Synthese Library 367

Marie I. Kaiser
Oliver R. Scholz
Daniel Plenge
Andreas Hüttemann *Editors*

Explanation in the Special Sciences

The Case of Biology and History

Springer

Explanation in the Special Sciences

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY, LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

VOLUME 367

For further volumes:
http://www.springer.com/series/6607

Marie I. Kaiser • Oliver R. Scholz • Daniel Plenge
Andreas Hüttemann
Editors

Explanation in the Special Sciences

The Case of Biology and History

Springer

Editors
Marie I. Kaiser
Philosophisches Seminar
Universität zu Köln
Köln, Germany

Oliver R. Scholz
Philosophisches Seminar
Westfälische Wilhelms-Universität Münster
Münster, Germany

Daniel Plenge
Philosophisches Seminar
Westfälische Wilhelms-Universität Münster
Münster, Germany

Andreas Hüttemann
Philosophisches Seminar
Universität zu Köln
Köln, Germany

ISBN 978-94-007-7562-6 ISBN 978-94-007-7563-3 (eBook)
DOI 10.1007/978-94-007-7563-3
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013955260

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

Stephan Berry has worked many years in the lab on problems from biophysics and biochemistry; today he is working as a freelance science author. His publications, both academic titles and introductory texts for lay readers, cover issues from biophysics and evolutionary science as well as history and archaeology. A common thread in these seemingly disparate areas is the notion of processes: What are the driving forces for processes, and how can we reconstruct their trajectories – within the cell or the organism, during historical change on the scale of societies, and on the global level of biological evolution?

Carl F. Craver is an associate professor in the Department of Philosophy and the Philosophy-Neuroscience-Psychology Program at Washington University in St. Louis. He is the author of *Explaining the Brain* (Clarendon Press) and the coauthor of *The Search for Mechanisms: Discoveries Across the Life Sciences* (University of Chicago Press).

Alexander Gebharter is a predoctoral research fellow at the Department of Philosophy (Düsseldorf Center for Logic and Philosophy of Science) at the Heinrich-Heine-University, Düsseldorf, and within the DFG Research Group "Causation, Laws, Dispositions, and Explanation at the Intersection of Science and Metaphysics". He obtained a degree (Mag. phil.) in philosophy at the Department of Philosophy at the University of Salzburg in 2010. His main research interests lie within the philosophy of science.

Doris Gerber is a Privatdozentin at the University of Tübingen, where she has received her Ph.D. in 2001 and her Habilitation in 2010. Her main areas of interest and competence are action theory, social philosophy, philosophy of the human and social sciences, philosophy of history, ethics and political philosophy. She has published several articles on these topics. Her recent book, *Analytische Metaphysik der Geschichte. Handlungen, Geschichten und ihre Erklärung* (Berlin 2012: Suhrkamp), focuses on the problem of historical explanation and the underlying metaphysical problems

Stuart Glennan is a professor of philosophy at Butler University in Indianapolis. He received his B.A. in philosophy and mathematics from Yale University and his Ph.D. in philosophy from the University of Chicago. His work is chiefly concerned with explanation, causation, and modeling. He is a leading advocate of the new mechanicism in the philosophy of science. His publications include "Rethinking Mechanistic Explanation" (*Philosophy of Science*), "Ephemeral Mechanisms and Historical Explanation" (*Erkenntnis*), and "Mechanisms, Causes and the Layered Model of the World" (*Philosophy and Phenomenological Research*).

Daniel Immerwahr is an assistant professor of history, specializing in twentieth-century US foreign relations, at Northwestern University. He received his B.A. from Columbia University, an additional B.A. from King's College, Cambridge, and his Ph.D. from the University of California, Berkeley. He has also served as a postdoctoral fellow at Columbia University's Committee on Global Thought. His research has appeared in *Modern Intellectual History*, the *Journal of the History of Ideas*, among other places.

Marie I. Kaiser is working as a postdoc at the University of Geneva, Switzerland. She has studied philosophy and biology at the University of Münster, Germany. In 2012, she received her Ph.D. from the University of Cologne, Germany, with a thesis on *The Ontic Account of Explanatory Reduction in Biology*. During her Ph.D. studies, she was a member of the DFG Research Group "Causation, Laws, Dispositions, and Explanation at the Intersection of Science and Metaphysics" and a visiting fellow at the Center for Philosophy of Science at the University of Minnesota, USA. Her main research interests are the philosophy of biology and the general philosophy of science. In particular, her work focuses on the concept of reductive explanation in biology, mechanisms, biological parthood, causal modeling in the life sciences, and philosophical issues raised by the sciences of complex systems.

Philip Kitcher is the John Dewey Professor of Philosophy at Columbia. He is the author of books on a wide range of topics, namely, the philosophy of mathematics, the philosophy of biology, the growth of science, the role of science in society, Wagner's *Ring*, and Joyce's *Finnegans Wake*. In 2011, he published two new books: *Science in a Democratic Society* (Prometheus Books) and *The Ethical Project* (Harvard University Press). His collection of essays, *Preludes to Pragmatism*, was published in September 2012 by Oxford University Press and *Deaths in Venice: The Cases of Gustav von Aschenbach* will appear from Columbia University Press in 2013.

Ulrich Krohs is a professor of philosophy of science and of nature at the University of Münster. He has studied biochemistry and philosophy, received his Ph.D. in biology at the Technical University of Aachen and his "Habilitation" in philosophy at the University of Hamburg. Most of his present research interests are located at the interface of both fields: philosophy and history of science; philosophy of nature and of technology; model theory; biomedical ethics; philosophy of cognition; and epistemology. Before moving to the University of Münster, he has taught at the

universities of Hamburg, Vienna, Bielefeld, and Bern, and held research fellowships at the Konrad Lorenz Institute for Evolution and Cognition Research and at the University of Pittsburgh's Center for Philosophy of Science.

Gerhard Müller-Strahl has studied medicine and physiology in Aachen and Paris, philosophy and mathematics at the universities in Göttingen, Leipzig, and Bochum. He has been practicing neurology in Lausanne (CHUV) and psychiatry in Witten-Herdecke. He has been performing biorheologic research in Los Angeles (USC) and in Göttingen at the Max-Planck Institute for Experimental Medicine, followed by a position as research group leader for cardiovascular physiology in Leipzig at the Carl-Ludwig Institute. At the universities in Bochum and Münster, his philosophical research has been focused on concepts of organisms (eighteenth through twentieth centuries), on structure and change of physiological theories, on neo-Kantianism, and on contemporary theories of explanation and causation with special reference to the life sciences.

Daniel Plenge studied *Geschichtswissenschaft* (history, historiography, historical science, and science of history) and philosophy in Münster. He is a predoctoral research fellow at the Department of Philosophy at the Westfälische Wilhelms-Universität Münster and a member of the DFG Research Group "Causation, Laws, Dispositions, and Explanation at the Intersection of Science and Metaphysics".

Alexander Reutlinger is a postdoctoral research fellow in the DFG Research Group "Causation, Laws, Dispositions, and Explanation at the Intersection of Science and Metaphysics", and currently a visiting fellow at the Center for Philosophy of Science (University of Pittsburgh). He obtained a Ph.D. from the University of Cologne in 2011. His research interests lie within philosophy of science and metaphysics. He is particularly interested in semantic and metaphysical questions concerning causation, objective probabilities, and (ceteris paribus) laws in the social and life sciences and in physics. Fairly recently, he got interested in epistemological and metaphysical aspects of explanation, laws, and emergence in complex systems.

Oliver R. Scholz is a professor of philosophy at the University of Münster, Germany, and a principal investigator in the DFG Research Group "Causation, Laws, Dispositions, and Explanation at the Intersection of Science and Metaphysics". His research interests include epistemology, metaphysics, and the philosophy of science. In recent years, he has focused on social epistemology (testimony; expertise), the philosophy of history, and the philosophy of the social sciences. He is the author of *Bild, Darstellung, Zeichen* (1991, 3rd edition 2009) and *Verstehen und Rationalität* (1999, 2nd edition 2001), and he has published articles on many topics, inter alia in *Synthese*, *Erkenntnis*, and *Grazer Philosophische Studien*.

Gerhard Schurz is a professor of philosophy at the University of Düsseldorf, where he holds the chair of theoretical philosophy and is head of the DCLPS (Düssseldorf Center for Logic and Philosophy of Science), which hosts four DFG research projects. His major areas of research are philosophy of science, epistemology, logic, cognitive science, and meta-ethics. Among his books are *The Is-Ought*

Problem (Kluwer 1997), *Einführung in die Wissenschaftstheorie* (Wissenschaftliche Buchgesellschaft, 3rd ed. 2011), *Reliable Knowledge and Social Epistemology* (ed. with M. Werning, Amsterdam 2009), *Evolution in Natur und Kultur* (Spektrum 2011), and *Philosophy of Science: A Unified Approach* (to appear with Routledge 2013).

Daniel Steel is an associate professor in the Department of Philosophy at Michigan State University. His research focuses on evidence and causal inference, especially as these topics pertain to the social and biological sciences. He is the author of *Across the Boundaries: Extrapolation in Biology and Social Science* (2008) and coeditor with Francesco Guala of *The Philosophy of Social Science Reader* (2011). In addition, he is the author of numerous articles that have appeared in such journals as *Philosophy of Science*, *Philosophy of the Social Sciences*, *British Journal for the Philosophy of Science*, and *Biology and Philosophy*.

Aviezer Tucker works at the University of Texas at Austin. He is the author of *Our Knowledge of the Past: A Philosophy of Historiography* (Cambridge University Press 2004) and editor of the *Blackwell Companion to the Philosophy of History and Historiography* (Wiley-Blackwell 2008). He conducts research on the philosophy of historiography, the philosophy of science, and political philosophy and their interactions.

Derek D. Turner is an associate professor of philosophy and a fellow of the Goodwin-Niering Center for the Environment at Connecticut College. He is the author of two books on historical science: *Making Prehistory: Historical Science and the Scientific Realism Debate* (2007) and *Paleontology: A Philosophical Introduction* (2011), both with Cambridge University Press. His recent papers on philosophical issues in paleontology have appeared in *Biology and Philosophy* and *Theory in Biosciences*.

Chapter 1
Introduction: Points of Contact Between Biology and History

Marie I. Kaiser and Daniel Plenge

Abstract The goal of this introductory chapter is to show how debates about scientific explanation in the philosophy of biology and in the philosophy of history can be conjoined to stimulate and enrich each other. We draw attention to two major points of contact: first, it seems as if historical explanations are not restricted to the historical science, but can be found for instance in the biological science as well (most notably in evolutionary biology). This raises the question of what it is that makes a scientific explanation "historical". Second, in philosophy of biology and in philosophy of history we recently find an emphasis on the role of mechanisms and mechanistic explanation in both sciences. But are there really such things as historical or social mechanistic explanations, and how do they relate to biological mechanistic explanations? In this paper we introduce different answers to these questions and argue that they enable and demand a joint and mutually stimulated discussion, by philosophers of history and by philosophers of biology.

Keywords Historical explanation • Narrative explanation • Historical law • How-possible explanation • Social mechanism • Mechanistic explanation • Historical science

It might seem a surprising project with disputable merits to edit a volume that aims to conjoining the debates about explanation in the philosophy of biology and in the philosophy of history. People with these kinds of reservations may have in mind

M.I. Kaiser (✉)
Philosophisches Seminar, Universität zu Köln, Richard-Strauss-Str. 2, 50931 Köln, Germany
e-mail: kaiser.m@uni-koeln.de

D. Plenge
Philosophisches Seminar, Westfälische Wilhelms-Universität Münster, Domplatz 23, 48143 Münster, Germany
e-mail: daniel.plenge@uni-muenster.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_1,

the opposition between nature and history[1] or between natural history and human history.[2] They might also remember that countless philosophers have argued that history or historiography is a part of what is often called "the humanities" (*Geisteswissenschaften*, *sciences humaines*), rather than being a part of "the sciences." They may reminisce about old battles on scientific versus hermeneutic approaches or debates on explanation versus *Verstehen* (understanding) in the philosophy of the social sciences. Some might even believe that "natural history" is an ontological contradiction and that its replacement by "historical science," "science of history," or "natural historiography" would not be any better in methodological respects. Without much exaggeration, even the following picture seems to have admirers in some quarters: unlike scientists (e.g., biologists) who produce empirically tested and at least approximately true theories about the world which exists independently of us, historians are men and women of letters who do not engage in scientific theory construction but in the *writing* of history, that is, in the writing of some form of literature or what is frequently called "narratives." Historians do not scientifically reconstruct or model the independently existing world. Rather, they are often said to construct it altogether by their writing, which is why their mode of comprehension is said to be fictional and not scientific.[3]

> History, some say, is an art, not a science. (Louch 1969, p. 61)

We believe that this picture is flawed and that there are good reasons for setting aside the reservations one might have against our project of conjoining biology and history/historiography. Accordingly, we speak of "the historical science"[4] and treat it as a part of the so-called special sciences,[5] just as the biological sciences.

---

[1]The term "history" is ambiguous. It refers at least to three different things: first, to something ontic (e.g., to the history of an object); second, to some discipline (e.g., historiography); and third, to the results obtained by some scholar and/or its presentation in form of a text (e.g., the Cambridge History of x).

[2]Oppositions like these may be due to the fact that history/historiography is traditionally concerned only with the study of human, cultural, or social phenomena. Hence, traditional philosophy of history has not included the philosophy of *natural* history so far (but this may change in the future; see, e.g., this volume, Part III and IV, and Cleland 2002, 2009, 2011).

[3]"Postmodern" philosophers of history might be said to come close to this caricature (see, e.g., Jenkins 1991; Munslow 2007). Even scholars who do not believe in a fundamental difference between history/historiography and the sciences constantly use phrases such as the "writing of history" when referring to what historians do or to history as a discipline (the most recent example is Leuridan and Froeyman 2012, p. 172). For the most recent and explicit oppositions to such expressions, see Kosso (2001) and Tucker (2004).

[4]However, we are aware of the fact that expressions such as "historical science" or "science of history" (contrary to the term "Geschichtswissenschaft" in German) are hardly ever used in the philosophy of history. This fact is remarkable, but, as one might be willing to say, due to the history of the field.

[5]We use the term "special science" merely because it is an established way to refer to everything else except physics. Apart from that, we are not completely happy with this term because it might convey the implicit message that disciplines like biology and history/historiography are "special" and thus inferior to physics.

This does not imply to blur the differences between these two disciplines. We agree that there are significant disparities among these two fields (e.g., concerning the role of experiments, the nature of the "empirical data," or the kinds of theories/generalizations that are developed). And we are aware of the fact that a much more elaborate discussion about questions such as "In which respect is history/historiography a science?" and "What is historical science?" is needed than the one we can provide here (cf. Kitcher and Immerwahr, this volume, Chap. 14; for a discussion of the peculiarities of historical sciences see Scholz, this volume, Chap. 11 and Tucker, this volume, Chap. 16). However, we are convinced that treating biology and history/historiography as siblings, rather than strangers, enables us and the contributors to this volume to establish fruitful connections between the two disciplines and to work out relevant differences.

There are *two major points of contact* between the debates about explanation in the philosophy of biology and in the philosophy of history that we think are worth being emphasized: first, the question of whether *historical explanations* can be found in biology and what it is that makes an explanation "historical" in character and, second, the recent emphasis on mechanisms and *mechanistic explanation* that can be observed in both fields. We successively elaborate these two points of contact in the next sections. In doing so, we introduce significant questions and theses that enable and, as we think, *demand* a joint and mutually stimulated discussion, by philosophers of history *and* by philosophers of biology.

1.1 Historical Explanation in Biology

The first point of contact is the thesis that *historical explanations* are not restricted to the (human) historical sciences but can also be found in other sciences, for instance, in cosmology, geology, paleontology, and also in the biological sciences, particularly in evolutionary biology (see, e.g., Goudge 1961; Mayr 1982; Rosenberg 2001, 2006; see Scholz, this volume, Chap. 11, on the spectrum of the historical sciences).

Some philosophers of biology, most notably Alex Rosenberg (2001, 2006), even claim that *all* biological explanations are (at least implicitly) "historical" in character. Rosenberg's argument relies on two main assumptions: first, on the controversial claim that, in biology, there exists only one law, namely, the "principle of natural selection" (2006, p. 150), which is a *historical* law (cf. Reutlinger, this volume, Chap. 6),[6] and, second, on the thesis that all explanations require the description of laws in order to be explanatory. From this Rosenberg concludes

[6]Reutlinger (this volume, Chap. 6) examines the question of whether and in which sense biological generalizations can be characterized as being "historical" and "contingent."

that all biological explanations must, at least implicitly, refer to the principle of natural selection and hence that "biological explanation is historical, all the way down to the molecules" (2006, p. 152). According to Rosenberg, the principle of natural selection comes into play as soon as an explanation refers to biological types: since biological types are functionally individuated and since functions must be understood etiologically (2006, pp. 17–20), any reference to biological types implicitly invokes evolutionary theory (more specifically, the description of *past* evolutionary processes). Even biological explanations, such as the molecular explanation of how DNA is replicated during cell division, implicitly appeal to evolutionary theory because they contain statements about biological types (e.g., DNA polymerase, nucleotides) which are individuated with reference to their past selective effects. Rosenberg concludes:

> Any subdiscipline of biology ... can uncover at best historical patterns, owing to the fact that (1) its kind vocabulary picks out items generated by a historical process, and (2) its generalizations are always open to being overtaken by evolutionary events. (2006, p. 153)

We do not share Rosenberg's radical view that *any* biological explanation is an (at least implicit) evolutionary explanation and thus a historical explanation. However, what is interesting about his view is the tight connection between evolutionary and historical explanation that he and others envisage. The overall question to which authors like Rosenberg provide an affirmative answer is:

Do there exist types of explanation in biology (e.g., in evolutionary biology) that are *historical*?

