationalism) deny this.[13] I do not intend to suggest that some kind of foundationalism, which entails that there are properly basic beliefs, is false. Rather, I would like to point out that one should not simply assume the truth of foundationalism or some other kind of epistemological theory that implies that there are properly basic beliefs without some kind of argument.

Second, even if it were true that a theory about, say, rationality, has to make an exception for itself, we have not been given a reason to embrace scientism rather than a rival theory of rationality. One could equally well embrace a theory that says, for instance, that a belief is rational if one has no good reason to think that it is false or unreliably produced, and that it amounts to knowledge if there is no such reason *and* it is reliably produced.

Third and most importantly, it is simply false that a theory about, say, rational belief or knowledge has to make an exception for itself. Take a foundationalist theory that says that certain of our beliefs are properly basic, for example, when they are reliably produced by a properly functioning mechanism that aims at truth, and that some of our beliefs based on linguistic, epistemic, and metaphysical

intuitions meet this criterion. One might then also claim that one *knows* this particular theory about knowledge on the basis of one's properly basic beliefs about particular cases of belief. That theory would meet its own criteria and would, thus, *not* have to make an exception for itself.

One may reply that *certain kinds* of epistemological theses *have to* make an exception for themselves. According to Adam Elga (2010), any consumer-rating magazine that would *not* rate itself as the No. 1 consumer-rating magazine would be inconsistent, for if it is *not* No. 1, the reader has insufficient reason to trust the consumer-ratings found in the magazine. Consumer-ratings magazines, therefore, have to be dogmatic about the correctness of the epistemic advice they give. This reply is important, for epistemological scientism gives epistemic advice, and if one epistemological theory that gives epistemic advice can properly be dogmatic about its own correctness, then why could another one not be dogmatic?

The analogy that Elga gives, however, fails for at least two reasons. First, a consumer-ratings magazine need *not* be dogmatic with respect to its own correctness. If consumer ratings show that it is *not* the best consumer-ratings magazine, the magazine could be simply be stopped. Or it could be continued. After all, its results might still be entirely correct – consumer-ratings simply say what consumers prefer. And even if they are not *entirely* correct (correct about everything), they might still give good advice in many cases and, therefore, be sufficiently reliable. Second, and more importantly, the consumer-ratings magazine does *not* formulate a general rule while dictating that it is itself an exception to that rule, whereas scientism, on the response under consideration, does so. Maybe advice needs to be dogmatic in some sense with regard to its own correctness. It does not follow that views that give advice but make an exception for themselves are *not* unduly ad hoc.

Third Response: Is Scientism Pragmatically Justified?

A third response grants that we cannot *rationally believe* scientism, but claims that we should nonetheless adopt it, because it is *pragmatically* justified: working with it – that is, believing it and acting on that belief – gives such good results that we should embrace it, even if we cannot rationally believe it. This means that one would either irrationally *believe* scientism or – for all we know, rationally – *accept* scientism, that is, work with scientism, adopt it as a policy without believing it, merely assume it for the sake of argument.[14]

It seems to me, though, that the idea that scientism is pragmatically justified suffers from at least two problems that *are* fatal. First, imagine that we did *not* accept scientism, assume it, or work with it, but that we *did* assume, accept, or work with a somewhat different thesis, namely the rather *un*controversial thesis that natural science leads to all sorts of rational beliefs. It seems that an acceptance or assumption along *those* lines would have the exact same good results as accepting or assuming scientism. We can have the same observations, experiments, inductions,

abductions, deductions, theories, models, and so forth when we reject scientism. Thus, even though natural science has indeed been impressively successful, that provides us with no good pragmatic reasons to embrace scientism rather than an epistemological thesis that ascribes a positive epistemic status *both* to the deliverances of natural science *and* to beliefs from non-scientific sources.

Now, one could, of course, reply that what I have pointed out is compatible with the idea that *both* views – scientism and the view that natural science leads to all sorts of rational beliefs and knowledge – are pragmatically justified. In that case, scientism would still be pragmatically justified. This is, of course, true, but the problem is that if both views are pragmatically justified, then, *ceteris paribus*, we have no reason to prefer scientism over the rival view. Scientism would, thereby, lose its bite, since it would then be arbitrary whether one adopts scientism or some rival view.

Another response to this objection is that scientism and the more modest idea that natural science leads to all sorts of rational beliefs might equally *lead to the acquisition of true* beliefs, but that scientism has the additional advantage that it also *avoids* or *helps to abandon* false beliefs because it discards as unreliable non-scientific sources of belief. If that were true, then, one might think, scientism would be *more instrumental* in reaching the twofold goal of believing truths and not believing falsehoods than certain rival views. The problem is that this *might be* the case, but that it might equally be the case that if we adopt scientism, we *abandon all sorts of true beliefs* that we would hold if we rejected scientism. All depends on how convincing the arguments regarding the (un)reliability of specific sources of beliefs, such as the introspection of phenomenal states, are going to be and, as I pointed out above, such arguments are highly controversial.

Second, if we were to embrace scientism merely for pragmatic reasons, we would *realize* that we have done so and our having done so would, therefore, fail to make a difference to which beliefs we hold – except for such trivial beliefs as the belief that we have adopted scientism for pragmatic reasons. If we only *assume for the sake of argument* or *act as if* certain beliefs from non-scientific sources are not rational, we will automatically continue to hold them, since that as such does not change the evidential basis for those beliefs.[15] For example, if, on the basis of introspection, I hold certain beliefs about the phenomenal states that I am in, or if I hold certain metaphysical beliefs, and then assume *merely for the sake of argument* that those beliefs are not rational, I can reach a certain conclusion about those arguments. But since I know I have reached that conclusion merely for the sake of argument and without any change in my evidence for these beliefs, I will, inevitably, continue to hold these introspective and metaphysical beliefs. We will, normally, only abandon beliefs if we *actually take ourselves to have* good reason to think – rather than assuming for the sake of argument – that those beliefs are false, irrational, unreliably formed, or some such thing. That would make scientism pointless, for the very idea of scientism is that we should hold only those beliefs that are based on natural scientific inquiry. Of course, if scientism were not only

pragmatically justified, but also *epistemically justified* because we have good reason to think that it is true, then that *would* probably lead us to abandon many of our beliefs, since we would then come to believe that they are not rational. However, that would also lead us back to the problems discussed in the two previous sections, so that we would still face the argument from self-referential incoherence.

Conclusion

In this chapter I have done a bit of philosophical judo: I have employed scientism's own weight against it. I have argued that scientism – the idea that only science delivers rational belief – is self-referentially incoherent, where arguments to that effect can be phrased as *reductios*. I also argued that the three main options that seem available to the adherent of scientism all fail: that on which we can rationally believe scientism to be true on a scientific basis, that on which scientism is an exception to scientism, and that on which scientism is pragmatically justified.