Answering this question with "yes" presupposes at least a rough idea about what a *historical explanation* is. In other words, it requires that the following question is answered:

What makes an explanation a specifically *historical* explanation?

Unsurprisingly, there is no consensus in the philosophy of biology about what a historical explanation is (or, to speak with Craver, what the "norms" are that distinguish historical from nonhistorical explanation; this volume, Chap. 2). Rosenberg, for instance, sides with Hempel (1942) and argues that historical explanations in biology are explanatory not because they redescribe the explanandum or because they link the explanans to the explanandum through the operation of implicit necessary truths about rational action (2001, p. 748). Historical explanations in biology rather explain because they (at least implicitly) appeal to the only biological law that we have, namely, to the principle of natural selection. Hence, Rosenberg agrees with Hempel that most historical explanations are incomplete "explanation

sketch[es]" (1942, p. 42) that do not explicitly refer to laws but invoke them as background information.[7]

> Historical explanations are *sketches of covering-law explanations* that implicitly appeal to historical laws.[8]

Contrary to Hempel and Rosenberg, Thomas A. Goudge (1961), the first philosopher of biology who addressed this issue, denies that historical explanations in evolutionary biology are explanatory because they deduce the explanandum event from a law or set of laws (e.g., the principle of natural selection). Instead, he characterizes them as *narratives* which show "how existing states of affairs are the result of the combined action of sequences of past events" (1961, p. 68).[9] For example, the eyespot on the wings of peacock butterflies is explained by the story of how certain events have led to the selection of this trait in populations of peacock butterflies.[10] Goudge stresses that in evolutionary biology, explanations are not covering-law explanations but rather "narrative explanations" (1961, p. 75) that establish an "intelligible, broadly continuous series of occurrences which leads up to the event in question" (1961, p. 77). According to Goudge, evolutionary biologists do the same as historians do when they explain: they tell a "likely story" (1961, p. 75), that is, they represent a number of possible events in an intelligible, coherent sequence.

> Historical explanations are *narrative* explanations.

At this point one might query whether the picture that philosophers of biology like Rosenberg and Goudge draw is an adequate view of what historians do and how they explain (and one might wonder which of them is right). Reason enough to have

[7] In philosophy of history, the notion of an explanation sketch is one of the most negatively connotated doctrines. The reason is that it seems to imply the immaturity of history that produces "mere" sketches of explanations, rather than complete explanations. But, of course, others believe that the doctrine of explanation sketches shows how "scientific" history was even in 1942 and that it did not and does not provide "mere" fables.

[8] In philosophy of history, connections between the alleged historicity of laws (or generalizations) and a specific type of historical explanation were already drawn by Terence Ball (1972, p. 184): "An historical explanation (…) is (…) one in which at least one 'law' (or better, perhaps, quasi-law) in the explanans is tensed or temporally located."

[9] A similar view can be found in Hull (1975, 1989).

[10] The events that are described in the explanans include, for instance, the predators' eating of butterflies without spots, the predators' being scared off by some of these butterflies due to their wind spots, and the predators' being hunted by owls that have eyes resembling the wing spots of peacock butterflies.

a look at what philosophers of history say about this issue. In philosophy of history, the question of what counts as a *historical explanation* has been a frequent matter of dispute. Hence, this seems to be one point at which the philosophy of biology can benefit from the philosophy of history (and *vice versa*).

The understanding of "historical explanation" that is most prevalent is that historical explanations are those explanations that are offered by people who are legitimately called historians. However, this thesis either is uninformative (if historians happen to be those people working in, say, departments of history) or calls for a clarification of notions such as "historical science," "historical studies," or "historical method."[11]

Another suggestion as to what makes an explanation historical has been provided by Gordon Graham. He argues that:

> "[A] historical explanation is one which explains a fact by *giving its history*." (Graham 1983, p. 65, our emphasis)

But this answer raises follow-up questions. Most importantly, it leaves open what it means to describe the "history" of the explanandum, and what the explanandum of a historical explanation is at all. It therefore seems as if this answer only shifts the focus of the question from the needed explication of "historical" to a specification of "history."

One possible way to get a more specific notion of "historicity" in theories of historical explanation has been to reserve specific explananda for such explanations. Thus, various philosophers of human historical science identified *historical explanations* with explanations of individual actions.

> Historical explanations are *explanations of individual actions.*

In this line of thought, the famous philosopher of history, William Dray, stepped into Robin Collingwood's (1994 [1946]) shoes by claiming that

> The objects of historical study are fundamentally different from those, for example, of the natural sciences, because they are the actions of beings like ourselves. (Dray 1957, p. 118)

[11] Some scholars even hold the view that there is just no specifically historical type of explanation (e.g., Hempel 1942; White 1943). May Brodbeck is responsible for one of the most famous quotes in this context: "There is no such thing as 'historical explanation', only the explanation of historical events." (1962, p. 254) However, it remains unclear what a historical event is and whether there is something special about historical events that makes them different to, say, natural events. If there exist specifically historical events, one might even argue that the fact that there is a class of explanations that explain specifically historical events suffices to call them specifically historical explanations.

In the classic Hempel-Dray-Scriven debate and in this variety of philosophy of history, it is controversial whether such explanations contain empirical laws (e.g., Hempel 1962, 1963), truisms or "normic statements" (e.g., Scriven 1959), or "principles of action" (Dray 1957, 1963), and whether such explanations are varieties of causal explanation or whether they are "reason explanations" sui generis (for an overview, see Dray 2000). One might want to claim that the "intentional" or "rational" character of social or historical phenomena makes them special in a more significant way. The primary problem with such positions has been stated variously. For many philosophers and social scientists, several social or historical phenomena or changes in social systems are neither intended nor rational but unintended outcomes of myriads of perhaps rational individual actions.

Be that as it may, the question remains of whether action explanation provides us with material that leads to an adequate understanding of the concept of *historical* explanation.[12] We are skeptical. The reason is that, although it is a contingent truth that history/historiography traditionally has been restricted to the study of human, cultural, or social phenomena, restricting the concept of historical explanation to the explanation of one type of phenomenon or events (namely, to individual human action explanations) seems to be too arbitrary. One might easily find arguments for Dray's thesis that the kinds of phenomena that historians or social scientists investigate are fundamentally different from phenomena of the natural world (e.g., the assertion that formations of rocks, which are the result of some "historical" processes, do not think about their "history," since they do not think at all).[13] Nevertheless, a concept of historical explanations that identifies them with explanations of singular actions of humans is too narrow to be convincing. Such a narrow concept would have at least two implausible consequences: first, explanations of social phenomena, such as wars, inequality, or economic decline, would not count as *historical* because they do not explain *singular* actions (see also Sect. 1.2); second, explanations of events that involve no humans at all (e.g., the explanation of the extinction of the dinosaurs) would be excluded from the set of historical explanations, too. And many "historical" phenomena are, of course, inseparably mixed, that is, natural-social or biological-social (e.g., climate change

[12]One of Dray's papers on the topic has the revealing, yet ambiguous title "The Historical Explanation of Actions Reconsidered"; see Dray (1963). Whereas Graham (1983, Chap. 4) is right in discussing Dray's position as a paradigm for debates about historical explanation, Dray meanders between an understanding of historical explanation as (i) a label that encompasses all explanations that historians develop and as (ii) a type of explanation that is historical. For another example of this problem, see Martin (1977).

[13]One possible source of arguments in favor of the existence of a fundamental difference between social/historical and natural phenomena is the debate about what has often been called the question of "naturalism," that is, the question of whether human or social phenomena are of a kind that prohibits their being studied "scientifically." For a famous contribution to this debate, see Bhaskar (1979).

or the "Black Death"). Thus, explanations of singular human actions might be *an* important subtype of historical explanation but not *the only* existing kind of historical explanation.[14]

Let us return to the idea that what makes explanations specifically *historical* is their *narrative* character. As we have seen, narrative explanations are at times characterized as describing the continuous series of events by which the explanandum event came about. This claim cannot only be found in philosophy of biology but is popular in philosophy of history as well.[15] The "model of a continuous series" of events, proposed by William H. Dray, is similar to the views about narrative explanations expressed by philosophers of biology, such as Goudge. According to Dray, these models explain an event by enabling the enquirer to "trace the course of events by which it [the explanandum event] came about" (Dray 1957, p. 68). This view is frequently presented as an alternative to the view that explanations require the description of general laws (i.e., must be covering-law explanations).[16] However, just as its counter position, the idea that historical explanations are narratives faces objections, too. For instance, the question arises of what exactly a "historical narrative" (Hull 1975, p. 253) or a "likely story" (Goudge 1961, p. 75) is and what makes them explanatory, rather than merely descriptive.[17] This conceptual vagueness is particularly surprising in light of the huge amount of narrativisms that are on the market in philosophy of history.

---

[14] Somebody who rejects this claim is, for example, J. O. Wisdom (1987).

[15] Despite the popularity of this view, it is important to note that not every historian holds that historical explanations (i.e., explanations in historiography) are narratives (cf. Hull 1975, p. 254). Furthermore, one should notice that various meanings of the term "narrative" are used in the literature. Classics of this genre are Dray (1954) and White (1963). While some of the early narrativists believed, roughly, that "narratives" track event sequences, more recent narrativists understand the concept of a narrative and of a narrative explanation in a much broader sense, that is, as referring to literary features, artistic means, and rhetorical devices applied in "history" or to any kind of text that creates "meaning." For recent critical discussion, see, for example, Day (2009), Frings (2008), Murphey (2009), and Brzechczyn (2009). For an anthology of the debate about narratives in philosophy of history, see Roberts (2001).

[16] As is well known, Dray's claim that such explanations by descriptions of continuous series do not require or imply laws was countered by Maurice Mandelbaum (1961). Later Mandelbaum (1977) uses the same notion (i.e., "continuous process") in explanatory contexts or what he terms "concrete causal analysis."

[17] Famously, Hempel argued against similar views by writing that "the mere enumeration in a yearbook of 'the year's important events' in the order of their occurrence clearly is not a genetic explanation of the final event or of anything else" (Hempel 1962, p. 23). "Genetic explanation" is Hempel's model of what he assumes to be an "explanatory procedure, which is widely used in history" (1965, p. 447). He explicates this model as follows:

> In order to make the occurrence of a historical phenomenon intelligible, a historian will frequently offer a 'genetic explanation' aimed at exhibiting the principal stages in a sequence of events which led up to the given phenomenon (Hempel 1962, p. 21).

These stages were, of course, to be covered loosely by laws. Saliently enough, this model is not very different from what Arthur Danto (1965) later referred to as narrative explanation.

Another interesting point of discussion concerning narrative explanations emerged in philosophy of biology. Several authors, most notably Stephen J. Gould and Richard C. Lewontin (1979), have claimed that adaptive explanations in evolutionary biology must be more than "just-so-stories," that is, more than merely plausible stories about how a trait could *possibly* have evolved in a possible environment. Accordingly, one could argue that:

> Historical explanations must be more than "just-so-stories" or *how-possible explanations*.

In line with this, Rosenberg argues that historical how-possible explanations must be "made adequate" by converting them into historical why-necessary explanations (2006, pp. 47–55). In Rosenberg's writings, it remains unclear how exactly this transformation shall proceed. What he does mention is that it includes the filling up of crucial links in the causal chains of the original explanation. Rosenberg's argumentation suggests that historical explanations are adequate only if they consist in more than in the telling of possible stories. Rather than describing what *could* have happened, adequate historical explanations, so Rosenberg, tell the story of what has *actually* happened and why this must have happened. In short, historical explanations are adequate only if they are why-necessary explanations. In a similar vein, Glennan claims that "to the extent that a narrative fails to show the necessity of the outcome, it fails to explain" (2010, p. 262). More generally, Craver (this volume, Chap. 2) argues that a philosophical theory of explanation must distinguish how-actually explanations from mere how-possibly models.

However, this view is far from being uncontested – among philosophers of biology *and* among philosophers of history. In the philosophy of biology, some authors defend the view that the evolutionary explanations that are given in practice are often not more than how-possible explanations. For example, Schurz (this volume, Chap. 7) argues that evolutionary explanations are considered to be adequate only if they specify at least some plausible mechanisms (of variation and of selection). Since these plausible mechanisms need not be empirically confirmed to a high degree, he concludes that evolutionary explanations are often mere how-possible explanations, rather than full causal explanations. In the philosophy of history, it is disputed in which sense historians can even explain by showing that an event was necessary, since history is claimed to be "contingent" (Little 2010).[18] Moreover, it is questioned whether historians can access enough evidence to fill in the links of the causal chains that lead up to an event (cf. Tucker, this volume, Chap. 16). If it turns out that historians have to provide such information to explain, facing the fact that often they cannot do so again leads to a philosophical scenario in which it would be challenged whether historians provide explanations at all.

[18] Stephan J. Gould wrote: "the central principle of all history – *contingency*" (2000, p. 283).

Rosenberg's claim that one can get historical why-necessary explanations by adding causal information to historical how-possible explanations gives rise to another interesting question: Are *historical explanations* a special kind of causal explanations, or are they opposed to them? Let us consider the former alternative:

Historical explanations are special kinds of *causal explanations*.

In philosophy of history, the terms "narrative explanation," "historical explanation," and "causal explanation" are often used interchangeably. We find an example for this even beyond the disciplinary boundaries. Daniel Athearn, who cannot be claimed to have been preoccupied with history or biology, writes of "narrative (historical or causal) explanation" (1994, p. 5). He argues in favor of a place for narrative explanations even outside the special sciences (e.g., in physics): "to produce explanations in science is to produce narrative causal explanations" (1994, p. 61). These explanations, which he also calls "productionistic explanations" (1994, p. 59), seem to be similar to the mechanistic explanations that have been frequently discussed in philosophy of biology during the last decade (see Sect. 1.2).[19] Although causation is an understudied field in philosophy of history and although causal accounts of explanation are far less frequent than one might expect, quite a few authors argue that explanations in history are causal explanations (see, e.g., Gerber, this volume, Chap. 9, and the discussion of causal explanations of historical trends by Turner, this volume, Chap. 12). Connections between narratives and causal explanations are sometimes also drawn.[20]

However, other philosophers of history have argued that causation has no central place in history and that historical/narrative explanations are *opposed to* causal explanations (this literature ranges from, e.g., Louch (1969) to Gorman (2007) more recently). A particularly sharp contrast between historical/narrative explanations, on the one hand, and causal explanations, on the other hand,[21] is drawn by those scholars who paint the picture that we referred to at the beginning: scientists gather empirical data and develop theories about the world, whereas historians are men and women of letters who *write* history; accordingly, what has often been called "historical interpretation" is supposed to be fundamentally different from any

[19] "A 'productionistic' explanation is a causal explanation of which the only essential components are events arising out of one another in succession and/or giving rise to (in a perfectly innocent and literal sense) the fact, entity, or phenomenon that the particular story explains." (Athearn 1994, p. 59)

[20] For accounts using causation in theorizing about explanations in history, see, e.g., Topolski (1976), Mandelbaum (1977), McCullagh (1998), Day (2009), Frings (2008), and Murphey (2009).

[21] Of course, back in those days the enemy of those who endorsed a strong opposition between history and natural science was the covering-law model of explanation.

scientific causal explanation.[22] Moreover, debates about hermeneutic understanding versus causal explanation are another central and classic place where such contrasts are emphasized (for an overview, see Martin 2000). Kitcher and Immerwahr (this volume, Chap. 14) argue that many of these older debates and recent revivals are mistaken because they battle the wrong philosophies of science in general and of explanation in particular. Further counterpositions to this tradition can be found in Scholz, Turner, Glennan, Steel, and Tucker (this volume, Chaps. 11, 12, 13, 14, 15 and 16).

To sum up, the claim that there also exist historical explanations in other sciences (e.g., in biology) gives rise to the question of what makes an explanation specifically historical. We have shown that a satisfying answer to this question is missing (although several answers are discussed) and that philosophy of biology and philosophy of history can fruitfully work together to specify the concept of a historical explanation and to determine its scope.

1.2 Mechanistic Explanation in the Historical Sciences

The *second* major point of contact is the increasing attention to mechanisms and mechanistic explanations that can be observed in both fields: in the philosophy of biology (e.g., Machamer et al. 2000; Craver 2007; Glennan 1996, 2002, this volume, Chap. 13; Bechtel 2006, 2008; see also Gebharter and Kaiser, this volume, Chap. 3, and Müller-Strahl, this volume, Chap. 5) and in the philosophy of history, of historical sociology (e.g., Norkus 2005, 2007; Lloyd 1986; see also Plenge, this volume, Chap. 10), and in social science in general.[23]

The "new mechanistic philosophy" (Skipper and Millstein 2005, p. 327) has been primarily developed with regard to the life sciences. Accordingly, most proponents of the mechanistic account concede that mechanistic explanations are an important but not the *only* kind of biological explanation (e.g., there might be what Krohs calls "semiotic explanations," too; this volume, Chap. 4). However, in recent years, there has been a tendency to extend the boundaries of the scope of the mechanistic account. Most notably, Stuart Glennan (2010, this volume, Chap. 13) argues that historical explanations also fall under the category of mechanistic explanations – even if they describe mechanisms that are less stable than other

[22] A paradigmatic example of such a view is the following: "The starting-point of the present study is the claim, common to almost all critical philosophers of history [sic!], that historical study aims at a kind of understanding quite different from that which is characteristic of the natural sciences" (Gallie 1964, p. 11). In a similar vein, this idea can be found in Mink (1966).

[23] Many authors do not differentiate between historical and social sciences. Although such terminology can be criticized, we neither want to take a stand on whether history and social science are distinct fields or whether they are closely related, nor do we take a stance on whether they have differing methods or not. For discussions concerning these points, see Glennan (this volume, Chap. 13) and Tucker (this volume, Chap. 16).

mechanisms (so-called ephemeral mechanisms). At the same time there emerged a, mostly independent, debate about mechanisms and mechanistic explanations in the philosophy of history, in philosophy of the social sciences, and within history/historiography and sociology itself.[24]

One might want to dispute that this is a fruitful point of contact by arguing that what philosophers of the historical and the social sciences mean by mechanisms and mechanistic explanation is different from the concept of mechanism and mechanistic explanation that is established in the philosophy of biology.

> Biological mechanisms are *fundamentally different* from historical/social mechanisms.

To support this claim, one might, for instance, point to the putatively categorical difference between the mechanism of a clock or of photosynthesis, on the one hand, and the mechanism of a particular children's birthday party, of a social thing like a university, or mechanisms that could have been responsible for the fall of the Roman Empire,[25] on the other hand. This line of argumentation might be motivated by the intuition, similar to Dray's position hinted at above, that the explanation of human actions and of social phenomena cannot be "of a piece with the explanation of the working of clocks or other mechanical devices" (Norkus 2005, p. 372). However, we think that the questions of whether there is a fundamental difference between biological mechanisms and historical or social mechanisms[26] and whether mechanistic explanations encompass biological, historical, and social explanations as well are still open for discussion. One might convincingly argue that there is a fundamental difference between the social and the natural world, or the arguments in favor of the entanglement of the social and the natural world might turn out to be more plausible (cf. Steel's investigation of "coupled human and natural systems (CHANS)," this volume, Chap. 15). Future discussion will show. We see no convincing arguments for nipping the discussion in the bud.

Mechanistic explanations, as they are understood in the philosophy of the life sciences, are descriptions of how the components of a mechanism are organized and how they interact with each other in order to bring about the explanandum phenomenon (cf. Machamer et al. 2000; Craver 2007; Glennan 1996, 2002; Bechtel 2006, 2008). For instance, the phenomenon of muscle contraction is mechanistically explained by describing how certain molecules and cell organelles (e.g., cal-

[24]The literature in this field is large and still growing (see, e.g., Hedström and Swedberg 1998; Tilly 2004, 2008; Schmid 2006; Manicas 2006; Demeulenaere 2011; Wan 2011).

[25]For a detailed analysis of the mechanisms that might be relevant to explain the histories of the Roman Republic and the Roman Empire, see Berry (this volume, Chap. 8).

[26]For some suggestions on how to explicate the concepts historical mechanism and social mechanisms, compare Glennan (this volume, Chap. 13) and Plenge (this volume, Chap. 10).

cium ions, myosin and actin filaments, the sarcoplasmic reticulum, tropomyosin molecules) interact with each other in a certain way (or, as others prefer to say, perform certain activities or operations, e.g., binding, releasing, tipping over, converting) so that they together produce the shortening of the muscle fiber. Some proponents of the mechanistic account (e.g., Craver, this volume, Chap. 2) argue that the causal mechanisms in the world itself, rather than our representations of them, are the explanations. Though there is considerable consent with regard to the main features of biological mechanism, there is also a lot of disagreement among the mechanists. What, for instance, is the ontological nature of the components of mechanisms (e.g., Müller-Strahl develops a mechanistic ontology for disease entities, in which the concept of a mechanistic base occupies center stage; this volume, Chap. 5)? Must mechanisms produce a phenomenon regularly, or can there be mechanisms that bring about a phenomenon only once? Must the parts of a mechanism be located on a lower ontological level than the mechanism as a whole, or can there be such thing that Craver (2007) calls "etiological mechanistic explanations," too? Is the mechanistic view committed to a special theory of causation, for example, one that accounts for the "productive" character of activities? Finally, how are biological mechanisms adequately represented (e.g., Gebharter and Kaiser argue that biological mechanisms can be represented by causal graph theory and that the resulting quantitative, probabilistic models are useful for certain scientific purposes; this volume, Chap. 3)?

We can now ask whether there are (specific types of) historical explanations that are similar to the mechanistic explanations that can be found in biology and that might be termed mechanistic explanations as well. In other words, an interesting working hypothesis is:

> There exist *mechanistic explanations* in the historical sciences, too.

This hypothesis is supported by the examples of historical mechanistic explanation that are presented by Berry (this volume, Chap. 8). But in what follows, we focus on the philosophical arguments that have been or can be provided in favor or against this hypothesis. We discuss three lines of argumentation. Two of them support the above thesis; the other one denies that the concept of mechanistic explanation can be applied to historical explanation[27] as well.

First, Glennan (2010 and this volume, Chap. 13) claims that historical explanations (or, as he sometimes calls them, "historiographic explanations") describe mechanisms, too. This claim presupposes that the notion of a mechanism

[27] In this context, "historical explanation" can be understood either as a specific type of explanation or as the explanations that historians typically give (which leaves open which types of explanation they offer).

is understood in a broad way. Glennan accepts Phyllis McKay Illari and Jon Williamson's general definitions of mechanisms, according to which:

> [a] mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (2012, p. 120)

Glennan argues that what makes "historical mechanisms" (2010, p. 260) similar to other kinds of mechanisms is that they bring about events, too (namely, "historical events," 2010, p. 264), and that descriptions of how the parts of historical mechanisms interact with each other explain the historical event. In Glennan's view, the only major difference between historical mechanisms and mechanisms for, say, DNA replication is that the former are less stable than the latter. He argues that, whereas biologists and other natural scientists study relatively stable systems, the mechanisms that figure in historical explanations are "ephemeral and capricious" (2010, p. 251). This means that the specific configuration of the parts of a historical mechanism (i.e., their coming together) is short-lived and may be contingent.[28] For this reason, Glennan refrains from calling such mechanisms "systems" but claims that they are better conceived of as "processes." However, Glennan emphasizes that despite the ephemeral nature of historical mechanisms, the interactions between their parts have a robust and reliable nature, too. Hence, they can also be described by "direct, invariant change relating generalizations" (2010, p. 260; see also 2002) – just as in case of traditional mechanisms. Glennan concludes that historical explanation is a subtype of mechanistic explanation.

Historical explanations are descriptions of *ephemeral mechanisms*.

Glennan stresses that even the characterization of historical explanations as narrative explanations does not render this thesis implausible since narratives are nothing but descriptions of ephemeral mechanisms.

Second, bringing together the debate about explanation in philosophy of biology and in philosophy of history reveals an interesting similarity, namely, the one between *mechanistic explanations* in biology and historical explanations as *narrative explanations* understood in a narrow way (this similarity is also recognized by Glennan 2010, this volume, Chap. 13). As we have pointed out before, some scholars characterize explanations in history/historiography as historical narratives, for instance, as "models of continuous series" (Dray 1957) of events that bring about the explanandum event. In Goudge's words (recall Sect. 1.1), a narrative explanation

[28] On this basis, one might even argue that the mechanism of natural selection is an ephemeral historical mechanism. This is exactly what Skipper and Millstein (2005) deny in their renowned paper. On the contrary, Illari and Williamson (2010) claim that mechanistic explanations by protein synthesis and by natural selection are more closely analogous than they appear.

establishes an "intelligible, broadly continuous series of occurrences which leads up to the event in question" (1961, p. 77). Some authors specify this claim by pointing out that historical narratives explain an event by "integrating it into an organized whole" (Hull 1975, p. 273) or that "[t]he aim is to make the sequence of events intelligible as a relatively independent whole" (Goudge 1961, p. 75). Claims like these are prevalent in the philosophy of history. It seems as if historical narrative explanations *understood in this way* are quite similar to the mechanistic explanations or models that are typical of the life sciences. The similarity rests on the fact that in both cases what is important for the explanation is some kind of integration of something into a "whole." Hence, one could argue that:

> Historical explanations are mechanistic explanations because they are *narrative* explanations.

We think that especially two possible analogies between mechanistic explanations in biology and narrative explanations in history/historiography understood in this way are worth being examined more closely: first, that explanations of both kinds specify part-whole relations, and second, that both of them explain a phenomenon (or event) by describing how a continuous sequence of events (or processes) brings about (or leads to) the explanandum phenomenon. We elaborate these aspects one after another.

An important idea, in the literature about mechanisms in the life sciences, is that mechanistic explanations "span multiple levels" (e.g., Craver 2007, p. 163; Gebharter and Kaiser, this volume, Chap. 3). This means that they explain a particular behavior or feature of a mechanism as a whole by appealing to the entities and activities that compose the mechanism (which are said to be located on a lower level of organization). In other words, mechanistic explanations require that a certain kind of *part-whole relations* is specified, namely, the ones between a mechanism and its components (e.g., between the mechanism for protein synthesis and the ribosomes, the mRNAs, the amino acids, their binding, moving, and linking). Interestingly, the idea that something is integrated into a whole and that it is important to figure out what belongs to this whole and what does not seems to be central to the concept of a narrative explanation in history, too. According to a prominent view, historical narratives also explain a particular event or phenomenon by representing only those events that are *relevant* (Hull 1975, p. 274), that is, that together form a *coherent, continuous whole* that culminates in the event to be explained (Goudge 1961, pp. 73–75). Hence, both mechanistic explanations in biology and narrative explanations in history/historiography seem to be models that involve representations of part-whole or constitutive relations.

One might challenge this analogy by denying that historical narrative explanations appeal to part-whole relations and thus span multiple levels. The argumentation could proceed as follows: even if narratives represent some "historical

process"[29] as being an integrated whole, this does not imply that there really exists a whole (i.e., the historical process) in the world that is located on a higher ontological level than the events that compose it.[30] These processes, which historians are supposed to investigate, are mere sequences of events, and the event to be explained is just the final event/the end of this sequence (e.g., the outbreak of a war), rather than being located on a higher ontological level. Moreover, the putative "wholes" represented in historical explanations are not as robust as biological mechanisms and do not regularly and repeatedly lead to the explanandum event. Thus, they are not real wholes after all. Although this line of argumentation has some convincing aspects, it also seems to overstate the ontological differences between biological mechanisms and historical "processes" (for a view that emphasizes the similarities, see Glennan 2010, this volume, Chap. 13). In addition, there might be convincing arguments available for why historical processes should be conceived as wholes that are located on a higher ontological level and that bring about or lead to the explanandum event (even if most of them do not regularly do so).[31] For instance, historians may be social realists and believe that there exist social processes (e.g., economic decline) or social systems (like the Roman Empire or the Credit Suisse) and that they are in a sense located on a higher ontological level than individuals and their actions. Furthermore, individualist historians might want to argue that when it comes to the explanation of historical events and social processes, one has to focus on a lower level. That is, one has to go down to the level of interacting people in order to explain the behavior of the whole (for recurring debates around notions such as "social process," "social structure," and "social system" and "history," see Plenge, this volume, Chap. 10). If historians want to explain the stability of some system, they might even want to refer to social processes that are somewhat regular (i.e., not unique and not totally contingent), like production in a factory or training in a sports team. Furthermore, biological processes like gastrulation, neurulation, or other developmental processes are also mere sequences of certain kinds of events (which, however, proceed regularly). Nevertheless, they are often referred to as wholes (sometimes even as mechanisms) that are located on a higher level of organization as their parts. So why should historical processes not be characterized

[29] Most philosophers of history use terms such as "historical process" in an innocent way. However, at second sight, it becomes clear that these concepts can be problematic (e.g., because they might imply a difference between "historical" processes and something else, e.g., "natural" processes) and need to be specified.

[30] Actually, "narrativist" philosophers of history would claim that the "wholes" historians claim to investigate are literally artifacts constructed in the narrative, which do not represent anything real. For counterpositions to "narrativism," see Gerber (this volume, Chap. 9) and Scholz (this volume, Chap. 10).

[31] Even if such arguments were not at hand, one could still stick to the claim that narrative explanations in history/historiography are a special kind of mechanistic explanations. One only needs to agree with Carl Craver that there exist types of mechanistic explanations that are not *constitutive* mechanistic explanations but rather *etiological* mechanistic explanations (i.e., descriptions of the antecedent causes of the explanandum event; Craver 2007, p. 107).

as wholes, too? In current philosophy of history, ontological inquiries like this are not of high repute. However, some authors in this volume take steps towards rehabilitating ontological issues in philosophy of history (cf. Gerber, this volume, Chap. 9; Plenge, this volume, Chap. 10; and Scholz, this volume, Chap. 11).

The second respect in which the concept of a narrative explanation (in the sense explicated above) seems to be similar to the concept of a mechanistic explanation is that both of them stress the importance of describing the *continuity* between the components of a mechanism or historical process. Revealing this continuity is essential to the explanatory power of both kinds of explanation. Mechanistic explanations in biology represent how one stage of a mechanism gives rise to another and how one activity of an entity causes another activity of another entity (e.g., how the transport of the mRNA from the nucleus into the cytoplasm enables the binding of the ribosome subunits, which in turn causes the start of the translation). Similarly, narrative explanations in history/historiography describe how one event leads to another via one or many processes (e.g., how the implementation of a new policy by constructing new social systems and thereby instigating myriads of individual activities leads some social groups into disaster). What some philosophers of history call "continuous series of events" is called "productive continuity" (Machamer et al. 2000, p. 3) by philosophers of biology.[32]

In sum, revealing the similarities between mechanistic explanations in biology and narrative explanations in history/historiography seems to be a promising, although not unproblematic, way to question the traditional opposition between a "scientific" way of representing and explaining the world, on the one hand, and a specific "historical" mode of describing and understanding the world, on the other.

Third, a possible challenge to the assumption that historical explanation is a special kind of mechanistic explanation is the claim that historical explanations explain *particular* events, whereas mechanistic explanations explain how a certain *type* of event or behavior (also called phenomenon) is *regularly* produced by a mechanism. In short, one might hold that:

> The explananda of historical explanations are *tokens*, whereas mechanistic explanations explain *types*.

According to this view, there would be no mechanistic explanation at all; neither of how Michael bumped his Ferrari into Ralf's Toyota nor of why Michael's Ferrari with which he won the Monaco Gran Prix in 1999 worked properly (by contrast,

[32]The only important difference in this context, which should not be swept under the table, concerns the *organization* of the parts of biological mechanisms and of historical processes. Whereas the events described in narrative explanations are always ordered sequentially (at least if they are token events), the entities and activities described in mechanistic explanations in biology frequently do exhibit more complex forms of organization (like positive and negative feedback).

both explanations could perhaps be characterized as historical explanations). Instead, a mechanistic explanation describes, for instance, how the Ferrari as a type of car works or how Ferraris like Michael's behave.

This challenge should be taken particularly serious because the view that historians are, by essence, concerned with the description of idiographic detail or concrete phenomena is widespread in the philosophy of history and elsewhere. It is often claimed that they would otherwise lose membership in their profession.[33] Following Popper,[34] Gordon Graham contrasts historical explanation with what he calls "theoretical explanation" (1983, 48f).[35] According to Graham, theoretical scientists are, in contrast to historians, concerned with disclosing general patterns or with finding out and explaining how things regularly work.[36] Aviezer Tucker (2004; this volume, Chap. 16) adopts a similar position. One of his main theses is that what historians explain is *token evidence* (e.g., particular documents or fossils) and *token events* (i.e., events that are "unique and unrepeatable"; Sober 1988, p. 78), like the Rise of Rome or the assassination of Kennedy. Contrary to the historical sciences, so Tucker, the "theoretical sciences" are not concerned with token evidence and events, but rather with theoretical types of replicated evidence and repeated events.[37]

Other authors adopt a more pluralistic position and allow for a diversity of explananda of historical explanation. In this line, Leuridan and Froeyman (2012) distinguish three possible kinds of explananda (or "aspects"; 2012, p. 183) of historiographic explanations: singular events, types of events (which they call "general historical events"; 2012, p. 183), and historical evidence. Hence, the above thesis that the explananda of historical explanations are restricted to tokens, whereas the explananda of mechanistic explanations are types, is far from being uncontested. This objection might even be strengthened. We can ascribe such objections also to Mario Bunge. He explicitly states that there exists "historical explanation[s] of laws" (1998, p. 43), which he characterizes as one of two existing types of "mechanismic explanation."