If what I have argued is correct, scientism will be tenable only in a substantially weaker variety which says that certain epistemic beliefs – beliefs about rationality and about knowledge – are rational, as well as certain linguistic and epistemic intuitions that are needed to back up one's scientism by argument. This is unavoidable, but deeply problematic for scientism for at least two reasons. First, scientism would have to count as rational certain beliefs that are not even remotely based on science, such as the view that a belief is justified only if it is produced by a sufficiently reliable process or the belief that a belief is rational only if it fits one's evidence. Surely, this goes against the spirit of scientism. Second, if linguistic and epistemic beliefs are allowed in, then exactly why should other beliefs, such as metaphysical beliefs, be excluded, that is, discarded as being irrational? Scientism, then, should not only be cast as a significantly weaker claim than it usually is. It should also be accompanied by a criterion and a defence of that criterion that is different from the thesis of scientism itself. This is needed in order to exclude belief sources that are in many ways similar to those sources of belief that are needed to get scientism started in the first place if it is to avoid the argument from self-referential incoherence.

Acknowledgements

For their helpful comments on earlier versions of this chapter, I would like to thank Lieke Asma, Jeroen de Ridder, Søren Harnow, Miriam Kyselo, Ard Louis, Kelvin McQueen, Nikolaj Nottelmann, Chris Oldfield, Esben Nedenskov Peterson, Pieter van der Kolk, Hans van Eyghen, René van Woudenberg, and (other) audience members of presentations at the University of Southern Denmark and the annual Dutch Research School of Philosophy conference at the Vrije Universiteit Amsterdam, as well as the two editors of this volume, Kevin McCain and Kostas Kampourakis. Publication of this chapter was made possible through the support

of a grant from the Templeton World Charity Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Templeton World Charity Foundation.

Notes

1 As regards psychology, see, for instance, Dennett (1991) and (2003). The debate about the methods of economy has been raging for decades; for several references, see Hayek (1979).
2 For more on the relation between the two, see Peels (2019).
3 There are a few exceptions; e.g., Lasonen-Aarnio (2010).
4 Interview by Sholto Byrnes in the *New Statesman*, April 10, 2006.
5 See Matt Warman, "Stephen Hawking Tells Google 'Philosophy Is Dead'", *The Telegraph*, May 11, 2011. He makes the same point in almost the same words in Hawking and Mlodinow (2010, p. 5).
6 One may, of course, wonder whether science is even possible without employing these sources of belief. This issue is up for debate and involves a variety of challenging issues; since it is a complex topic and since I have already discussed it in detail elsewhere (see Peels 2018), I will leave it aside here.
7 For a similar position about the untrustworthiness of introspection, see Dennett (1991) and (2003).
8 For a similar claim, see Atkins (1995).
9 It seems to me that this squares well with how 'self-referential incoherence' is usually defined; see, for instance, Boyle (1972, p. 25); Mavrodes (1985).
10 See, for instance, Benton (2011); Smithies (2012); Turri (2011).
11 See, for instance, many of the essays in Jack and Roepstorff (2003) and (2004). See also my criticisms in Peels (2016).
12 Kim (1988, p. 390), has made this point in much more detail.
13 See, for instance, Haack (2009).
14 For a detailed account of the distinction between belief and acceptance, see Cohen (1992).
15 As I have argued elsewhere, this is a general problem with all belief-policies that are *not* themselves beliefs (see Peels 2013).

References

Almeder, R. (1998). *Harmless Naturalism: The Limits of Science and the Nature of Philosophy*. Peru, IL: Open Court.
Atkins, P.W. (1995). Science as Truth. *History of the Human Sciences*, 8, pp. 97–102.
Benton, M.A. (2011). Two More for the Knowledge Account of Assertion. *Analysis*, 71, pp. 684–687.
Boyle Jr., J.M. (1972). Self-Referential Inconsistency, Inevitable Falsity, and Metaphysical Argumentation. *Metaphilosophy*, 72, pp. 25–42.
Cohen, L.J. (1992). *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.
Crick, F. (1966). *Of Molecules and Men*. Seattle: University of Washington Press.
Dennett, D.C. (1991). *Consciousness Explained*. London: Penguin Press.

Dennett, D.C. (2003). Who's on First? Heterophenomenology Explained. *Journal of Consciousness Studies*, 10, pp. 19–30.

De Ridder, J. (2014). Science and Scientism in Popular Science Writing. *Social Epistemology Review and Reply Collective*, 3, pp. 23–39.

Elga, A. (2010). How to Agree about How to Disagree. In R. Feldman and Ted A. Warfield, eds., *Disagreement*. Oxford: Oxford University Press, pp. 175–186.

Fumerton, R. (1994). Skepticism and Naturalistic Epistemology. *Midwest Studies in Philosophy*, XIX, pp. 321–340.

Haack, S. (2009). *Evidence and Inquiry: A Pragmatist Reconstruction of Epistemology*. 2nd ed. Amherst, NY: Prometheus Books.

Hawking, S, and L. Mlodinow. (2010). *The Grand Design*. New York: Bantam Books.

Hayek, F.A. (1979). *The Counter-Revolution of Science: Studies on the Abuse of Reason*. Indianapolis: Liberty Fund.

Jack, A.I., and A. Roepstorff, eds. (2003). Trusting the Subject: The Use of Introspective Evidence in Cognitive Science. *Journal of Consciousness Studies*, 1, p. 10.

Jack, A.I., and A. Roepstorff, eds. (2004). Trusting the Subject: The Use of Introspective Evidence in Cognitive Science. *Journal of Consciousness Studies*, 2, 11, pp. v–xxii.

Kim, J. (1988). What is Naturalized Epistemology? In J.E. Tomberlin, ed., *Philosophical Perspectives*, 2. Atascadero, CA: Ridgeview, pp. 381–406.

Lasonen-Aarnio, M. (2010). Unreasonable Knowledge. *Philosophical Perspectives*, 24, pp. 1–21.

Mavrodes, G. I. (1985). Self-Referential Incoherence. *American Philosophical Quarterly*, 22, pp. 65–72.

Neurath, O. (1987). Unified Science and Psychology. Originally published in 1932, in B. McGuinness, ed., *Unified Science*. Dordrecht: Kluwer, pp. 1–23.

Peels, R. (2013). Belief-Policies Cannot Ground Doxastic Responsibility. *Erkenntnis*, 78, pp. 561–569.

Peels, R. (2016). The Empirical Case against Introspection. *Philosophical Studies*, 173, 2461–2485.

Peels, R. (2018). The Fundamental Argument against Scientism. In M. Boudry and M. Pigliucci, eds., *Science Unlimited? The Challenges of Scientism*. Chicago: Chicago University Press, pp. 165–184.

Peels, R. (2019). A Conceptual Map of Scientism. In J. de Ridder, R. Peels, and R. van Woudenberg, eds., *Scientism: Prospects and Problems*. New York: Oxford University Press, pp. 28–56.

Quine, W.V.O. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Rosenberg, A. (2011). *The Atheist's Guide to Reality: Enjoying Life without Illusions*. New York: W.W. Norton.