> Historical explanation (. . .) consists in the tracing of the *evolution of a law*, by showing how it arose in the course of time from patterns characterizing earlier stages in an evolutionary process, as when a new pattern of social behavior is given a historical explanation. (Bunge 1998, vol. II, 38. On Bunge's theory of mechanisms and mechanismic explanation, see his 1997 and 2004)

---

[33] Even thinkers such as Mandelbaum (1977) advocated this position.

[34] "Now the sciences which have an interest in specific events and their explanation may, in contradistinction to the generalizing sciences, be called the historical sciences." (Popper 1974, 447f).

[35] Goudge points to a similar difference when he distinguishes between "systematic" and "historical" modes of explanation (1961, p. 62).

[36] Of course, this does not preclude that these patterns or processes may be "historical" in the sense that they are the result or outcome of some preceding process (e.g., adaptive evolution).

[37] The thesis that historical explanations explain singular occurrences is popular among narrativists, too. They frequently emphasize the *uniqueness* of the explananda of historical explanations (e.g., Goudge 1961, p. 77).

However, in order to reject the thesis that historical explanations explain solely tokens, one need not quarrel about Bunge's use of the term "law." It suffices to point out that sometimes historians are interested in explaining more than unique, singular occurrences, namely, *types* of events (like historical patterns or trends; see also Little 2010). For instance, they explain why absolutist states did well in collecting taxes, or how a Roman emperor managed the finances.

But don't we implicitly alter the question at this point? One might argue that what some token historians do does not affect the answer to the question what a *historical* explanation (as a specific type of explanation) is. In other words, one might say that if a historian explains a historical regularity or law by modeling the mechanism that is responsible for that behavior, he simply does not explain "historically" but provides a different kind of explanation. Accordingly, the set of all historical explanations would overlap but not coincide with the set of all explanations that historians provide. However, the thesis that historians also offer nonhistorical explanations is at least a debatable conclusion (for a "liberal" theory of scientific explanation that stresses the diversity of significant research questions in science *and* history alike, see Kitcher and Immerwahr, this volume, Chap. 14).

The above thesis that the explananda of historical explanations are tokens, whereas mechanistic explanations explain types, can also be criticized by questioning its second part. That is, one might argue that although mechanistic explanations explain types of phenomena more frequently, this must not and is not solely the case. The explanatory practice in the life sciences reveals mechanistic explanations of singular occurrences, too (cf. Glennan 2010, this volume, Chap. 13). Examples are mechanistic explanations of how the genetic disease of a particular patient causes certain symptoms or of how a particular mutation brought about a third leg on the back of an individual *Drosophila melanogaster*.

In conclusion, the aim of this section was not to judge whether historical explanations are a special kind of mechanistic explanations or not. Rather, we wanted to show that much can be said in favor of it but also to hint at the problems with such a view. All in all, the questions of what a specifically historical explanation is and what the similarities and differences between historical explanations and mechanistic explanations in the life sciences are constitute a promising field for future philosophical research.

1.3 Conclusion

In our view, the project of bringing together debates about explanation from philosophy of biology and from philosophy of history is a fruitful one, and the reservations that one might have against it can be rebutted. In order to show this, we identified some major points of contact between these disciplines which had not been obvious at first sight.

The claim, for instance, that some biological explanations (e.g., evolutionary explanations) are historical in character requires an answer to the question of

what makes an explanation specifically historical. Is it the fact that they (at least implicitly) invoke historical laws? Or is their historical character due to their status of being narrations, that is, descriptions of continuous series of events that together form an intelligible whole? Examining the peculiarities of historical explanations gives rise to further interesting questions that lie at the intersection between philosophy of biology and philosophy of history. Are historical explanations "just-so stories" or mere how-possible explanations that should be avoided in the biological science? Or are they special kinds of causal explanations and, if yes, what makes them special?

The second major point of contact that we identified concerned the debate about mechanisms. Is it plausible to claim that both special sciences, the biological and the historical sciences, aim at discovering mechanisms and provide mechanistic explanations? Is the difference between biological mechanisms and historical/social mechanisms just one of degree (e.g., different degrees of stability), or do they constitute fundamentally different kinds of mechanisms (if there can be found mechanisms in history at all)? What are the similarities and differences between narrative explanations in historical science and mechanistic explanations in biology, and do the similarities warrant characterizing historical explanation as a subtype of mechanistic explanation?

We do not claim that these are the only overlaps of interests, debates, and problems. Rather, we hold that they provide a good starting point for discussion. Many of the issues that we raised are developed more thoroughly in the contributions to this volume. Moreover, the contributions address several equally interesting topics that we could not approach in this introduction.

References

Athearn, D. (1994). *Scientific nihilism. On the loss and recovery of physical explanation*. Albany: State University of New York Press.

Ball, T. (1972). On 'historical' explanation. *Philosophy of the Social Sciences, 2*, 181–192.

Bechtel, W. (2006). *Discovering cell mechanisms. The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (2008). *Mental mechanisms. Philosophical perspectives on cognitive neuroscience*. New York/London: Taylor and Francis Group.

Bhaskar, R. (1979). *The possibility of naturalism. A philosophical critique of the contemporary human sciences*. Brighton: The Harvester Press.

Brodbeck, M. (1962). Explanation, prediction, and "imperfect" knowledge. In F. Herbert & M. Grover (Eds.), *Minnesota studies in the philosophy of science, Volume III: Scientific explanation, space, and time* (1962nd ed., pp. 231–272). Minneapolis: University of Minnesota Press.

Brzechczyn, K. (2009). Between science and literature: The debate on the status of history. In K. Brzechczyn (Ed.), *Idealization XIII: Modeling in history*. Amsterdam: Rodopi.

Bunge, M. A. (1997). Mechanism and explanation. *Philosophy of the Social Sciences, 27*, 410–465.

Bunge, M. A. (1998). *Philosophy of science* (Vol. 2). New Brunswick: Transaction.

Bunge, M. A. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences, 34*, 182–210.

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science, 69*, 474–496.
Cleland, C. E. (2009). Philosophical issues in natural history and its historiography. In A. Tucker (Ed.), *A companion to the philosophy of history and historiography* (pp. 44–62). Oxford: Blackwell.
Cleland, C. E. (2011). Prediction and explanation in historical natural science. *British Journal of Philosophy of Science, 62*(3), 551–582.
Collingwood, R. G. (1994[1946]). *The idea of history*. Oxford: Oxford University Press.
Craver, C. F. (2007). *Explaining the brain. Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon.
Danto, A. C. (1965). *Analytical philosophy of history*. Cambridge: Cambridge University Press.
Day, M. (2009). *Competing explanations. Exclusion and importance in historical accounts*. Saarbrücken: VDM Verlag.
Demeulenaere, P. (Ed.). (2011). *Analytical sociology and social mechanisms*. Cambridge: Cambridge University Press.
Dray, W. H. (1954). Explanatory narrative in history. *The Philosophical Quarterly, 4*, 15–27.
Dray, W. H. (1957). *Laws and explanation in history*. Oxford: Clarendon.
Dray, W. H. (1963). The historical explanation of actions reconsidered. In S. Hook (Ed.), *Philosophy and history. A symposium* (pp. 105–135). New York: New York University Press.
Dray, W. H. (2000). Explanation in history. In J. H. Fetzer (Ed.), *Science, explanation, and rationality. Aspects of the philosophy of Carl G. Hempel*. Oxford: Oxford University Press.
Frings, A. (2008). Erklären und Erzählen: Narrative Erklärungen historischer Sachverhalte. In J. Marx & A. Frings (Eds.), *Erzählen, Erklären, Verstehen. Beiträge zur Wissenschaftstheorie und Methodologie der Historischen Kulturwissenschaften* (pp. 129–164). Berlin: Akademie-Verlag.
Gallie, W. B. (1964). *Philosophy and the historical understanding*. London: Chatto & Windus.
Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*(1), 49–71.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(3S), 342–353.
Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*, 251–266.
Gorman, J. (2007). *Historical judgement. The limits of historiographic choice*. Stocksfield: Acumen.
Goudge, T. H. (1961). *The ascent of life. A philosophical study of the theory of evolution*. London: Allen & Unwin.
Gould, S. J. (2000). *Wonderful life. The burgess shale and the nature of history*. London: Vintage.
Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society London B, 205*, 581–598.
Graham, G. (1983). *Historical explanation reconsidered*. Aberdeen: Aberdeen University Press.
Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.
Hempel, C. G. (1942). The function of general laws in history. *The Journal of Philosophy, 39*, 35–48.
Hempel, C. G. (1962). Explanation in science and in history. In R. G. Colodny (Ed.), *Frontiers of science and philosophy* (pp. 7–34). Pittsburgh: Pittsburgh University Press.
Hempel, C. G. (1963). Reasons and covering laws in historical explanation. In S. Hook (Ed.), *Philosophy and history. A symposium* (pp. 143–163). New York: New York University Press.
Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: The Free Press.
Hull, D. L. (1975). Central subjects and historical narratives. *History and Theory, 14*, 253–274.
Hull, D. L. (1989). *The metaphysics of evolution*. Albany: State University of New York Press.
Illari, P. M., & Williamson, J. (2010). Function and organization: Comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences, 41*, 279–291.

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*(1), 119–135.
Jenkins, K. (1991). *Re-thinking history*. London: Routledge.
Kosso, P. (2001). *Knowing the past. Philosophical issues of history and archaeology*. Amherst, NY: Humanity Books.
Leuridan, B., & Froeyman, A. (2012). On lawfulness in history and historiography. *History and Theory, 51*, 172–192.
Little, D. (2010). *New contributions to the philosophy of history*. Dordrecht: Springer.
Lloyd, C. (1986). *Explanation in social history*. Oxford: Blackwell.
Louch, A. R. (1969). History as narrative. *History and Theory, 8*, 54–70.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Mandelbaum, M. (1961). Historical explanation. The problem of 'covering laws'. *History and Theory, 1*, 229–242.
Mandelbaum, M. (1977). *Anatomy of historical knowledge*. Baltimore: Johns Hopkins University Press.
Manicas, P. T. (2006). *A realist philosophy of social science. Explanation and understanding*. Cambridge: Cambridge University Press.
Martin, R. (1977). *Historical explanation. Re-enactment and practical inference*. Ithaca: Cornell University Press.
Martin, M. (2000). *Verstehen: The uses of understanding in social science*. New Brunswick: Transaction.
Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Harvard University Press.
McCullagh, C. B. (1998). *The truth of history*. London: Routledge.
Mink, L. O. (1966). The autonomy of historical understanding. *History and Theory, 5*(1), 24–47.
Munslow, A. (2007). *Narrative and history*. Basingstoke: Palgrave Macmillan.
Murphey, M. G. (2009). *Truth and history*. Albany: State University of New York Press.
Norkus, Z. (2005). Mechanisms as miracle makers? The rise and inconsistencies of the 'mechanismic approach' in social science and history. *History and Theory, 44*, 348–372.
Norkus, Z. (2007). Troubles with mechanisms: Problems of the 'mechanistic turn' in historical sociology and social history. *Journal of the Philosophy of History, 1*, 160–200.
Popper, K. R. (1974). *The open society and its enemies. Volume II: The high tide of prophecy: Hegel, Marx, and the aftermath*. London: Routledge.
Roberts, G. (Ed.). (2001). *The history and narrative reader*. London: Routledge.
Rosenberg, A. (2001). How is biological explanation possible? *British Journal for Philosophy of Science, 52*, 735–760.
Rosenberg, A. (2006). *Darwinian reductionism. Or, how to stop worrying and love molecular biology*. Cambridge: University of Chicago Press.
Schmid, M. (2006). *Die Logik mechanismischer Erklärungen*. Wiesbaden: VS-Verlag.
Scriven, M. (1959). Truisms as the grounds for historical explanations. In P. Gardiner (Ed.), *Theories of history. Readings from classical and contemporary sources* (pp. 443–475). New York: Free Press.
Skipper, R. A., & Millstein, R. L. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in the History and Philosophy of Biological and Biomedical Sciences, 36*, 327–347.
Sober, E. (1988). *Reconstructing the past: Parsimony, evolution, and inference*. Cambridge, MA: MIT Press.
Tilly, C. H. (2004). Social boundary mechanisms. *Philosophy of the Social Sciences, 34*, 211–236.
Tilly, C. H. (2008). *Explaining social processes*. Boulder: Paradigm.
Topolski, J. (1976). *Methodology of history*. Dordrecht: D Reidel.
Tucker, A. (2004). *Our knowledge of the past. A philosophy of historiography*. Cambridge: Cambridge University Press.

Wan, P. Y.-z. (2011). *Reframing the social. Emergentist systemism and social theory*. Farnham: Ashgate.
White, M. G. (1943). Historical explanation. *Mind, 52*, 212–229.
White, M. G. (1963). The logic of historical narration. In S. Hook (Ed.), *Philosophy and history. A symposium* (pp. 3–31). New York: New York University Press.
Wisdom, J. O. (1987). *Philosophy of the social sciences I: A metascientific introduction*. Aldershot: Avebury.

Part I
General Issues on Explanation

Chapter 2
The Ontic Account of Scientific Explanation

Carl F. Craver

Abstract According to one large family of views, scientific explanations explain a phenomenon (such as an event or a regularity) by subsuming it under a general representation, model, prototype, or schema (see Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421–441; Churchland, P. M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge: MIT Press; Darden (2006); Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation* (pp. 331–496). New York: Free Press; Kitcher (1989); Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25). My concern is with the minimal suggestion that an adequate philosophical theory of scientific explanation can limit its attention to the format or structure with which theories are represented. The *representational subsumption view* is a plausible hypothesis about the psychology of understanding. It is also a plausible claim about how scientists present their knowledge to the world. However, one cannot address the central questions for a philosophical theory of scientific explanation without turning one's attention from the structure of representations to the basic commitments about the worldly structures that plausibly count as explanatory. A philosophical theory of scientific explanation should achieve two goals. The first is *explanatory demarcation*. It should show how explanation relates with other scientific achievements, such as control, description, measurement, prediction, and taxonomy. The second is *explanatory normativity*. It should say when putative explanations succeed and fail. One cannot achieve these goals without undertaking commitments about the kinds of ontic structures that plausibly count as explanatory.

C.F. Craver (✉)
Department of Philosophy, Washington University in St. Louis, One Brookings Drive, St. Louis 63130-4899, MO, USA
e-mail: ccraver@artsci.wustl.edu

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_2,

Representations convey explanatory information about a phenomenon when and only when they describe the ontic explanations for those phenomena.

Keywords Scientific explanation • Models • Representation • Mechanism • Laws • Demarcation • Normativity

2.1 Introduction

According to one large family of views, scientific explanations essentially subsume a phenomenon (or its description) under a general representation (see Hempel 1965; Kitcher 1981, 1989; Churchland 1989; Bechtel and Abrahamsen 2005; Machamer et al. 2000). Authors disagree about the precise form that these representations should take: For Carl Hempel they are generalizations in first-order logic; for Philip Kitcher they are argument schemas; for Bechtel and Abrahamsen they are mental models; for Churchland they are prototype vectors; and for Machamer, Darden, and Craver they are mechanism schemas. Here, my focus is on the basic assumption that the philosophical dispute about scientific explanation is, or should be, about the representational form that such explanations take. While this *representational subsumption view* (RSV), in all of its guises, will likely be part of any theory of human understanding, the RSV is precisely the wrong place to begin developing a philosophical theory of explanation. Or so I shall argue.

Two philosophical objectives have been central to the philosophical debate over the nature of scientific explanation for over 50 years. The first is *explanatory demarcation*: the theory should distinguish explanation from other forms of scientific achievement. Explanation is one among many kinds of scientific success; others include control, description, measurement, prediction, and taxonomy. A theory of explanation should say how explanatory knowledge differs from these others and should say in virtue of what particular kinds of knowledge count as explanatory. The second goal is *explanatory normativity*. The theory should illuminate the criteria that distinguish good explanations from bad. The term "explanation" should not be an empty honorific; the title should be earned. A philosophical theory of explanation should say when the title is earned. My claim is that in order to satisfy these two objectives, one must look beyond representational structures to the ontic structures in the world. Representational subsumption, in other words, is insufficient as an account of scientific explanation. The fundamental philosophical dispute is ontic: it concerns the kinds of ontic structure that ought to populate our explanatory texts, whatever their representational format.

Some caveats will hopefully prevent misunderstandings. First, I do not claim that one can satisfy all of the normative criteria on explanatory models, texts, or communicative acts by focusing on ontic explanations alone. Clearly, there are questions about how one ought to draw diagrams, organize lectures, and build elegant and useable models that cannot be answered by appeal to the ontic structures themselves. The ontic explanatory structures are in many cases too complex, reticulate, and laden with obfuscating detail to be communicated directly.

Scientific explanations are constructed and communicated by limited cognitive agents with particular pragmatic orientations. These topics are interesting, but they are downstream from discussions of what counts as an explanation for something else. Our abstract and idealized representations count as conveying explanatory information in virtue of the fact that they represent certain kinds of ontic structures (and not others). Second, my topic is independent of psychological questions about the kinds of explanation that human cognitive agents tend to produce or tend to accept. Clearly, people often accept as explanations a great many things that they should reject as such. And people in different cultures might have different criteria for accepting or rejecting explanations. These facts (if they are facts) would be fascinating to anthropologists, psychologists, and sociologists. But they are not relevant to the philosophical problem of stating when a scientific explanation ought to be accepted as such. In the view defended here, scientific explanation is a distinctive kind of achievement that cultures and individuals have to learn to make. Individual explanatory judgments, or cultural trends in such, are not data to be honored by a normative theory that seeks to specify when such judgments go right and when they go wrong. Finally, I do not suppose that there is one and only one form of scientific explanation. Though at times I adopt a specifically causal-mechanical view of explanation (see Craver 2007), and so will describe the ontic structures involved in explanation as causal or mechanistic, I intend the term *ontic structure* to be understood much more broadly. Other forms of ontic structure might include attractors, final causes, laws, norms, reasons, statistical relevance relations, symmetries, and transmissions of marks, to name a few. The philosophical dispute about explanation, from this ontic perspective, is about which kinds of ontic structure properly count as explanatory and which do not.

I proceed as follows. In Sect. 2.2, I disambiguate four ways of talking about explanation: as a communicative act, as a representation or text, as a cognitive act, and as an objective structure. The goals of that discussion are to distinguish these senses of explanation and to highlight some distinctive conceptual contributions that the ontic conception makes to our speaking and thinking about explanations. In Sect. 2.3, I illustrate how appeal to ontic explanations is essential for marking several crucial normative dimensions by which scientific explanations are and ought to be evaluated: the distinction between how-possibly and how-actually explanations, the distinction between phenomenal descriptions and explanations, the difference between predictive and explanatory models, and the requirement that explanatory models should include all and only information that is explanatorily relevant to the phenomenon one seeks to explain. In Sect. 2.4, I review how these normative dimensions long ago raised problems for Hempel's covering-law model, the once dominant idea that explanations are arguments (texts) with a description of the *explanandum* phenomenon as their conclusion. In Sect. 2.5, I use Churchland's PDP model of explanation as an exemplar of psychologistic theories to illustrate how cognitivist models of explanation presuppose, rather than satisfy, the normative distinctions laid out in Sect. 2.4. In Sect. 2.6, I show how the ontic conception provides a satisfyingly simple answer to the question: How can idealized models explain?

2.2 The Ambiguities of "Explanation"

Consider four common modes in which people (including scientists) talk about explanation. Suppose, thinking of one's neuroscience professor, one says:

> (S1) Jon explains the action potential (Communicative Mode).

One might imagine Jon in front of a classroom, writing the Hodgkin and Huxley model of the action potential on a chalkboard. Alternatively, we might imagine him writing a textbook that walks, step-by-step, through the complex mechanisms that give rise to action potentials. Explanation, so understood, is a communicative act. It involves an explainer, an audience, and a text (a lecture or book, in this case) that conveys information from the explainer to the audience. If everything goes right, Jon manages in his lecture to convey information about action potentials to an audience, and the audience comes to understand how action potentials are produced.

Explanatory communications of this sort might fail in at least three ways. First, Jon might successfully deliver a false explanation. He might explain (incorrectly) that action potentials are produced by black holes in the endoplasmic reticulum. We can imagine excited students understanding Jon's lecture and dutifully reporting it back on the exam. Jon explains the action potential to the class (i.e., he gave them a model of action potential generation), but the explanation is false. In a second kind of failure case, Jon unsuccessfully delivers a true explanation. He might, for example, give an impeccably accurate lecture about the action potential but leave his students completely confused. The lecture fails, we might suppose, because it presupposes background knowledge the students lack, or because it is delivered in a language the students are unprepared to handle. Finally, we might imagine Jon delivering an impeccably organized and conversationally appropriate lecture to undergraduate students who, because they are distracted by other plans, fail to understand what Jon is telling them. The explanation fails as a communicative act, but it is not Jon's fault. His audience just did not get it.

In contrast to this communicative mode, we sometimes talk about explanation in the *ontic mode*, as a relation among features of the world. One says, for example, that:

> (S2) The flux of sodium (Na^+) and potassium (K^+) ions across the neuronal membrane explains the action potential (Ontic Mode).

S2 is not at all like S1. In cases like S2, the items in the subject position are not intentional creatures (like Jon); they are states of affairs. And no text about a topic is transmitted from an explainer to an audience. The explanatory relation described in S2 is not properly fleshed out in terms of the delivery of information

via a text to an audience. There is no text, no representation, no information (in the colloquial sense).[1] It would appear, in fact, that S2 could be true even if no intentional creature knows or ever knew the fact that S2 expresses. The term "explains" in S2 is synonymous with a description of the kinds of factors and relations (ontic structures) that are properly taken to be explanatory; as noted above, examples include causes, causal relevance, components, laws, and statistical relevance relations. Wesley Salmon expressed precisely this contrast as follows:

> The linguistic entities that are called 'explanations' are statements reporting the actual explanation. Explanations, in this [ontic] view, are fully objective and, where explanations of nonhuman facts are concerned, they exist whether or not anyone ever discovers or describes them. Explanations are not epistemically relativized, nor (outside of the realm of human psychology) do they have psychological components, nor do they have pragmatic dimensions. (Salmon 1989, p. 133)

Salmon credits Coffa (1974) with this insight and notes that even Hempel, at times, could be read as embracing the view that laws themselves (rather than law statements or generalizations, which are representations of laws) provide ontic explanations for explanandum events and regularities. As Coffa explains, Hempel's deductive-nomological formulation of the covering-law model is susceptible of either an ontic or an epistemic interpretation. However, his inductive-statistical formulation has, at bottom, an irreducible epistemic component. This is because Hempel defines the relevant probabilities in such explanations relative to the presumed background knowledge of the scientists. For Coffa, the need to relativize what counts as an explanation to what people know or believe was a major strike against the account. In his view, "no characterization of inductive explanation incorporating that feature [epistemic relativization] can be backed by a coherent and intelligible philosophy of explanation" (1974, p. 57). The ontic mode captures this objective way of talking about explanation.[2]

It is worth emphasizing that the term "explanation" in S1 and S2 is ambiguous, as revealed by the inability to meaningfully combine the two sentences into one, as in:

(S1 + 2) Jon and the flux of Na^+ and K^+ ions across the neuronal membrane explain the action potential.

[1] I do not know precisely how to specify the kind of mind-dependence I intend to exclude without also excluding causal interactions involving intentional phenomena that seem to me perfectly legitimate in explanations: that Jill ducked because she saw the looming object. Nor do I intend to exclude notions of information fully specified in causal or statistical form, and so independently of human interpretation. Yet perhaps I have said enough to gesture in the direction of a more adequate formulation.

[2] The German verb "erklären" is not ambiguous like the English word "explanation." The verb contains the idea of "making clear," which automatically suggests the communicative or representational mode.

If we think of explanation primarily in the communicative mode, (S1 + 2) appears odd because Na+ and K+ ions do not deliver lectures or produce diagrams with the intention of delivering information to an audience. If we think of explanation in the ontic mode, as, for example, a matter of producing, constituting, or otherwise being responsible for the *explanandum* phenomenon, then (S1 + 2) appears odd because it appears to assert that Jon causes, produces, or otherwise is responsible for the generation of action potentials (generally), which is clearly false. For these reasons, it would be a kind of conceptual mistake to think that an analysis of "explanation" in the sense expressed in S1 could serve as an analysis of "explanation" in the sense expressed in S2. This is not to say the two are unrelated. In particular, whether Jon has provided his class with a correct explanation of the action potential would seem to depend on whether Jon's lecture correctly indicates how action potentials are produced or constituted. The endoplasmic black hole hypothesis makes this clear. That is, whether or not Jon's explanatory communicative act (described in S1) fails in the first sense described above will depend on whether his text matches (to a tolerable degree) the patterns of causation, constitution, and responsibility that *in fact* explain the production of action potentials (as described in S2).

To explore this connection a bit further, consider a third mode of thinking about explanations. Where S1 places Jon, the communicative agent, in the subject position, and S2 places worldly states of affairs in the subject position, this third way of speaking puts Jon's explanatory text in the subject position:

(S3) The Hodgkin-Huxley (HH) model explains the action potential (Textual Mode).

The HH model of the action potential, one of the premier theoretical achievements in the history of neuroscience, is a mathematical model that describes how the membrane voltage of a neuron changes as a function of ionic conductances and how ionic conductances change as a function of voltage and time. My point does not turn on the fact that a specifically *mathematical* model appears in the subject position; rather, S3 is meant to apply generally to any text: it might be an article, book, cartoon, diagram, film, graph, or a lecture. A text, in this sense, is a vehicle for conveying intentional content from a communicator to an audience. Hodgkin and Huxley communicated their understanding of the current-voltage relations in neuronal membranes to the rest of us in the form of a compact mathematical representation from which we (the audience) might extract a wealth of pertinent information about this topic.

Yet it would be a mistake to put John and the HH model together as the conjoined subjects of a sentence such as:

(S1 + 3) Jon and the HH model explain the action potential.

because they explain the action potential in different ways: Jon as a communicative agent, and the HH model as a communicative text. It would not be a confusion of this sort to assert that Jon, Hodgkin, and Huxley explained the action potential to the class (perhaps Jon invited some illustrious guests). Nor would it be confused to claim that Jon's lecture, the equivalent circuit diagram, and the HH model explained the action potential. In these last two sentences, the term "explanation" applies univocally to the three objects listed in the subject position.

It would be a confusion, however, to put the HH model and ionic fluxes together as the conjoined subjects of a sentence such as:

(S2 + 3) Ionic fluxes and the HH model explain the action potential.

for they again explain the action potential in different ways: ionic fluxes, as states of affairs that produce or constitute action potentials, and the HH model as a communicative text. Equations do not produce action potentials, though action potentials and their mechanisms can be described using equations. The HH model might be included in the explanatory text, but the equation is neither a cause nor a constituent of action potentials. This confusion is propagated by those who think of the HH model as a "law" that "governs" the action potential (e.g., Weber 2005) rather than as a mathematical generalization that describes how some of the components in the action potential mechanism behave (see Bogen 2005; Craver 2006, 2007). To put the point the other way around, it would be wrong to claim that Jon used ionic fluxes to explain action potentials to the class (unless, e.g., he were to illustrate the process of diffusion by placing dye in the bottom of a beaker, in which case the demonstration becomes a "text" that is intended to convey information to a class).

Finally, let us consider a more mentalistic way of speaking about explanation (or, less awkwardly, about understanding). We might think that a cognitive agent explains/understands a phenomenon by activating a mental model that in some sense fits the phenomenon to be explained. Churchland speaks of explanations, as I discuss below, as involving the activation of a prototype in the connectionist networks of one's brain. Similarly, Bechtel and Abrahamsen insist that explanation is "essentially a cognitive activity." Directly contrary to the reading in S2, they claim that what figures in explanation is not "the mechanisms in the world" but "representations of them" (Bechtel and Abrahamsen 2005, p. 425). To express this very reasonable thought, we should recognize a fourth way of speaking:

(S4) Jon's mental representation of the mechanism of the action potential explains the action potential (Cognitive Mode).

S4 is no doubt a bit strained to the native English speaker's ear. Typically, in these situations, we would speak not of explanation but of understanding. Jon, in this case, *understands* the action potential when Jon can activate a mental representation of the requisite sort and, for example, answer questions about how the action potential might differ depending on different changes in background conditions, ion concentrations, distributions of ion channels, cell morphology, and the like. But let us put this worry aside until the next section in order to draw out some important differences between S4 and the others.

The subject position of S4 is occupied by a mental representation. The subject is not a cognitive agent but, as it were, a part or sub-process of the agent's cognitive architecture. The mental representation itself has no communicative intentions of the sort that Jon has. And it is hard to make sense of the idea that such a representation has an audience with which it is attempting to communicate. Though mental representations are said to influence one another, subsume one another, and the like, it would be an illegitimately homuncular sort of thought to say that, for example, one mental representation understands the other. There is nobody "in John's head" to read the representation and understand it. And it would be wrong to say that the mental representation explained something to Jon (in the sense of S1), since Jon, as the possessor of the mental representation, already understands it quite well. Thus, it seems to me something of a category mistake to assert that:

(S1 + 4) Jon and his mental representations explained the action potential.

though it is at least plausible to say that Jon is able to explain the action potential to the class in virtue of his having a set of stored mental representations about action potentials, the mental representations and Jon explain in different ways. Likewise, for reasons we have already discussed, it would be a mistake to assert that:

(S2 + 4) Jon's mental representations and ion fluxes explain the action potential.

Jon's mental representations do not cause or produce action potentials (unless they drive him to do some electrophysiological experiments). And although ionic fluxes are certainly involved in the production of Jon's mental representations (given that such representations must be implemented somehow in neural architectures), the ionic fluxes in Jon's brain do not subsume action potentials, as do his abstract representations of the action potential mechanisms. S4 is clearly closest to the textual reading of explanation statements:

(S3 + 4) Jon's mental representations and the HH model explain the action potential.

Rightly or wrongly, many think of mental representations as texts or images written in the mind or in our neural architectures. If so, it is easy to think that mental representations and scientific representations might play the same kind of role and might be involved in the same kind of explanatory process. There are important differences, however, between these two kinds of explanation. First, the HH model might explain the action potential even if Jon never learns it. In S4 we are concerned with cognitive achievements of a single mind, not with the explanatory advance of a science (as appears to be the concern of S3). The HH model "covers" or "subsumes" many features of the action potential regardless of whether Jon ever hears about it. For the HH model to be relevant to Jon's understanding of the action potential, let us allow, he has to form an internal representation of something like the HH model and activate it. But the HH model itself does not need to be "activated" to count as an explanation. Even if (implausibly enough) there is a brief moment in time when nobody in the world is thinking about the HH model in relationship to action potentials, there remains a sense in which the HH model continues to explain the action potential during our cognitive slumbers (if, indeed, the HH model is an explanation of the action potential; a topic to which I return below).

The simple point is that the term "explanation" has four common uses in colloquial English: (1) to refer to a *communicative act*, (2) to refer to a cause or a factor that is otherwise responsible for a phenomenon (the *ontic* reading), (3) to refer to a *text* that communicates explanatory information, and (4) to refer to a *cognitive* act of bringing a representation to bear upon some mysterious phenomenon. These uses are no doubt related. Explainers (we might suppose) understand a phenomenon in virtue of having certain *cognitive* representations, and they use explanatory *texts* (such as the HH model) to represent *ontic* explanations (such as the production of action potentials by ionic fluxes) in order to *communicate* that understanding to an audience. Though these senses of "explanation" are subtly related to one another, they are not so subtly *different* senses of explanation. It would be a mistake to conflate them.

In particular, there is an especially clear line between S2, on the one hand, and S1, S3, and S4 on the other. S1, S3, and S4 each depend in some way on the existence of intentional agents who produce, interpret, manipulate, and communicate explanatory texts. Jon's communicative act of explanation presupposes a communicator and an audience. The HH model is a scientific text produced, learned, and applied by intentional agents in the act of discovering, explaining, and understanding action potentials. It is called a model in part because it is a representation that intentional creatures can use for the purposes of making inferences about a worldly system. And Jon's internal representation, or mental image, is likewise dependent for its existence on Jon's being the kind of creature that thinks about things. I suppose

it is possible to understand the term "model" in a more technical, logical, or set-theoretic sense, but even in this technical reading the notion depends for its existence on creatures that are able to, for example, form inferences and apply general frameworks in specific instances.

S2, the ontic mode of thinking about explanation, does not depend on the existence of intentional agents in this way. A given ontic structure might cause, produce, or otherwise be responsible for a phenomenon even if no intentional agent ever discovers as much. This ontic way of talking about explanation allows us to express a number of reasonable sentences that would be strained, if not literal nonsense, if our thinking about explanation were tied to the modes expressed in S1, S3, and S4. Here are some examples:

(A) Our world contains undiscovered phenomena that have explanations.
(B) There are known phenomena that we cannot currently explain (in the sense of S1, S3, or S4) but that nonetheless have explanations.
(C) A goal of science is to discover the explanations for diverse phenomena.
(D) Some phenomena in our world are so complex that we will never understand them or model them, but they have explanations nonetheless.