Ross, D., J. Ladyman, and D. Spurrett. (2007). In Defence of Scientism. In J. Ladyman, D. Ross with D. Spurrett and J. Collier, *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press, pp. 1–65.

Schwitzgebel, E. (2011). *Perplexities of Consciousness*, Cambridge, MA: MIT Press.

Smithies, D. (2012). The Normative Role of Knowledge. *Noûs*, 46, pp. 265–288.

Stenmark, M. (2001). *Scientism: Science, Ethics and Religion*, Aldershot: Ashgate.

Turri, J. (2011). The Express Knowledge Account of Assertion. *Australasian Journal of Philosophy*, 89, pp. 37–45.

Wilson, E.O. (1975). *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press.

19

HOW ARE THE UNCERTAINTIES IN SCIENTIFIC KNOWLEDGE REPRESENTED IN THE PUBLIC SPHERE?

The Genetics of Intelligence as a Case Study

Kostas Kampourakis

Introduction

Evelyn Fox Keller (2000) famously described the 20th century as the century of the gene. The gene concept was coined in the beginning of that century by Wilhelm Johannsen who noted that "The word gene is completely free from any hypothesis; it only expresses the established fact, that at least many properties of an organism are conditioned by special, separable and thus independent 'conditions', 'foundations', 'dispositions'" (translated in Roll-Hansen, 2014, p. 4). From this view, by the end of the 20th century, Keller noted, the gene had ended up being considered "… the guarantor of intergenerational stability, the factor responsible for individual traits, and, at the same time, the agent directing the organism's development" (Keller, 2000, pp.144–145). In popular culture, the powers of the gene reached another dimension, as Dorothy Nelkin and Susan Lindee (2004, p.16) noted: "Clearly the gene of popular culture is not a biological entity. … The gene is, rather, a symbol, a metaphor, a convenient way to define personhood identity, and relationships in social meaningful ways." The subsequent research in genomics since the 1990s until today has produced an enormous understanding of the complexities of genome structure, regulation and development, but at the same time it has shown that whereas genes are certainly important, they are nevertheless not the mythical, powerful entities that we might have thought them to be (Kampourakis, 2017).

Nevertheless, in the public sphere the concept of gene seems to retain its mythical powers. A quick internet search reveals several examples. For instance, a 2014 article in the *Guardian* was titled "'Happy gene' may increase chances of romantic relationships."[1] The title of a 2015 article in the *New York Times* suggested that "Infidelity lurks in your genes."[2] And there is more. Several authors have argued for

the pervasiveness of such views in the public sphere (see, e.g., Hubbard and Wald, 1997; Nelkin and Lindee, 2004; Heine 2017). One major issue concerning genes and their public representation is what I have called *genetic fatalism* (Kampourakis 2018, p.18). This conception comprises three others, which are often conflated: genetic essentialism, the idea that genes inside us specify who we are; genetic determinism, the idea that this is done regardless of the environment; and genetic reductionism, the idea that if we want to understand why we are the way we are, we have to study our genes (see Kampourakis, 2017, p. 6). Therefore, media messages like those above might have a large impact, as recent research has shown that notions of genetic fatalism, especially essentialism, align well with core human intuitions (Heine, 2017). One might wonder why, given the current state of knowledge about genes and genomes, the mythical view of genes as being all-powerful entities has not ceased to exist.

Dorothy Nelkin has provided a detailed account of how scientific findings, and their impact, have been represented in the media during the 20th century. A major conclusion she reached was that whereas the media can play an important role in enhancing the public understanding of science, and indeed there exist examples of science reporting that is thoughtful and accurate, they have nevertheless often failed to achieve this:

> "… too often science in the press is more a subject for consumption than for public scrutiny, more a source of entertainment than for information. Too often science is presented as an arcane activity outside and above the sphere of normal human understanding, and therefore beyond our control.
> (Nelkin, 1995, p.162)

Nelkin also noted a key difference between the habits of mind of scientists and journalists. Whereas scientists consider new research findings as tentative and provisional, such findings are exactly what journalists find newsworthy; established research is for them old news and consequently less interesting (p.165). Therefore, the same results can be perceived and represented differently. As a result, one might think that accounts of the mythical powers of genes are simply bad journalism; the people who attempted to translate the original scientific reports, either exaggerated or misunderstood the original findings.

However, it isn't necessarily so. A detailed study of public representations of genes in the popular press has shown that the discourse about genes around the end of the 20th century has not been more deterministic than before, if determinism is broadly defined as "the assignment of exclusive influence over human outcomes to genes". Condit has actually concluded that "In most periods, most sources have concluded that genes and the environment interact to produce human characteristics" (Condit, 1999, p.210). Studies on how the media represent genetic research have also shown that, overall, they do not make exaggerated claims. For instance, a study of 627 newspaper articles published in Canada, the United States,

the United Kingdom and Australia found that only 11% of them had moderately to highly exaggerated claims. These newspaper articles reported on 111 articles published in scientific journals. The majority of the newspaper articles made no claims (63%) or slightly exaggerated claims (26%). An interesting finding was also that only 15% of the newspaper articles and 5% of the scientific articles discussed costs or risks of the research, with the vast majority of them discussing its benefits (Bubela & Caulfield, 2004). This study was generally in agreement with previously published research on the topic.

Perhaps then the hype we sometimes see in the representation of genetics research is not due to bad reporting. An interesting view is that "… the spectacle of science is not simply an epiphenomenal artefact, tacked on to 'real science', but is, rather, part of the epistemic core of scientific cultures and scientific work" (Steinberg, 2015, pp.2–3). This simply means that when science becomes a spectacle, as in the case of genes acquiring mythical powers, it is not only because of an external misinterpretation, but also because of features inherent in the scientific knowledge itself. In many cases, the mythical powers of genes and the accuracy of DNA analyses are taken for granted in media representations, without explicit discussions of the uncertainties involved. For instance, the relation between a particular DNA sequence X and a particular disease D is probabilistic. This means that whereas more people with X are expected to exhibit D compared to people who do not have X, there will certainly exist people with X who do not have D as well people with D who do not have X. Relations between traits and diseases are probabilistic, not deterministic.

This chapter is devoted to a qualitative analysis of how scientific findings are communicated to the public by the media. In particular, I am looking at how the findings of an article about the genetic basis of human intelligence published in the prestigious journal *Nature Genetics* (Sniekers et al., 2017) were reported in internet media reports, in particular in the online editions of widely read newspapers, as well in news-devoted websites. My aim is to explore the different ways in which the original message stated in the scientific article was transmitted, translated or distorted while being communicated to the broader public. The chapter begins with a description of the findings of the original article, continues with a description of the various reports, and concludes with implications of representing scientific knowledge in the public sphere. I must note that this is only a case study that aims at highlighting potential problems with the public representation of scientific research and arriving at some general conclusions and recommendations.

A Case Study: A 2017 Meta-Analysis on the Relation between Genes and Intelligence

On May 22, 2017, an article was published on the website of the journal *Nature Genetics* with the title: "Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence" (Sniekers

et al., 2017). *Nature Genetics* is a prestigious scientific journal. According to its aims and scope:

> *Nature Genetics* publishes the very highest quality research in genetics. It encompasses genetic and functional genomic studies on human and plant traits and on other model organisms. Current emphasis is on the genetic basis for common and complex diseases and on the functional mechanism, architecture and evolution of gene networks, studied by experimental perturbation.[3]

The impact of the journal is considered to be very high, having a 5-year impact factor of 32.197 (for comparison, a very good journal in the humanities would have an impact factor of less than 3). The impact factor is certainly only one among several different metrics used to evaluate scientific research (see Chapter 10 for details), and one might even question its value. However, impact factors presently matter a lot if only because scientists strive to have their findings published in high impact factor journals and because these are the journals that their colleagues mostly read.

This article reported a meta-analysis of data from genome wide associated studies (GWAS). Such studies look for associations between specific sites on DNA and specific conditions (traits or diseases). To achieve this, GWAS use dense maps of variations at the level of single nucleotides (single nucleotide polymorphisms or SNPs) that cover the human genome, in order to look for differences in the frequencies of alleles (different versions of the same gene) between different groups of individuals, for instance people with a certain condition of interest and people without that condition. The basic assumption is that a significant difference in the frequency of a variant between these two groups, for instance that people with a condition are significantly more likely to have a SNP than people without the condition, would indicate that the corresponding region of the genome contains DNA sequences that somehow affect that condition (Kampourakis, 2017, pp.115–119). Sniekers et al. (2017) used data from previously published studies and unpublished data for 78,308 unrelated individuals (19,509 were younger than 18 years old and 58,799 were between 18 and 78 years old) from Australia, the Netherlands, Sweden, the United Kingdom, and the United States. They performed an analysis that included 12,104,294 SNPs. The results of the GWAS analysis supported the conclusion that 22 genes were implicated in intelligence. The researchers also noted that their "… calculations show that the current results explain up to 4.8% of the variance in intelligence …" (Sniekers et al., 2017, p.1109).

There are two important points that must be clarified before we proceed. The first one is how intelligence was conceptualized in this study. In general, intelligence is a concept that is far from simple to define, as different kinds of intelligence can be perceived in different cultures (Sternberg, 2004; Cocodia, 2014). In December 13, 1994, a statement on intelligence was published in the *Wall Street*

Journal, under the title "Mainstream Science on Intelligence", signed by 52 prominent experts on intelligence (reprinted in Gottfredson, 1997, p.13):

1. Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – "catching on," "making sense" of things, or "figuring out" what to do.
2. Intelligence, so defined, can be measured, and intelligence tests measure it well. They are among the most accurate (in technical terms, reliable and valid) of all psychological tests and assessments. They do not measure creativity, character, personality, or other important differences among individuals, nor are they intended to.

The researchers in the Sniekers et al. (2017), study followed this definition and used data on intelligence acquired with a variety of measures. In particular, these data either stemmed from calculating Spearman's *g* (general factor of intelligence) or other measures of intelligence (WISC-II, Moray House Test No. 12, Multidimensional Aptitude battery – Full-scale IQ, SON-R, Composite IQ score, WISC-R, WAIS-R, Fluid intelligence – touchscreen, Fluid intelligence – web-based) that are known to correlate highly with *g* (p.1107). The details are impossible to present here, but in what follows I assume that intelligence as defined above was measured in a reliable and valid manner in this study, because my focus is on its public representation.

The second important point is the exact meaning of the statement that the current results explain up to 4.8% of the variance in intelligence. This statement means that the researchers were able to explain 4.8% of the differences in intelligence among the particular people studied based on the genetic differences among them (that is, the differences among the DNA sites studied). In other words, 4.8% of the differences in intelligence as measured among these particular people could be attributed to differences in the genes studied, whereas the remaining 95.2% of the differences in intelligence among them should be attributed to other factors. This does not in any way mean that intelligence does not have a genetic basis; however, it does mean that more genes than those considered in the meta-analysis and other non-genetic, such as environmental, factors are implicated. If people differ from one another in intelligence, a very small part of this difference is due to differences in the genes analyzed in this study.

The researchers also employed another method, called a genome-wide gene association analysis (GWGAS). This method is based on converging evidence from multiple genetic variants in the same gene. As a result, it can yield novel genome-wide significant signals, on a gene-based level, which are not necessarily picked up by a standard GWAS that looks for variants across the genome. The GWGAS

method identified 47 genes associated with intelligence, whereas the GWAS analysis had identified 22 genes, as mentioned above. However, 17 of these genes were overlapping, that is they were identified by both methods. As a result, the total number of genes associated with intelligence that were identified in this meta-analysis were 22 (GWAS) + 47 (GWGAS) – 17 (overlapping) = 52 genes. From these genes, 12 had been previously identified in older studies, whereas 40 were new. Interestingly, among the 47 genes identified by the GWGAS to be associated with intelligence, 15 were also found to be associated with educational attainment. Among these genes, there were four for which strong associations were found. Three of these genes are involved in neuronal function: *SHANK3* in synapse formation; *DCC* in axon guidance; and *ZFHX3* in myogenic and neuronal differentiation. The fourth gene, *BMPR2*, is involved in embryogenesis and bone formation, and has also been linked to pulmonary arterial hypertension (Sniekers, 2017, p.1110).

Overall, the researchers made an important step forward in understanding the genetic background of intelligence, by identifying several new genes somehow associated with it. But they also explicitly stated that the identified genes only explain a small amount of the observed variation. They explicitly acknowledged the modesty of their findings, by concluding the article with the following sentence: "These findings provide starting points for understanding the molecular neurobiological mechanisms underlying intelligence, one of the most investigated traits in humans." (p.1112). This important research marks the beginning of understanding the genetics of intelligence, according to the researchers themselves. The question then becomes: how was this message communicated to the public? Did the various media succeed in transmitting an accurate message?

The Media Reports of the 2017 Genome-Wide Association Meta-Analysis

The Criteria Used for the Analysis of the Media Reports

The Sniekers et al. (2017) article was published online on May 22, 2017. In the days that followed, the news was widely discussed on the Internet. In an attempt to examine how these reports represented the findings and their significance, I used the phrase "genes intelligence May 2017" for a Google search, which resulted in 30 reports, published between May 22, 2017, and October 16, 2017. These reports were published in a variety of media, from widely read periodicals such as the *New York Times* and *The Guardian*, to specialized websites reporting on science. In order to evaluate these media representations, I analyzed their texts with very specific criteria in mind:

1. Whether the concept of intelligence was defined or clarified. Given that it is far from simple to define intelligence, it is useful to see whether the reports

defined, or at least clarified, the main concept of interest. Sniekers et al. (2017) did not provide a definition for intelligence, but cited articles that provided a detailed discussion. In my analysis I am looking for an explicit definition / implicit reference to the concept, or lack thereof.