If we tie our thinking about explanation to the existence of creatures that are able to represent, communicate, and understand phenomena, each of these sentences is awkward or nonsensical. If one allows for an ontic way of thinking about explanation, however, each of these sentences is relatively straightforward and non-elliptical. (A) concerns aspects of the world that nobody has ever represented or that nobody ever will represent. If explanation requires representation by an intentional agent, then this should not be possible. (B), (C), and (D) also recognize a distinction between whether or not a phenomenon has an explanation, on the one hand, and whether anyone knows or can otherwise construct the explanation for it, on the other. (A)–(D) are very natural things to say.

More importantly, (A)–(D) indicate an asymmetric direction of fit between the representation-involving ways of talking about explanation and the ontic mode. In particular, it would appear that the adequacy of our communicative acts, our scientific texts, and our mental models depends in part on whether they correctly inform us about the features of the world that cause, produce, or are otherwise responsible for the phenomena we seek to explain. While Jon might be able to convey his endoplasmic black hole model of the action potential to his students (and thus to explain his model to them), Jon would not thereby explain the action potential to them. His putative explanation would merely leave them confused, whether they know it or not. If we treat this black hole model as one of Jon's mental representations activated when he thinks of action potentials, then it seems right to say that although Jon thinks he understands the action potential, he is deeply mistaken about this; in fact, he has only the illusion of understanding the action potential. And the same can be said of false models; they might vary considerably in the accuracy with which they describe the explanation for the action potential. If the philosophical topic of explanation is to provide criteria of adequacy for scientific explanations, then the ontic conception is indispensable: explanatory

communications, texts, and representations are evaluated in part by the extent to which they deliver more or less accurate information about the ontic explanation for the *explanandum* phenomenon.

2.3 Adequate Explanations and the Ontic Conception

In many areas of science, explanatory texts are taken to be adequate to the extent that they correctly describe the causes (etiological explanations) or the underlying mechanisms (constitutive explanations) responsible for the phenomenon one seeks to explain (the *explanandum* phenomenon) (See Machamer et al. 2000; Craver 2007; Bechtel and Abrahamsen 2005). In such areas of science, successful models contain variables that stand for causally relevant properties or features of the system and represent the appropriate relations among those variables. Successful communication of explanatory information (as opposed to misinformation) conveys information about those causally relevant features and their relations. And finally, one understands (rather than misunderstands) the *explanandum* phenomenon to the extent that one correctly grasps the causal structure of the system at hand.

The importance of truth to scientific explanation generally is recognized in the commonplace distinction between a *how-possibly model* and a *how-actually* model (Dray 1957; Machamer et al. 2000). Gastric ulcers might have been caused by emotional stress (as it was once thought), but they are in fact caused by *Helicobacter pylori* bacteria (see Thagard 1999). Action potentials might have been produced by a distinctive form of animal electricity, but they are in fact produced by fluxes of ions across the cell membrane. The earth might have been at the center of the solar system with the moon, sun, and planets revolving around it, but it is not. One might form elegant models describing these putative causes and constitutive mechanisms, and one might use such models to predict various features of the *explanandum* phenomenon, and such models might provide one with the illusion that one understands how an effect is brought about or how a mechanism works. However, there is a further fact concerning whether a plausible explanation is in fact the explanation.

To claim that truth is an essential criterion for the adequacy of our explanations, one need not deny that paradigmatically successful explanatory models explicitly make false assumptions or presume operating conditions that are never seen in reality. An explanatory model in physics might assume that a box is sliding on a frictionless plane. An explanatory model in electrophysiology might presume that an axon is a perfect cylinder or that the membrane obeys Ohm's law. A physiologist might model a system in a "wild-type" organism by presuming that all individual organisms in the wild type are identical. Such idealization is often required in order for one to form a parsimonious yet general description of a wide class of systems. Yet this undeniable fact about scientific models need not lead one to abandon the not so subtle difference between models that incorrectly describe how something might have worked from those that describe, more accurately, how it in fact works.

In other words, whatever we want to say about idealization in science, it should not lead us to the conclusion that there is no explanatory difference between a model that describes action potentials as being produced by ionic fluxes, on the one hand, and one that describes it as being produced by black holes, on the other. Perhaps then, the appropriate distinction is not between how-possibly and how-actually, but between how-possibly and how-actually within the limits of idealization.

Yet my point about the centrality of the ontic conception to our criteria of explanatory adequacy goes beyond the mere claim that our explanations should be true (or approximately true). Not all true models are explanatory. Models can be used to describe phenomena, to summarize data, to calculate undetected quantities, and to generate predictions (see Bogen 2005). Models can play any or all of these roles without explaining anything. Models can fall short as explanations because (1) they are purely descriptive or *phenomenal models*, (2) they are purely *predictive models*, (3) they are mere *sketches* of the components and activities of a mechanism with gaps and question marks that make the explanation incomplete, or (4) the model includes explanatorily irrelevant factors. Consider these in turn.

(1) *Phenomenal Models.* Scientists commonly draw a distinction between models that merely describe a phenomenon and models that explain it. Neuroscientists such as Dayan and Abbott, for example, distinguish between purely descriptive mathematical models, models that "summarize data compactly," and mechanistic models, models that "address the question of how nervous systems operate on the basis of known anatomy, physiology, and circuitry" (2001, p. xiii). Mechanistic models describe the relevant causes and mechanisms in the system under investigation. The distinction between purely descriptive, phenomenal models and mechanistic models is familiar in many sciences. Snell's law describes how light refracts as it passes from one medium to another, but the law does not explain why the path of light changes as it does. To explain this principle, one must appeal to facts about how light propagates or about the nature of electromagnetic phenomena. (Of course, one might explain the angle of refraction of a beam of light by appeal to the fact that the light crossed between two media in which it has different velocities. However, we are interested here in explaining why light generally bends when it passes from one medium to the next. Snell's law tells us that our beam of light is not alone in exhibiting this mysterious behavior, but it does not tell us why light generally behaves this way).

Precisely the same issue arose with respect to the HH model of the action potential. As part of building their "total current equation," Hodgkin and Huxley (1952) generated equations to model how the conductance of a neuronal membrane to sodium and potassium changes as a function of voltage during an action potential. The equations are surprisingly accurate (approximately true), but they leave it utterly mysterious just how the membrane changes its conductance during an action potential. Hodgkin and Huxley are explicit about this explanatory limitation in their model. To explain these conductance changes, scientists needed first to discover the membrane-spanning channels

that open and close as a function of voltage (Bogen 2005, 2008; Craver 2006, 2007, 2008; Hille 2001). The signature of a phenomenal model is that it describes the behavior of the target system without describing the ontic structures that give rise to that phenomenon.

(2) *Purely Predictive Models.* Explanatory models often allow one to make true predictions about the behavior of a system. Indeed, some scientists seem to require that explanatory models must make new predictions. Yet not all predictively adequate models are explanatory. A model might relate one effect of a *common cause* to another of its effects. For example, one might build a model that predicts the electrical activity of one neuron, A, on the basis of the activity of another neuron, B, when the activities of both A and B are in fact explained by the activity in a "parent" neuron, C, that synapses onto both A and B (while A and B have no influence on each other). A model might relate *effect to cause.* One might, that is, build a model that predicts the behavior of neuron C in the above example on the basis of the behavior of neurons A or B. One can infer that a brain region is active on the basis of changes in the ratio of oxygenated to deoxygenated hemoglobin in the vasculature of that brain region. This law-like correlation makes functional magnetic resonance imaging of the brain possible. Yet nobody to my knowledge believes that the changes in oxygenation explain neuronal activity in these brain regions. The explanation runs the other way around; changes in neural activation cause (and so explain) changes in regional blood flow. Finally, a model might relate two events that follow one another in a regular *sequence* but that, in fact, have no explanatory connection. One can predict that the ballgame will begin from the performance of the national anthem, but the performance of the national anthem does not explain the start of the game. The point of these examples is that models may lead one to expect a phenomenon without thereby explaining the phenomenon. These judgments of scientific common sense seem to turn on the hidden premise that explanations correctly identify features of the ontic structures that produce, underlie, or otherwise responsible for the *explanandum* phenomenon (see Salmon 1984). Expectation alone does not suffice for explanation.

(3) *Sketches.* A third dimension for the evaluation of scientific models is the amount of detail that they provide about the causal structure of the system in question. A model might "cover" the behavior of a system at many grains of description. It might be a phenomenal model, as described above, in which case it serves merely as a description, rather than an explanation, of the system's behavior. At the other end of the spectrum, it might supply a fully worked out description of all of the components, their precise properties, their precise spatial and temporal organization, all of the background and boundary conditions, and so on. It is rare indeed that science achieves that level of detail about a given system, in part because a central goal of science is to achieve generalization, and such particularized descriptions foil our efforts to build generalizable models. Between these poles lies a continuum of grains of detail. A mechanism sketch is a model of a mechanism that contains crucial black boxes or filler terms that, at the moment, cannot be filled in with further details. For example, one might

sketch a model of memory systems as involving encoding, storage, and retrieval without having any precise ideas about just how memories are encoded in the brain, how or where they are stored, or what precisely it would mean to retrieve them. Such a sketch might be true, or approximately true, and nonetheless explanatorily shallow. One can deepen the explanation by opening these black boxes and revealing their internal causal structure. In doing so, one allows oneself to answer a broader range of questions about how the phenomenon would differ were one or the other feature of the mechanism changed (cf. Woodward 2003). This ability is typically taken to be an indirect measure of one's understanding of how the system works. The crucial point about sketches for present purposes is that the spectrum from phenomenal model to sketch, to schema, to fully instantiated mechanism is defined by the extent to which the model reveals the precise details about the ontic explanation for the phenomenon. Again, it would appear, the ontic explanation plays an asymmetric and fundamental role in our criteria for assessing explanations.

(4) *Relevance.* An explanatory text for a given phenomenon ought to include all and only the factors that are explanatorily relevant to the *explanandum* phenomenon. While it is true that people with yellow fingers often get lung cancer, the yellow fingers are explanatorily irrelevant to the lung cancer. A putatively explanatory model that included finger color as part of the explanation for Carla's lung cancer would be a deeply flawed explanatory model.

As discussed in the previous section, explanations are sometimes spoken of as communicative acts, texts (e.g., models), and representations. So conceived, explanations are the kinds of things that can be more or less complete and more or less accurate. They might include more or less of the explanatorily relevant information. They might be more or less deep. Conceived ontically, however, the term explanation refers to an objective portion of the causal structure of the world, to the set of factors that produce, underlie, or are otherwise responsible for a phenomenon. Ontic explanations are not texts; they are full-bodied things. They are not true or false. They are not more or less abstract. They are not more or less complete. They consist in all and only the relevant features of the mechanism in question. There is no question of ontic explanations being "right" or "wrong," or "good" or "bad." They just are.

The point is that norms about the contents of ontic explanations make an essential contribution to the criteria for evaluating explanatory communications, texts (models), and mental models. Good mechanistic explanatory models are good in part because they correctly represent objective explanations. Mere how-possibly models describe the wrong causes or wrong mechanisms, whereas how-actually models get it right. Phenomenal models describe the phenomenon without revealing the ontic structures that produce it. Merely predictive models describe correlations but not causal structures. Mechanism sketches leave out relevant portions of the causal structure of the world. The issue here is not merely that an explanation must

be true: predictive models, phenomenal models, sketches, and models containing irrelevancies might be true but explanatorily inadequate. The ontic structure of the world thus makes an ineliminable contribution to our thinking about the goodness and badness of explanatory texts. The traditional philosophical problem of explanation was to provide a model that embodies the criteria of adequacy for sorting good explanations from bad. One cannot solve that problem without taking the ontic aspect of explanation seriously.

Let me put this another way: the norms of scientific explanation fall out of a prior commitment on the part of scientific investigators to describe the relevant ontic structures in the world. Explanation, in other words, is intimately related to the other aspects of science, such as discovery and testing. The methods that scientists use to discover how the world works, the standards to which they hold such tests, are intimately connected with the goal of science to reveal the ontic structures that explain why the phenomena of the world occur and why they occur as they do. One cannot carve off the practice of building explanations from these other endeavors. These methods and products of the scientific enterprise hang together once one recognizes that science is committed, ab initio, to giving a more or less precise characterization of the ontic structure of the world.

The commitment to realism embodied in these claims can be justified on several grounds. It is justified in part because it makes sense of scientific-commonsense judgments about the norms of explanation. It is also justified by reference to the fact that an explanation that contains more relevant detail about the responsible ontic structures are more likely, all things equal, to be able to answer more questions about how the system will behave in a variety of circumstances than is a model that does not aim at getting the ontic structures that underlie the phenomenon right. This follows from the fact that such models allow one to predict how the system will behave, for example, if its parts are broken, changed, or rearranged and so how the mechanism is likely to behave if it is put in conditions that make a difference to the parts, their properties, or their organization. It is always possible (though never easy) to contrive a phenomenally adequate model post-hoc if and when the complete input-output behavior of a system is known. However, the critical question is how readily we can discover this input-output mapping across the full range of input conditions without knowing anything about the underlying mechanism. We are far more likely to build predictively adequate models when aspects of the mechanism are known. Finally, models that reveal objective causal structures automatically reveal knobs and levers in the world that might be used for the purposes of bringing parts of it under our control (Woodward 2003).

To illustrate the importance of the ontic aspect of explanation for developing a philosophical theory of scientific explanation, I now consider two models of explanation that, at least on some readings, neglect the importance of the ontic mode. I argue that they fail to embody the criteria of adequacy for scientific explanations because they focus their attention on representations rather than on the ontic structures those representations represent.

2.4 The CL Model

One systematic (though somewhat uncharitable) way of diagnosing the widely acknowledge failure of the CL model is to see it as emphasizing explanatory representations over the ontic structures they represent. This is not the only, nor even the most familiar, diagnosis. Others (such as Churchland and Bechtel) argue that the CL model fails because it insists on formulating explanations in propositional logic. Such critics respond to the shortcomings of the CL model by developing new representational frameworks that are more flexible and more cognitively realistic. If my diagnosis is correct, such revisions fail to address the core problems with the CL model as a theory of scientific explanation.

According to the CL model, explanations are arguments. The conclusion of the argument is a description of the *explanandum* phenomenon. The premises are law statements, canonically represented as universal or statistical generalizations, and descriptions of antecedent or boundary conditions. Explanation, on this view, is expectation: the explanatory argument shows that the description of the *explanandum* phenomenon follows, via an acceptable form of inference, from descriptions of the laws and conditions. In this sense, explanations show that the *explanandum* phenomenon was to be expected given the laws and the conditions. The emphasis is on the representational structures: *statements* of the laws, *descriptions* of the conditions, *entailment* relations, and human *expectations*.

I say that this characterization is somewhat uncharitable for two reasons. First, the CL model typically requires that the premises of the explanatory argument be true, that is, that the law statements describe real laws and that the descriptions of conditions are accurate. Second, and more fundamentally, the logical force with which the *explanandum* statement follows from the premises might be taken to mirror the sense in which the *explanandum* phenomenon had to happen or was more likely to happen given the laws and the initial conditions. One might more charitably interpret Hempel as suggesting that the inferential necessity in the argument mirrors or expresses the corresponding natural necessity in the world. And as Salmon (1989) pointed out, there are passages in Hempel's classic statement of the CL model that lend themselves to such an ontic interpretation: the laws, not law statements, explain. Be that as it may, Hempel does not appear to have recognized this ambiguity in his own writing, and it is certainly in the keeping with the program of logical empiricism to think that all the essential features of science could be captured with the expressive formalism of logic. The commitment to "natural necessity" in this putatively more charitable reading, in fact, does violence to Hempel's strongest empiricist convictions.[3]

[3]As Ken Aizawa (personal communication) notes, the CL model arguably can accommodate sentences (A)–(D) of Sect. 2.2. If one takes the CL model to equate explanation with rational expectability rather than rational expectation, then one can say that there are explanations to be discovered and explanations so complex that we will never know them.

Let me amplify a bit. On the most austerely empiricist interpretation of the CL model, it would be incorrect to say that the logical or inferential necessity of the argument "mirrors" a kind of natural necessity with which events follow the laws. According to this interpretation, the universal generalizations used to express universal laws are true summaries of events; they assert that all Xs that are F are, as a matter of fact, also G. There is no further thing, the necessity of a law, that makes it the case that all Xs that are F are also G. Likewise, one might understand probabilistic laws as asserting objective frequencies. If we count up all of the Xs that are F, we find as a matter of fact that some percentage of them are G. There need be no further fact that explains why G holds with this frequency in the population. In response to various counterexamples to the CL model, its defenders began to place more restrictions on the representations of laws. When it was objected that one could, according to this model, explain why a particular coin in Goodman's pocket is a dime on the basis of the claim that all the coins in Goodman's pocket are dimes, the response was to demand that laws make no reference to particulars, such as Goodman, or particular places, such as his pocket, or particular times, such as t = March 17, 1954 (see Ayer 1974). The formal structure of the representation, in other words, was called upon to block the counterexamples.