2. Whether the fact that the sample consisted of people of European ancestry only was mentioned and whether the respective limitations were discussed. Conclusions from studies that find associations in particular populations are valid for those populations only and cannot be extrapolated to other populations that might have a different genetic constitution. Therefore, this is a limitation of the study that should be explicitly discussed. It should be noted that Sniekers et al. (2017) were not explicit about this in their article; however, given that *Nature Genetics* is a journal read by experts, this should be self-evident to them, and therefore it makes sense that the authors simply mentioned it and did not provide a detailed explanation.
3. Whether the small amount of variation in intelligence explained by those genes was mentioned, and whether the implications of this were explained. Even though 52 genes might sound like a lot, they could only explain a small amount of variation in intelligence, estimated at 4.8%. This clearly suggests that several other (genetic and non-genetic) factors are implicated in intelligence. Sniekers et al. (2017) were explicit about this.
4. Whether the message that this is the beginning of our understanding of the genetics underlying intelligence was conveyed. Sniekers et al. (2017) were explicit that this is just the beginning of our understanding of the genetics of intelligence.
5. Whether there was an explicit statement against the idea of the genetic determinism of intelligence. Given the complexity of this trait, and the fact that several genetic and non-genetic factors are implicated, reporters should alert their readers about this fact, refrain from using simplistic determinist language, and – even better – explain this complexity.
6. Whether there was a reference or link to the original Sniekers et al. (2017) in *Nature Genetics*. This allows readers to directly access the original article, even though one would need an institutional subscription or to make a payment in order to read the full text.
7. Whether the reporters asked the opinion of experts who did not participate in the Sniekers et al. (2017) study. Asking external experts to evaluate a study is a good way to provide a more balanced view of the study rather than simply relying on what the researchers themselves have stated.

Main Findings from Analysis of the 30 Media Reports

Table 19.1 presents an overview of the analyzed reports, in chronological order (day of appearance). Overall, some of the reports are a lot better than others because they delve into the details of the findings and their implications, as well

TABLE 19.1 The extent to which each of the 30 analyzed media reports fulfilled the criteria set (✔: fulfilled: ✘: failed; E: explicit; I: implicit; AGD: against genetic determinism; COM: complexity).

	Media – article title/author (date) URL in endnote	*Intelligence explicitly defined OR implicitly clarified*	*European origin mentioned AND limitations explained*	*Small amount of variation mentioned AND explained*	*Beginning of understanding of intelligence*	*Statement against genetic determinism OR for complexity*	*References to the NG study*	*Comments from experts not involved in the NG study*
1.	The New York Times – *In 'Enormous Success,' Scientists Tie 52 Genes to Human Intelligence* / Carl Zimmer (May 22, 2017)[4]	✔ (E)	✔ / ✔	✔ / ✔	✔	✔ (AGD)	✔	✔
2.	The Guardian – *Scientists identify 40 genes that shed new light on biology of intelligence* / Ian Sample (May 22, 2017)[5]	✘	✘	✔ / ✔	✔	✔ (AGD)	✔	✔
3.	NBC News – Forty More Genes for Intelligence Discovered / Maggie Fox (May 22, 2017)[6]	✘	✔ / ✘	✔ / ✔	✔	✔ (COM)	✔	✘
4.	NEWSWEEK – *Scientists Discover Over 50 New Genes Linked to Intelligence Levels* / Hannah Osborne (May 22 2017)[7]	✘	✘	✔ / ✔	✔	✔ (AGD)	✔	✘
5.	Daily Mail online –*'Smart genes' account for 20% of our intelligence: 40 newly found genes suggests smart people are tall, thin and unlikely to smoke* / Shivali Best (May 22, 2017)[8]	✔ (I)	✘	✔ / ✘	✔	✘	✔	✘
6.	INTERNATIONAL BUSINESS TIMES – *Is intelligence genetic? 40 genes linked to IQ discovered* / Martha Henriques (May 22, 2017)[9]	✘	✔ / ✘	✘	✘	✔ (COM)	✔	✘

(*continued*)

TABLE 19.1 (Cont.)

	Media – article title/author (date) URL in endnote	*Intelligence explicitly defined OR implicitly clarified*	*European origin mentioned AND limitations explained*	*Small amount of variation mentioned AND explained*	*Beginning of understanding of intelligence*	*Statement against genetic determinism OR for complexity*	*References to the NG study*	*Comments from experts not involved in the NG study*
7.	METRO – *Newly discovered 'intelligence genes' could be the reason you're so smart* / Fiona Parker (May 22, May)[10]	✘	✘	✔ / ✘	✔	✘	✘	✘
8.	Science News (Magazine of the Society for Science and the Public)– *40 more 'intelligence' genes found* / Laura Sanders (May 22, 2017)[11]	✘	✘	✔ / ✔	✔	✔ (COM)	✔	✔
9.	Science Alert – *Scientists Have Identified 40 New Genes Linked to Intelligence* / Mike McRae (May 22, 2017)[12]	✘	✔ / ✘	✔ / ✘	✔	✔ (COM)	✔	✘
10.	RT News – *Scientists discover intelligence linked to 52 'smart genes'* / (May 23, 2017)[13]	✘	✔ / ✘	✔ / ✘	✔	✔ (AGD)	✔	✘
11.	Inc. – *Scientists Find 52 Genes That Are Directly Linked to Intelligence.* / Minda Zetlin (May 23, 2017)[14]	✘	✔ / ✔	✔ / ✘	✔	✔ (AGD)	✘	✘
12.	The Japan Times – *'Smart genes' account for 20% of intelligence: study* / AFP-JIJI (May 23, 2017)[15]	✘	✔ / ✘	✘	✔	✔ (COM)	✔	✘
13.	TECH TIMES – *Science Stumbles On 40 New Genes Linked To Intelligence* / Katrina Pascual (23 May 2017)[16]	✘	✔ / ✘	✔ / ✘	✔	✔ (AGD)	✔	✘
14.	PHYS.ORG – *Large study uncovers genes linked to intelligence* / Raffaele Ferrari (May 23, 2017)[17]	✔ (E)	✔ / ✘	✘	✔	✔ (COM)	✔	✘

15.	Science Daily – *New genetic roots for intelligence discovered* / author unidentified, source cited: Vrije Universiteit Amsterdam (May 23, 2017)[18]	✘	✘	✔ / ✘	✔	✘	✔	✘
16.	The Scientist – *Smarty Genes: Scientists have identified 40 new genes linked to human intelligence.* / Ashley P. Taylor (May 23, 2017)[19]	✘	✘	✔ / ✔	✔	✘	✔	✔
17.	NEW ATLAS – 52 genes associated with intelligence discovered / Rich Haridy (May 23rd, 2017)[20]	✔ (I)	✘	✔ / ✘	✔	✘	✔	✘
18.	THE NATION – *'Smart genes' account for 20pc of intelligence* / (May 24, 2017)[21]	✘	✔ / ✘	✘	✘	✘	✔	✘
19.	Live Science –*Your Intelligence Genes: 52 and Counting* / Stephanie Pappas (May 24, 2017)[22]	✘	✔ / ✘	✔ / ✘	✔	✔ (COM)	✔	✔
20.	SCIENCE WORLD REPORT – *Scientists Identify 52 Genes Linked to Intelligence* / Elaine Hannah (May 24, 2017)[23]	✘	✔ / ✘	✔ / ✘	✔	✘	✔	✘
21.	THE CUT – *Yes, There Is a Genetic Component to Intelligence* / Jesse Singal (May 25, 2017)[24]	✘	✘	✔ / ✔	✔	✔ (AGD)	✔	✘
22.	PLOS BLOGS – *Six things we learned from that massive new study of intelligence genes* / Tabitha Powledge (May 26, 2017)[25]	✔ (I)	✔ / ✔	✔ / ✘	✔	✔ (COM)	✘	✔
23.	ALZFORUM – *Massive GWAS Reveals 40 New "Intelligence" Genes* / Marina Chicurel (May 26, 2017)[26]	✔ (I)	✔ / ✘	✘	✔	✘	✔	✔
24.	The Science Times – *Around 80,000 People Lead to Identify 40 Intelligence Genes* / Jaden Jane (May 27, 2017)[27]	✔ (E)	✔ / ✘	✘	✘	✔ (COM)	✔	✘

(*continued*)

TABLE 19.1 (Cont.)

	Media – article title/author (date) URL in endnote	*Intelligence explicitly defined OR implicitly clarified*	*European origin mentioned AND limitations explained*	*Small amount of variation mentioned AND explained*	*Beginning of understanding of intelligence*	*Statement against genetic determinism OR for complexity*	*References to the NG study*	*Comments from experts not involved in the NG study*
25.	QUARTZ – *An inconvenient truth: A massive new study lays out the map of our genetic intelligence* / Olivia Goldhill (May 30, 2017)[28]	✔ (I)	✘	✔ / ✘	✔	✔ (AGD)	✔	✘
26.	BioNews – *Forty new genes linked to intelligence in humans* / Annabel Slater (May 30, 2017)[29]	✘	✔ / ✘	✔ / ✘	✔	✔ (COM)	✔	✘
27.	CBC – *We've found 50 genes for intelligence. Could that lead to discrimination?* / (June 03, 2017)[30]	✘	✘	✔ / ✘	✔	✘	✔	✘
28.	VOX – Scientists are finding more genes linked to IQ. This doesn't mean we can predict intelligence. / Brian Resnick (Jun 6, 2017)[31]	✘	✔ / ✔	✔ / ✘	✔	✔ (AGD)	✔	✘
29.	World Economic Forum – *Scientists just found 40 new genes that affect your IQ* / Callum Brodie (June 22, 2017)[32]	✘	✘	✔ / ✘	✔	✔ (AGD)	✘	✘
30.	BUSINESS INSIDER UK – Scientists discovered people who are highly-intelligent have 52 genes in common / Cheng Cheng and David Anderson (October 16, 2017)[33]	✔ (I)	✘	✔ / ✘	✔	✔ (AGD)	✘	✘

as because they explicitly explain the limitations and the problems of this particular study. It is important to note that the principal investigator of the Sniekers et al. (2017) study, Daniele Posthuma is quoted in many of these reports. This contributes significantly to the accuracy of the respective reports as she is explicit about the limitations and the implications of the findings of the study. In what follows I describe how many of the reports fulfilled the above criteria, and I provide representative quotes (the numbers within brackets refer to the number of the respective report in Table 19.1).

First of all, only 9 out of the 30 reports somehow referred to a definition of intelligence. Among those, only 3 reports provided explicit definitions, making reference to a "mental ability" [1]; "the ability to learn, understand or deal with new situations" or "the ability to apply knowledge to manipulate one's environment or to think abstractly" [14]; or "Intelligence is referred as the ability to learn, to understand and to deal up with new situations. Somehow, it is defined as the ability to apply the knowledge learned to think abstractly and manipulate an environment." [24]. All in all, in a report about scientific findings it is important to be precise about what one is talking about. What intelligence is might seem self-evident, but as this meaning varies among cultures it is necessary to specify what the discussion is about. Therefore, a definition of general intelligence and *g*, as well as of how it is measured, would have been useful.

Regarding the fact that the sample consisted of people of European ancestry only, 17 out of the 30 reports mentioned this fact but only 4 discussed this limitation. In particular, these reports noted that: "But other gene studies have shown that variants in one population can fail to predict what people are like in other populations. Different variants turn out to be important in different groups, and this may well be the case with intelligence." [1]; "For one thing, the researchers chose to limit their study to people of European descent because the same genes sometimes have different effects in people from different ethnic groups." [11]; "…all 78,308 study subjects were of European descent … Patterns of intelligence genes will probably be at least somewhat different in other ethnic groups." [22]; "The gene variations that produce the differences between Europeans aren't necessarily the same variations that produce differences among groups of different ancestry. So if you were to test the DNA of someone of African origin, and saw they lacked these genes, it would be incredibly irresponsible to conclude they had a lower capacity for intelligence."[28].

This is a very important point because whatever can be concluded about the correlation of the 52 genes and intelligence, and the possible impact of the former on the latter, is valid for this particular population only, and likely for people of similar ancestry, but not people of all ancestries. Therefore, it is noteworthy that only 4 reports mentioned this limitation, as these findings are not generalizable. It should be noted that "European" in this study was used to denote people from Australia, the Netherlands, Sweden, the United Kingdom, and the United States, not all European countries or nations. Of course, and this is a sensitive

issue, in many respects the genetic differences among humans are not really many. Therefore, the main point here is not that Africans or Asians have a significantly different genetic constitution from Europeans. However, there are differences in gene frequencies among different populations that sometimes are important for medical purposes. The main point is that findings from one sample as in the Sniekers (2017) are not automatically generalizable.

Another important point is the very small amount of variation (4.8%) in intelligence that the identified 52 genes could explain. As many as 24 out of the 30 reports mentioned this fact; however, only 7 of them actually explained what this means and what it entails. This was explained in various ways: "each variant raises or lowers I.Q. by only a small fraction of a point" [1]; "… most [genes] contributing only a minuscule amount to a person's cognitive prowess." [2]; "And although 52 genes sounds like a lot, they only explain a small part of the differences in intelligence between one person and another." [3]; "Following the study, the team used the findings to try to predict intelligence in another, independent sample. "The prediction estimate was only good for five percent of the variants in that sample …" "[4]; "Together, the genetic variants identified in the GWAS account for only about 5 percent of individual differences in intelligence, the authors estimate. That means that the results, if confirmed, would explain only a very small part of why some people are more intelligent than others." [8]; "After the study, the researchers tried to predict intelligence in an independent group of study participants based on the 52 genes they'd identified, and they were right only 5 percent of the time, *Newsweek* reported." [16] (in this case, the author simply repeated the conclusion made by another author already quoted earlier); "That means about 95 percent of the intelligence differences in these samples, at least as measured in this manner, did *not* come down to the genes the researchers examined, which leaves plenty of room for those concerned with environmental influences." [21].