But it appears that no amount of formal modification could block some very serious problems. In particular, the account could not satisfy the criteria of adequacy sketched in the previous section. First, the model does not, by itself, have machinery to distinguish phenomenal descriptions from explanations. Asked why a given X that is F is G (e.g., why a particular raven is black), the CL model famously appeals to the generalization that all Fs are Gs (e.g., all ravens are black). But one might reasonably object that such an explanation fails to discharge the request for explanation and, instead, merely lists the *explanandum* phenomenon as one of many phenomena, each of which is equally mysterious. Likewise, to explain why an action potential has a particular form, it does little to provide a generalized description of that form. One wants to know why action potentials have that form, not that all action potentials, in fact, have it. Clearly, what one wants is an account of the ontic structures, in this case mechanisms, that give rise to action potentials (see Bogen 2005; Craver 2006). Perhaps one could describe such mechanisms in terms of a series of law statements about the internal causal structure of the action potential. Indeed, one might see the HH model as offering a sketch of just such an account (I am not in favor of this way of talking, but will entertain it here to make my limited point). My limited point is that there is a difference between such a mechanistic model, which reveals internal causal structures, and a phenomenal model, which simply generalizes the phenomenon; and crucially, the difference between them is not a formal difference but a difference in what is being described. The mechanistic description describes parts and processes at a lower level than the action potential, and this shift in levels is not a formal difference in the representation but an ontic difference between a whole (the action potential) and its parts. To mark the difference between a phenomenal model and a mechanistic model (1 above), that is, one must appeal to the (quasi-mereological) structures of the world that relate the *explanans* to the *explanandum*.

Second, if one sticks with the austere empiricist reading of the CL model (i.e., one that is not supplemented with some sort of ontic difference between laws and accidental generalizations), then the CL model does not recognize a distinction between generalizations that are explanatory and generalizations that are not. It should not matter whether two causally independent effects of a common cause explain one another or whether an effect explains its cause, or whether one type of event is explained by another type of event that always (or regularly) precedes it in time. That is, the CL model in its austere empiricist form does not distinguish explanatory models from merely predictive models (2 above). Indeed, the very idea that explanation is expectation would appear to insist that any predictive model is ipso facto an explanatory model. This is why the "prediction-explanation symmetry thesis" was heavily debated by proponents and opponents of the CL model alike (Hempel abandoned it quickly). The difference between explanation and prediction, which is fundamental to providing an adequate account of explanation, seems to rely not on some feature of the way we represent the world but rather on some feature of the world that distinguishes explanatory relations from mere correlations.

Third, the CL model in its austere form does not appear to mark a distinction between sketches and more complete descriptions of a mechanism. So long as the model suffices to derive a description of the *explanandum* phenomenon, it counts as an explanation (full stop). Grant that a defender could reconstruct multilevel explanation as one finds in neuroscience and physiology by describing, as it were, laws within laws all the way down. That is not the issue. The issue before us is whether the CL model recognizes that by exploding black boxes and revealing internal causal structures one is, ipso facto, providing a deeper explanation. The model would become more and more complex, of course, as it includes more and more of the internal causal structure, but nothing in the formal structure of the model would indicate that the model was getting deeper. For that, one must appeal to features of the world that the model describes.

Finally, the CL model in its austere form does not recognize a difference between relevant and irrelevant explanatory factors. It is generally true that men who take birth control pills fail to get pregnant, and nothing in the formal structure of the CL model instructs us to jettison the irrelevant conjunct in the antecedent of this conditional. A similar problem arises for explanations of general laws. The predictive value of a model is unaffected (at least in many cases) by the inclusion of irrelevant detail. If one set of law statements and boundary conditions, K, entails another set, P, then the conjunction (K and S), where S is any arbitrary sentence that fails to contradict a member of K, also entails P. So the HH model plus Kepler's laws explains the HH model. Hempel called this the problem of irrelevant conjunction (Hempel 1965, p. 273, fn 33). This is a problem because it conflicts with the common scientific practice of filtering out irrelevant factors from explanations. A mechanistic explanatory model suffers, for example, if it includes irrelevant parts that are not in the mechanism, irrelevant properties that play no causal role, or irrelevant activities that are sterile in the mechanism. The important point for present purposes, however, is that it would appear that explanatory relevance is not a feature

of the formal structure of an argument but rather of the kinds of ontic structures that the representation describes.

These kinds of objection to the CL model are by now thoroughly familiar to philosophers of science. What is less familiar, I suppose, is the thought that these problems require for their solution that one shift one's focus away from the representational structures of explanatory texts to features of the systems they represent. This is the insight of the ontic conception of explanation. The solution to these puzzles, and so the fundamental tasks of providing a philosophical account of explanation, is not to be discovered by building elaborate theories about how explanatory information is represented. Though the question of how such information is or ought to be represented is interesting and worthwhile, it will not by itself answer the questions that a narrower, normative approach takes as distinctive of the philosophical problem of scientific explanation.

2.5 Churchland's Connectionist Account

If this diagnosis is correct, then one should find similar problems at work for those theories of scientific explanation that keep the representational subsumption view in place but change the format of the representation. As a representative of psychologistic models of explanation more generally, consider Paul Churchland's (1989) parallel distributed processing (PDP) account of explanation. Churchland objects to Hempel's model (and, in fact, the entire logical empiricist enterprise) on the ground that human cognitive agents (such as scientists) do not in fact think with the structures of first-order predicate logic. His revolutionary objective is to rebuild a model of science inspired by connectionist, or parallel distributed processing, theories of cognition rather than on twentieth-century advances in logic.

On Churchland's view, understanding is prototype activation in a connectionist network:

> Explanatory understanding consists in the activation of a particular prototype vector in a well-trained network. It consists in the apprehension of the problematic case as an instance of a general type, a type for which the creature has a detailed and well-informed representation. (Churchland 1989, p. 210)

When we understand a phenomenon, we assimilate it to a prototype and thereby generate novel features of the phenomenon from a few input features. The prototype stores a wealth of theoretical information about a phenomenon. Understanding, accordingly, is a matter of recognizing that a given phenomenon fits a more general prototype. Scientific explanation involves the construction of prototypes (such as the HH model, presumably) that can be so applied.

The first thing to notice about Churchland's model of understanding is that he does not say how those instances of prototype activation that constitute understanding are different from those that do not. Prototype-activation vectors are used to describe many aspects of brain function. Stored patterns of activation across

populations of neurons control balance, posture, and reaching; they produce and direct saccadic eye movements; and they regulate endocrine release and bodily fluid homeostasis. To put the point maximally bluntly: if the brain does it, it likely does it with activation vectors. So the idea that understanding involves the activation of prototype vectors tells us very little about the distinctive character of understanding.

To make this more concrete, consider the distinction between recognition and understanding. One can recognize Ike in a crowd without explaining anything about him. Suppose that one wants to understand why Ike is a bookie, or why Ike has only a junior high education. One cannot answer these questions by merely recognizing Ike. This is because Ike's surface features (his gait, his hair line, his shape), that is, the kinds of things that will show up in the visual Ike-recognition vector, are in most cases not explanatorily relevant to his professional and educational status. To drive the point home, it would appear that Churchland's model does not have a principled means for distinguishing phenomenal models, which merely describe the phenomenon to be explained, from explanatory models, which explain why the *explanandum* phenomenon is as it is.

In the years since Churchland's suggestion, cognitive scientists have learned more about the cognitive mechanisms of causal understanding. Churchland could add further content to his account by building details about how human cognitive systems discern and represent the relevant ontic structures that constitute *bona fide* understanding. Though it is no trivial matter to formulate such a theory, there can be no doubt that such a theory could, in fact, be implemented in a connectionist network, whatever it is. However, notice that building a model of the cognitive capacities that make bona fide understanding possible in creatures such as us requires one to say what the prototype vectors must be about in order to constitute bona fide understanding: that they are about causal structures, laws, statistical dependencies, mechanisms, or what have you. In other words, in order to say which specific cognitive capacities are relevant to our ability to understand the world, we must thrust our attention outward from representations to the ontic explanations that they must represent if they are to truly constitute understanding.

There is further reason to avoid equating scientific explanation with the abilities of individual cognitive agents. Some phenomena might be so complex that they overwhelm our limited (individual) cognitive systems. Perhaps a mechanism has so many parts with so many interactions that it is impossible for a single person to fully understand. Perhaps scientists must rely on computer simulations, graphical representations, and large compiled databases in order to build models that explain the complex phenomena in their domain. Perhaps human working memory is so limited that it cannot entertain all of the information explanatorily relevant to a given phenomenon (compare Rosenberg 1985, 1994). Mary Hegarty shows that even simple mechanisms overwhelm our processing capacities if they have over a handful of parts or if the interactions among them cannot be represented in two dimensions (Hegarty et al. 1988). For this reason, it seems inappropriate to model scientific explanation, which has no principled limit on its complexity, on the basis of individual human cognition, which is often quite limited. It would be wrong to

say that phenomena produced by very complex mechanisms (i.e., those that outstrip our cognitive capacities) have no explanation. The explanations exist even if our brains cannot represent them.

Suppose, though, we accept Churchland's PDP model as an adequate account of the psychology of human understanding. Can this psychological account do double duty as an account of the norms of scientific explanation? The inclusiveness of the PDP model (and the representational model in general) is again its primary drawback. The more permissive an account of explanatory representations, the less likely it is to fulfill the distinctions discussed above in Sect. 2.3. Churchland explicitly disavows interest in the norms of explanation (Churchland 1989, p. 198). However, the demands on a philosophical theory of explanation cannot be satisfied without thinking about norms for evaluating explanations. Consider Churchland's description of etiological causal prototypes:

> An etiological prototype depicts a typical temporal sequence of events, such as cooking of food upon exposure to heat, the deformation of a fragile object during impact with a tougher one, the escape of liquid from a tilted container, and so on. These sequences contain prototypical elements in a prototypical order, and they make possible our explanatory understanding of the temporally extended world. (Churchland 1989, p. 213)

But as discussed above, some temporal sequences are explanatory (if appropriately supplemented with the causal relations between the different events in the sequence), and some are not. An account of explanation should help one to distinguish the two. Churchland acknowledges this limitation: "Now just what intricacies constitute a genuine etiological prototype, and how the brain distinguishes between real causal processes and mere pseudoprocesses, are secondary matters I shall leave for a future occasion" (Churchland 1989, p. 214). Those who would develop a normative account of explanation, however, cannot avoid this question. The way to understand how brains distinguish causes from temporal sequences is to start by considering how causes differ from temporal sequences – that is, by examining the objective explanations in the world rather than the way that they are represented in the mind/brain. A similar point could be made about common cause structures and effect-to-cause explanations. That is, the model does not appear to have the resources to distinguish predictive models from explanatory models.

An equally fundamental problem arises when we consider the question of explanatory relevance. Grant that explanatory representations are prototypes and that explanation involves activating such prototypes. Different features of the phenomenon are relevant for different explanatory purposes. Suppose that Ike is a member of the gang, the Sharks; he is single and 30 years old; he weighs 210 lb; he has a junior high education; he is a bookie; he idolizes Johnny Ramone; and he plays guitar. To explain why he is a bookie, it would be relevant to note that he is a member of a gang and perhaps that he has a junior high education, but it would probably not be relevant to note that he weighs 210 lb or that he plays guitar. To explain why he plays guitar, it might be relevant to note that he is a single, 30-year-old male who idolizes Johnny Ramone, but not (I suppose) that he is a bookie or that he has a junior high school education. All of these features are in the Ike prototype (which,

if we know him well, contains innumerable other features of varying degrees of explanatory relevance to these phenomena). And all of these features are activated when we think of Ike. Yet only some of these features are relevant to explaining why he is a bookie, only some are relevant to explaining why he plays guitar, and few of the features in these two lists overlap.

What goes for Ike goes for the categories of science. Ion channels can be characterized along a number of dimensions: molecular weight, primary structure, voltage sensitivity, maximum conductance values, primary structure, and so on. Different features of a given type of ion channel are relevant for different explanatory purposes. An account of explanation that can be used to sort good explanations from bad should help to sort explanatorily relevant information from explanatorily irrelevant information. But the PDP account cannot be so used unless the activation-vector story is supplemented with an account of explanatory relevance. However, to supplement it, one will have to begin by assessing what explanatory relevance is, and this again thrusts our attention away from representation and out onto the ontic structures that good explanatory texts describe.

Hempel, who can be credited with initiating sustained philosophical discussion of the nature of scientific explanation, drew precisely the sharp line between explanation and understanding that I am here trying to make explicit:

> . . . man has long and persistently been concerned to achieve some understanding of the enormously diverse, often perplexing, and sometimes threatening occurrences in the world around him . . . Some of these explanatory ideas are based on anthropomorphic conceptions of the forces of nature, others invoke hidden powers or agents, still others refer to God's inscrutable plans or to fate.
>
> Accounts of this kind undeniably may give the questioner a sense of having attained some understanding; they may resolve his perplexity and in this sense 'answer' his question. But however satisfactory these answers may be psychologically, they are not adequate for the purposes of science, which, after all, is concerned to develop a conception of the world that has a clear, logical bearing on our experience and is capable of objective test. (Hempel 1966, pp. 47–48)

The point of this passage is to drive a wedge between the psychological mechanisms that give rise to the sense of intelligibility and understanding, on the one hand, and a properly philosophical theory of scientific explanation. The task is to develop an account of scientific explanation that makes sense of the scientific project of connecting our models to structures that can be discovered through experience and objective tests. In domains of science that concern themselves with the search for causes and mechanisms, this amounts to the idea that the norms of explanation fall out of a commitment by scientists to describe as accurately and completely as possible the relevant ontic structures in the world. Viewed in this way, our theories of scientific explanation cannot carve off those ontic structures as if they were expendable in the search for a theory of explanation: the norms of explanation fall out of the scientific commitment to describe those ontic structures.

2.6 Idealization and the Ontic Conception

Let us now turn attention to the role of idealization in scientific explanation. As a matter of historical record, explanatory texts are often idealized in the sense that they make false assumptions about the system they represent in order to make the texts more compact and elegant. To make matters worse, such texts appear to function as they do in our scientific communication largely because they describe the relevant ontic structures incorrectly. If so, one might be tempted to conclude that it is inappropriate to emphasize the ontic mode of explanation; scientific explanation essentially involves divorcing one's thought from the relevant ontic structures and providing representations that make the messy phenomena intelligible and useful to creatures like us.

The undeniable fact that scientific models are typically idealized is clearly most problematic for accounts of explanation that demand that a scientific explanation must subsume a description of the phenomenon under a true general representation, that is, for the strongest versions of the representational subsumption view. Hempel, for example, requires as a criterion of adequacy on explanatory arguments that the premises of the argument be true. And for this reason, his model of explanation rather famously has difficulty accommodating the ubiquitous practice of idealization. Hempel was committed to a representational view and to the idea that the representations in explanations have to be true, so it was a challenge for his view that explanatory models are (almost) always idealized.

Of course, the requirement that explanatory texts must be true is certainly reasonable. Even if one can subsume a description of the action potential under a model that posits the existence of black holes in the endoplasmic reticulum, and even if the model renders action potentials intelligible (i.e., the model gives people the sense of understanding how action potentials are produced), such a model simply cannot explain the action potential. The reason is plain: there are no black holes in the endoplasmic reticulum. The ideal of scientific explanation cannot be wholly severed from the criterion of truth lest we lose any grip at all on the idea that it is a scientific explanation rather than an intelligible tale of some other sort. The goal of building an explanatory text is not to provide the illusion of understanding, but rather to provide *bona fide* understanding. Entirely false explanatory texts offer only the former.[4]

Idealized models, however, are of interest because they are not entirely false: they bring to light aspects of the system under investigation that are difficult to see unless one makes false assumptions. Things are easier if one assumes, for example, that the axon is cylindrical, that the concentration of ions is everywhere uniform, and that the membrane obeys Ohm's law strictly. The explanatory text contains idealizing assumptions precisely because, in making such assumptions, one reveals aspects of

[4]The same point could be made in terms of empirical adequacy rather than truth, should that be preferred. Idealized theories, as I have described them, must be empirically inadequate in some respect; otherwise, there would be no basis for the claim that they contain false assumptions.

the ontic structure of the system that would otherwise be occluded. The idealizing model thus has the capacity to inform us about the ontic explanations for phenomena even if the model is not, strictly speaking, true.

Now, it would surely be a mistake to claim that a model has to be true to convey explanatory information. But conveying explanatory information about X and truly representing the explanation for X are not the same thing. Friends of the ontic conception should say that idealized models are useful for conveying true information about the explanation, but that they are not true representations of the explanation.

One benefit of clearly disambiguating the ontic mode from the communicative, representational, and cognitive modes of talking about explanation is that it allows us to divide labor on these matters. Terms like "true," "idealized," and "abstract" apply to representations or models. They do not apply to the ontic structures they represent (bracketing cases in which the ontic structures involved in the explanation are themselves representations). Once these are separated, the problem of idealization is clearly not a problem for philosophical theories of explanation; rather it is a problem for philosophical theories of reference. The question at the heart of the problem of idealization is this: What is required for a given representation to convey information about the ontic structure of the world? This is an important question, but it is a question about reference, not a question about explanation. We only invite confusion if we fail to keep these questions distinct.