This is another very important point that should have been explicitly discussed in all reports. If 52 genes can explain less than 5% of the variation in the intelligence measured in the particular sample, there should exist hundreds of additional genes that explain the remaining 45% of variation in intelligence, assuming that about 50% of this variation is explained by the variation in genes and another 50% by the variation in the environment in human populations (Plomin and Stumm, 2018). The most important implication of this is that a lot more research is required. It must be noted that all but three reports make this very crucial point. What is even more important is that the amount of variation in intelligence that genes explain, often described as heritability, is something very specific to a particular population and to a particular environment. Furthermore, heritability does not represent how genetic a particular characteristic is. To give a classic example, all humans have two arms, and of course there exist genes that are implicated in the development of our legs and arms. That we have arms and not wings is of course due to particular DNA sequences that we have. Differences in DNA that cause differences in the number of arms in humans are rare. Therefore, most of

these differences are explained in terms of differences in the environment (e.g., accidents leading to amputation, environmental influences during development, etc.). As a result, having two arms is a trait that has a very low heritability in human populations, even though there is definitely a strong genetic basis (Kampourakis, 2017, pp. 194–202).

Given that the 52 identified genes could only explain 4.8% of the variation in intelligence in a very specific population, it would have been desirable for the authors of the media reports to also take an explicit stance against naïve ideas of genetic determinism, that is that there is one or a few genes that make some people more intelligent than others. Whereas there is no question that intelligence has a genetic basis, its inheritance must be very complicated if hundreds or, perhaps thousands of genes, are implicated. This means that in contrast to the model of Mendelian genetics, still widely taught at secondary schools, readers need to understand the complexities and the probabilistic character of intelligence. To give another example for comparison, we now know from various studies that approximately 80% of the variation in height among individuals within a population is due to genetic factors. This means that height is more "inheritable" than intelligence, as differences in DNA can explain about 80% of differences in height and about 50% of differences in intelligence. Yet, even in that case we have not gone far in identifying the implicated genes; for instance, a study that used GWAS data from 253,288 individuals and identified 697 SNPs in 423 loci, explained 16% of the observed variation in height (Wood et al., 2014). This means that another 64% of the variation in height remains to be explained and that numerous genes should be implicated. Imagine then the difficulty in figuring out the causes of intelligence in which environmental factors seem to have as an important of a contribution as genes do.

However, only 11 reports included an explicit statement against genetic determinism (having the notation AGD in Table 19.1); another 10 simply highlighted the complexity of the phenomena underlying intelligence (with the notation COM in Table 19.1); and the remaining 9 reports did not explicitly address this issue. Given that the power of genes seems to be quite intuitive to non-experts, all reports should have explicit statements against genetic determinism, like these: "These genes do not determine intelligence, however." "Hundreds of other studies have come to the same conclusion, showing a clear genetic influence on intelligence. But that doesn't mean that intelligence is determined by genes alone. Our environment exerts its own effects, only some of which scientists understand well." [1]; "… scientists generally agree that a large proportion of intelligence is inherited—and therefore based on genetic factors. But intelligence is not a straightforward trait influenced by just a few genes. Rather, there are hundreds of genes that make up a complex web, the vast majority of which research has yet to identify. Environmental factors can influence intelligence too, including education and upbringing." [4]; "Environment is as important a factor in intelligence as genetics, the scientists note, and we already know quite a bit about how

a person's environment can affect his or her intelligence." [11]. It is important to refrain both from exaggerating the importance of genes and from downplaying it. As one author nicely put it, "just as it's important not to slip into naïve blank-slate-ism here, it's equally important not to fall for overzealous strands of genetic determinism, or the idea that genes are destiny."[21] But given the focus of the reports on genes, which often intuitively seem to be the causes *by default*, non-expert readers need to be reminded that when it comes to intelligence both genes and environment are important.

A related concern has to do with the titles of the reports. Certain titles were more accurate than others in conveying the message and the main conclusion of the study. Titles such as "In 'enormous success,' scientists tie 52 genes to human intelligence" [1], "Scientists identify 40 genes that shed new light on biology of intelligence" [2], "Scientists discover over 50 new genes linked to intelligence levels" [4], or "52 genes associated with intelligence discovered" [17], accurately convey the message that those genes are somehow related to intelligence, without implicitly exaggerating their impact. In contrast, titles such as "Newly discovered 'intelligence genes' could be the reason you're so smart" [7], "Scientists discover intelligence linked to 52 'smart genes'" [10], "Your Intelligence Genes: 52 and Counting" [19], or "Around 80,000 people lead to identify 40 intelligence genes" [24] might be perceived to imply that genes alone determine how intelligent we are. Correlation and causation are different, therefore reports should have clearly conveyed that the researchers in Sniekers et al (2017) found evidence for the former and not for the latter.

Finally, whereas 25 out of the 30 reports referred to the original article published in *Nature Genetics*, several of them referred to other sources as well. It is of course not clear whether the authors of these reports read the original article in detail, or simply reproduced the comments of the scientists involved in the study, or those of other reports. I should note that journalists, even if they have a background in science, might find the original article hard to understand as it is published in a very specialized scientific journal. Yet, this is no excuse in my view for disseminating messages and conclusions one does not understand. Several of the reporters quoted Danielle Posthuma, the principal investigator of the study, who conveyed very accurate messages. However, what most authors of the analyzed reports did not do, and what only 7 of them did, is to contact or quote experts who were not involved in the study. This can be journalists' way of implementing a peer review process; readers should be given the opportunity to draw their own conclusions by reading a report of what was done, what the scientists involved think about this, as well as what experts who did not participate in the study think.

Conclusions

Whereas the scientists involved in the Sniekers et al. (2017) study were (understandably) eager to report their contribution to improving our understanding

of the genetics of intelligence by identifying 40 new implicated genes to their peers, the media reports have the very different role of informing the public. The scientists publishing their work in *Nature Genetics* are writing for a very specialized audience who could look into the cited paper about a definition of intelligence, who understood what the fact that the sample was of European ancestry entailed for the conclusions of the study, that explaining 4.8% of the variation in intelligence is better than before but still low, and therefore a lot of research remains to be done. But none of this is necessarily self-evident for non-expert readers. Therefore, the journalists reporting the findings should have taken caution to explain all these, as was exemplarily done in the *New York Times* report by Carl Zimmer [1]. It must be noted that by examining the news articles in detail, not just looking for overt evidence of genetic fatalism, I have found that these news articles still contribute to notions of genetic fatalism. The reason for this is that these news articles should explicitly address important aspects of the NG study (Table 19.1) whereas they did not; in contrast, by not doing so, they convey implicit messages that might reinforce, instead of counteracting, notions of genetic fatalism.

It thus becomes clear that journalists with a background in science *can* very effectively communicate the news to the public. In my view, the most important issue in accurately communicating the findings of scientific research to the public is not only the research findings but also what the limitations and the resulting uncertainties of the research are. Therefore, the message of the reports should not have been that several genes tied to intelligence were identified but rather that some genes, among a lot more, were identified; the reports should have noted that the identified genes only explained a small amount of variation in intelligence; that the results are not necessarily generalizable to all human populations. Uncertainty is a major feature of scientific research (Kampourakis and McCain, 2019). Therefore, the public should become aware not only of what we know but also of what we still have to find out. To give but one example, a recent meta-analysis (Savage et al., 2018), again led by Danielle Posthuma, with data from 269,867 participants, identified 205 associated genomic loci, of which 190 are new, and 1,016 genes, of which 939 new. Scientists may be on the right path but there is still a long way to go, and the public must understand this.

Acknowledgements

Many thanks to David Kirby and Kevin McCain for useful comments on earlier versions of this chapter.

Notes

1 www.theguardian.com/science/2014/nov/20/happy-gene-romantic-relationships erotonin-romance

2 www.nytimes.com/2015/05/24/opinion/sunday/infidelity-lurks-in-your-genes.html?partner=rss&emc=rss
3 www.nature.com/ng/about/aims
4 www.nytimes.com/2017/05/22/science/52-genes-human-intelligence.html
5 www.theguardian.com/science/2017/may/22/scientists-uncover-40-genes-iq-einstein-genius
6 www.nbcnews.com/health/health-news/forty-more-genes-intelligence-discovered-n763071
7 www.newsweek.com/intelligence-genes-discovered-scientists-iq-clever-inherited-613348
8 www.dailymail.co.uk/sciencetech/article-4530428/Scientists-identify-FORTY-new-intelligence-genes.html
9 www.ibtimes.co.uk/intelligence-genetic-40-genes-linked-iq-discovered-1622846
10 metro.co.uk/2017/05/22/newly-discovered-intelligence-genes-could-be-the-reason-youre-so-smart-6654048/
11 www.sciencenews.org/article/40-more-intelligence-genes-found
12 www.sciencealert.com/researchers-discover-40-new-genes-connected-to-intelligence
13 www.rt.com/usa/389314-scientists-smart-genes-intelligence/
14 www.inc.com/minda-zetlin/scientists-find-52-genes-directly-linked-to-intelligence-and-there-are-probably-.html
15 www.japantimes.co.jp/news/2017/05/23/world/science-health-world/smart-genes-account-20-intelligence-study/#.WnBEomaZOfV
16 www.techtimes.com/articles/208037/20170523/science-stumbles-on-40-new-genes-linked-to-intelligence.htm
17 phys.org/news/2017-05-large-uncovers-genes-linked-intelligence.html
18 www.sciencedaily.com/releases/2017/05/170523083324.htm
19 www.the-scientist.com/?articles.view/articleNo/49491/title/Smarty-Genes/
20 newatlas.com/intelligence-genes-discovered/49650/
21 nation.com.pk/24-May-2017/smart-genes-account-for-20pc-of-intelligence
22 www.livescience.com/59252-intelligence-linked-to-52-genes.html
23 www.scienceworldreport.com/articles/59591/20170524/scientists-identify-52-genes-link-intelligence.htm
24 www.thecut.com/2017/05/genetics-intelligence.html
25 blogs.plos.org/onscienceblogs/2017/05/26/six-things-we-learned-from-that-massive-new-study-of-intelligence-genes/
26 www.alzforum.org/news/research-news/massive-gwas-reveals-40-new-intelligence-genes
27 www.sciencetimes.com/articles/16138/20170527/around-80-000-people-lead-to-identify-40-intelligence-genes.htm
28 qz.com/993356/a-massive-new-study-lays-out-the-map-of-our-genetic-intelligence/
29 www.bionews.org.uk/page_845044.asp
30 www.cbc.ca/radio/quirks/trump-exits-paris-accord-finding-genes-linked-to-intelligence-and-more-1.4143218/we-ve-found-50-genes-for-intelligence-could-that-lead-to-discrimination-1.4143354
31 www.vox.com/science-and-health/2017/6/6/15739590/genome-wide-studies
32 www.weforum.org/agenda/2017/06/scientists-just-found-40-new-genes-that-affect-your-iq/
33 uk.businessinsider.com/dna-genes-linked-human-intelligence-2017-10?r=US&IR=T

References

Bubela, T. M. and Caulfield, T. A. (2004). Do the print media "hype" genetic research? A comparison of newspaper stories and peer-reviewed research papers. *Canadian Medical Association Journal*, *170*(9), 1399–1407.

Cocodia, E. A. (2014). Cultural perceptions of human intelligence. *Journal of Intelligence*, *2*(4), 180–196.

Condit, C. M. (1999). *The Meanings of the Gene*. Madison: University of Wisconsin Press.

Gottfredson, L. S. (1997) Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23.

Heine, S. J. (2017). *DNA Is Not Destiny: The Remarkable, Completely Misunderstood Relationship Between You and Your Genes*. New York: Norton.

Hubbard, R., & Wald, E. (1997). *Exploding the Gene Myth: How Genetic Information Is Produced and Manipulated by Scientists, Physicians, Employers, Insurance Companies, Educations, and Law Enforcers*. Boston, MA: Beacon Press.

Kampourakis, K. (2017) *Making Sense of Genes*. Cambridge: Cambridge University Press.

Kampourakis, K. (2018). *Turning Points: How Critical Events Have Driven Human Evolution, Life, and Development*. Amherst, NY: Prometheus Books.

Kampourakis K. and McCain K. (2019). *Uncertainty: How It Makes Science Advance*. New York: Oxford University Press.

Keller, E. F. (2000). *The Century of the Gene*. Cambridge, MA: Harvard University Press.

Nelkin, D. (1995) *Selling Science: How the Press Covers Science and Technology* (Revised Edition). New York: WH Freeman.

Nelkin, D., & Lindee, S. M. (2004). *The DNA Mystique: The Gene as a Cultural Icon*. Ann Arbor: University of Michigan Press.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, *19*, 148–159.

Roll-Hansen, N. (2014). Commentary: Wilhelm Johannsen and the problem of heredity at the turn of the 19th century. *International Journal of Epidemiology*, *43*(4), 1007–1013.

Savage, J. E., Jansen, P. R., Stringer, S., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919.

Sniekers S., Stringer S., Watanabe K., et al. (2017) Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, *49*, 1107–1112.

Steinberg, D. L. (2015). *Genes and the Bioimaginary: Science, Spectacle, Culture*. Farnham, UK: Ashgate.

Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist, 59*(5), 325–338.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, *46*(11), 1173–1186.

INDEX