To say that a model is idealized is, ipso facto, to recognize a distinction between models that are true and models that are false. To say that a model is an idealization of an ontic explanation, after all, is to say that the model contains one or more false commitments about that ontic explanation. The very idea of an idealized model of an explanation commits one, at least implicitly, to the existence of an ontic explanation against which the model can be evaluated. It is more sensible to say that idealized models convey explanatory information in virtue of making false assumptions that bring certain truths about the ontic explanation to light. If we say, in contrast, that false models explain, we are left scratching our heads about how a false model could be an explanation of anything at all. Our heads will itch, however, only if we are committed first and foremost to the idea that explanations are representations. But that is to get things backward. The explanations are in the world. The scientist's task is to describe them. And they can use any number of representational tools to convey that explanatory information clearly and effectively. If we give up on the representational subsumption view as the heart of our philosophical theories of explanation, the problem of idealization then finds its proper home in semantics.

2.7 Conclusion

The central tasks for a philosophical theory of scientific explanation are (a) to demarcate explanation from other kinds of scientific achievement and (b) to articulate the norms that distinguish adequate explanations from inadequate explanations.

In this chapter, I have argued that the term "explanation" is ambiguous, having at least four senses, and that one might construct a theory adequate to one of these senses without in the process constructing a theory that is adequate to the others. I have argued that the philosophical theory of explanation depends fundamentally on an ontic conception of explanation, that is, on a view about the kinds of structures in the world that count as legitimately explanatory. Appeal to such structures is required to distinguish how-possibly from how-actually explanations, phenomenal models from mechanistic models, merely predictive models from explanatory models, sketches from complete-enough explanations, and relevant from irrelevant explanatory factors.

Just as representational views of explanation, on their own, cannot provide an account of the norms underlying a philosophical analysis of scientific explanation, an account that addresses those norms leaves work to be done by representational theories. Not all of the facts in an ontic explanation are salient in a given explanatory context, and for the purposes of communication, it is often necessary to abstract, idealize, and fudge to represent and communicate which ontic structures cause, constitute, or otherwise are responsible for such phenomena. Such topics are the proper province of psychologistic theorizing about scientific explanation and work in the philosophy of reference. But these topics are separate from the classic philosophical topic of the nature of scientific explanation.

Acknowledgments I thank Andreas Hütteman, Marie I. Kaiser, Alex Reutlinger, and other members of the philosophy community at the Universtität zu Köln for support and discussion during the writing of this chapter. I also thank Ken Aizawa, Kevin Amidan, Justin Garson, and Jim Tabery for feedback on earlier drafts. This chapter was delivered at Duke University, and I am grateful to Robert Brandon, Andrew Janiak, Karen Neander, Alex Rosenberg, and Walter Sinnott-Armstrong for helpful comments.

References

Ayer, A. J. (1974). What is a law of nature. In M. Curd & J. A. Cover (Eds.), *Philosophy of science: The central issues* (pp. 808–825). New York: WW Norton.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421–441.

Bogen, J. (2005). Regularities and causality: Generalizations and causal explanations. *Studies in the History and Philosophy of Biology and the Biomedical Sciences, 36*(2), 397–420.

Bogen, J. (2008). Causally productive activities. *Studies in History and Philosophy of Science: Part A, 39*(1), 112–123.

Churchland, P. M. (1989). *A neurocomputational perspective: The nature of mind and the structure of science*. Cambridge: MIT Press.

Coffa, J. A. (1974). Hempel's ambiguity. *Synthese, 28*(2), 141–163.

Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*(3), 355–376.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon.

Craver, C. F. (2008). Physical law and mechanistic explanation in the Hodgkin and Huxley model of the action potential. *Philosophy of Science, 75*(5), 1022–1033.

Darden, L. (2006). *Reasoning in biological discoveries*. New York: Cambridge University Press.
Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge: MIT Press.
Dray, W. (1957). *Laws and explanations in history*. Oxford: Oxford University Press.
Hegarty, M., Just, M. A., & Morrison, I. R. (1988). Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning. *Cognitive Psychology, 20*(2), 191–236.
Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation* (pp. 331–496). New York: Free Press.
Hempel, C. G. (1966). *Philosophy of natural science*. Englewood Cliffs: Prentice Hall.
Hille, B. (2001). *Ion channels of excitable membranes*. Sunderland: Sinauer.
Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology, 117*(4), 500–544.
Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, 48(4), 507–531.
Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation*. Minneapolis: University of Minnesota Press.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Rosenberg, A. (1985). *The structure of biological science*. Cambridge: Cambridge University Press.
Rosenberg, A. (1994). *Instrumental biology or the unity of science*. Chicago: University of Chicago Press.
Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Salmon, W. C. (1989). Four decades of scientific explanation. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation, Minnesota studies in the philosophy of science* (Vol. 18, pp. 3–219). Minneapolis: University of Minnesota Press.
Thagard, P. (1999). *How scientists explain disease*. Princeton: Princeton University Press.
Weber, M. (2005). *Philosophy of environmental biology*. Cambridge: Cambridge University Press.
Woodward, J. (2003). *Making things happen*. New York: Oxford University Press.

Part II
Explanation in the Biological Sciences

Chapter 3
Causal Graphs and Biological Mechanisms

Alexander Gebharter and Marie I. Kaiser

Abstract Modeling mechanisms is central to the biological sciences – for purposes of explanation, prediction, extrapolation, and manipulation. A closer look at the philosophical literature reveals that mechanisms are predominantly modeled in a purely qualitative way. That is, mechanistic models are conceived of as representing how certain entities and activities are spatially and temporally organized so that they bring about the behavior of the mechanism in question. Although this adequately characterizes how mechanisms are represented in biology textbooks, contemporary biological research practice shows the need for *quantitative*, *probabilistic* models of mechanisms, too. In this chapter, we argue that the formal framework of causal graph theory is well suited to provide us with models of biological mechanisms that incorporate quantitative and probabilistic information. On the basis of an example from contemporary biological practice, namely, feedback regulation of fatty acid biosynthesis in *Brassica napus*, we show that causal graph theoretical models can account for feedback as well as for the multilevel character of mechanisms. However, we do not claim that causal graph theoretical representations of mechanisms are advantageous in all respects and should replace common qualitative models. Rather, we endorse the more *balanced view* that causal graph theoretical models of mechanisms are useful for some purposes while being insufficient for others.

The order of authorship is alphabetical; both authors contributed equally to this chapter.

A. Gebharter (✉)
Heinrich-Heine-Universität Düsseldorf, Düsseldorf Center for Logic and Philosophy of Science, Universitätsstraße 1, 40225 Düsseldorf, Germany
e-mail: alexander.gebharter@phil.hhu.de

M.I. Kaiser
Philosophisches Seminar, Universität zu Köln, Richard-Strauss-Str. 2, 50931 Köln, Germany
e-mail: kaiser.m@uni-koeln.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_3,

Keywords Causal graph theory • Modeling • Mechanism • Probabilistic model • Quantitative model

3.1 Introduction

The search for mechanisms that underlie the phenomena under study is ubiquitous in many biological fields. Physiologists seek to find the mechanism for muscle contraction, cancer scientists try to discover the mechanisms that cause cell proliferation, and ecologists aim at elucidating the various mechanisms that bring about the maintenance of species diversity – just to mention a few examples. In the last 15 years, the philosophical literature on mechanisms has dramatically increased. Among the major proponents of the "new mechanistic philosophy" (Skipper and Millstein 2005, p. 327) are Carl Craver (2007), William Bechtel (2006, 2008), Stuart Glennan (2002, 2005), Lindley Darden (2006, 2008), and Peter Machamer et al. (2000). According to the mechanist's view, scientific practice consists in the discovery, representation, and manipulation of mechanisms. Scientific explanations are (exclusively or primarily) conceived as mechanistic explanations, that is, as descriptions of how the components of a mechanism work together to produce the phenomenon to be explained.[1]

Our primary interest in this chapter is the modeling of biological mechanisms. How are, can, and should mechanisms be represented? Are certain kinds of models of mechanisms advantageous with regard to particular scientific purposes like explanation, understanding, prediction, or manipulation? Previous philosophical literature on this topic (e.g., Glennan 2005; Craver 2007; Bechtel 2008) regards mechanistic models as being primarily *qualitative* representations. According to the mechanist's view, adequate models of mechanisms describe all and only those factors that contribute to bringing about the mechanism's behavior of interest (i.e., the "constitutively relevant" factors; cf. Craver 2007, pp. 139–159). These factors include the entities (or objects) that compose the mechanism, the activities (or operations or interactions) that these entities engage in, and the spatial and temporal organization of the entities and activities (i.e., how the entities are spatially distributed, which position shifts of entities take place, which activities initiate which other activities to what time). These qualitative models of biological mechanisms are typically depicted by diagrams (cf. Perini 2005), which scientists sometimes call "cartoon models" (Ganesan et al. 2009, p. 1621). Diagrams make it easier to understand how the steps of a mechanism together bring about the behavior in question. Hence, the representations of mechanisms that can be found in common biology textbooks are typically qualitative models.

[1]Of course, one need not subscribe to all the details of the mechanistic view of science in order to acknowledge the importance of mechanisms to wide areas of biology.

However, biological research practice is much more diverse than what is depicted in biology textbooks. Whereas the models of mechanisms which are designed for textbooks aim at providing explanations and promote understanding, modeling strategies that are pursued in contemporary scientific practice, by contrast, serve multiple purposes. Besides offering explanation, models of mechanisms are also used, for instance, to make (quantitative or qualitative) predictions, to guide hypotheses building in scientific discovery, and to design manipulation experiments or even computer simulations. In some research contexts what will be needed are not purely qualitative models of mechanisms, but rather models that contain quantitative, probabilistic information. These models often have the virtue of being closer to the experiments and studies that are actually carried out in biological research practice. It is due to this closeness that probabilistic and quantitative models often allow for more usable predictions, in particular when it comes to predicting the probabilities of certain phenomena of interest under specific manipulations. Another advantage of models of mechanisms that combine qualitative with quantitative, probabilistic information might be that they allow for the integration of qualitative (e.g., molecular) studies and probabilistic (e.g., ecological or evolutionary) studies in a certain biological field. This is, for example, an urgent issue in epigenetics where the laboratory experiments performed by molecular epigeneticists and the observational studies and computer simulations conducted by ecologists and evolutionary biologists need to be brought together (cf. Baedke 2012).

With this chapter, we respond to the need of contemporary biology for models of mechanisms that include quantitative, probabilistic information. We argue that the formal framework of causal graph theory is well suited to provide us with probabilistic, (often) quantitative representations of biological mechanisms.[2] We illustrate this claim with an example from actual biological research, namely, feedback regulation of fatty acid biosynthesis in *Brassica napus*. Modeling this example allows us to show how causal graph theory is able to account for certain features of biological mechanisms that have been regarded as problematic (e.g., their multilevel character and the feedback relations that they frequently contain). However, besides the virtues our analysis of this case study also reveals which difficulties causal graph theoretical modeling strategies face when it comes to representing mechanisms. As a result, we argue for the *balanced view* that, even though causal graph theoretical models of mechanisms have advantages with respect to particular scientific purposes, they also have shortcomings with respect to other purposes.

We start with an introduction of the basic formal concepts of causal graph theory (Sect. 3.2). In Sect. 3.3, we present what can be regarded as the major characteristics of biological mechanisms, namely, their multilevel character, their two kinds of components, and the spatial and temporal organization of their components. Section 3.4 deals with the case study that is central to our analysis:

[2]In certain areas of neuroscience, causal graph modeling is already prevalent (cf. the work of Karl J. Friston, Michael D. Lee, Eric-Jan Wagenmakers, Josh Tenenbaum, and others).

the mechanism for feedback inhibition of ACCase by 18:1-ACP in *Brassica napus*. In Sect. 3.5, we discuss how this mechanism (as well as one of its submechanisms) can be modeled by using causal graph theory. In doing so, we also address the possible objection that causal graph theory can account neither for the feedback relations that many biological mechanisms contain nor for the fact that mechanisms are frequently organized in nested hierarchies. On the basis of this analysis, we can then specify, on the one hand, the virtues and, on the other hand, the shortcomings of modeling biological mechanisms within a causal graph framework (Sect. 3.6).

3.2 Causal Graph Theory

Causal graph theory is intended to model causality in a quite abstract and empirically meaningful way; it therefore provides principles which connect causal structures to empirical data. While causal structures are represented by graphs, empirical data is stored by means of probability distributions over sets of statistical variables. In this section we will introduce the basic formal concepts needed to investigate the question of whether a causal graph framework is capable of representing mechanisms. We start by giving some notational conventions and remarks concerning statistical variables and probability distributions (Sect. 3.2.1) before providing definitions for "probabilistic dependence" and "probabilistic independence" (Sect. 3.2.2). We introduce the concept of a causal graph (Sect. 3.2.3) and illustrate how such a causal graph, complemented by a probability distribution, becomes a causal model (Sect. 3.2.4).

3.2.1 Statistical Variables and Probability Distributions

A *statistical variable* X is a function that assigns exactly one of at least two mutually exclusive properties/possible values of X ("$val(X)$" designates the set of X's possible values) to every individual in X's domain D_X. Statistical variables can be used in a way quite similar to predicate constants. "$X(a) = x$" (where "a" is an individual constant), for instance, can be read as the token-level statement "individual a (e.g., a particular *Drosophila* fly) has property x (e.g., red eye color)" and "$X(u) = x$" (where "u" is an individual variable) as the type-level statement "having property x." Formulae like "$X(u) = x$" can be abbreviated as "$X = x$" or, even shorter, as "x" whenever reference to individuals u is not needed. For the sake of simplicity, we shall only use discrete variables, that is, variables X whose set of possible values $val(X)$ is finite. Continuous quantities can be captured by discrete variables whose values correspond to the accuracy of the used measurement methods.

Given a statistical variable X or a set of statistical variables X, then Pr is a *probability distribution* over X if and only if Pr is a function assigning a value $r_i \in [0,1]$ to every $x \in val(X)$, so that the sum of all assigned r_i equals 1.

Since probability distributions should be capable of storing empirical data, we interpret probabilities as objective probabilities, that is, as inductively inferred limit tendencies of observed frequencies.

3.2.2 Probabilistic Dependence and Independence Relations

Given a probability distribution Pr over variable set V, conditional probabilistic dependence between two variables X and Y can be defined in the following way:

(1) $DEP_{Pr}(X,Y|M)$ if and only if there are x, y, and m so that $Pr(x|y,m) \neq Pr(x|m)$, provided $Pr(y,m) > 0$.[3]

Read "$DEP_{Pr}(X,Y|M)$" as "X and Y are probabilistically dependent conditional on M." According to definition (1), two variables X and Y are probabilistically dependent conditional on M if the probability of at least one value of one of these two variables is probabilistically sensitive to at least one value of the other variable in at least one context $M = m$. So "probabilistic dependence" is a quite weak notion. "Probabilistic independence," on the other hand, is a very strong notion. If two variables X and Y are probabilistically independent conditional on M, then there is not a single X-value x and not a single Y-value y so that x is probabilistically sensitive to y in any context $M = m$. Conditional probabilistic independence ($INDEP_{Pr}$) is defined as the negation of conditional probabilistic dependence:

(2) $INDEP_{Pr}(X,Y|M)$ if and only if for all x, y, and m, $Pr(x|y,m) = Pr(x|m)$, provided $Pr(y,m) > 0$.

Unconditional probabilistic dependence/independence ($DEP_{Pr}(X,Y)$/$INDEP_{Pr}(X,Y)$) turns out to be a special case of conditional probabilistic dependence/independence; it can be defined as conditional probabilistic dependence/independence given the empty context $M = \varnothing$:

(3) $DEP_{Pr}(X,Y)$ if and only if $DEP_{Pr}(X,Y|\varnothing)$.
(4) $INDEP_{Pr}(X,Y)$ if and only if $INDEP_{Pr}(X,Y|\varnothing)$.

[3]The condition $Pr(y,m) > 0$ is needed because $Pr(x|y,m)$ is defined as $Pr(x,y,m)/Pr(y,m)$ and division by 0 is undefined.

3.2.3 Graphs and Causal Graphs

Let us turn to the concept of a causal graph. A *graph* G is an ordered pair $\langle V,E\rangle$, where V is a set of so-called vertices (which are statistical variables in causal graphs) while E is a set of so-called edges. Edges may be all kinds of arrows (e.g., "$\rightarrow$," "$\cdots>$," and "$\leftrightarrow$") or undirected links ("—") representing diverse binary relations among objects in V. Two variables in a graph's variable set V are called *adjacent* if and only if they are connected by an edge. A chain of $n \geq 1$ edges connecting two variables X and Y of a graph's variable set V is called a *path* between X and Y. A path of the form $X \rightarrow \ldots \rightarrow Y$ is called a *directed path* from X to Y. Whenever a path contains a subpath of the form $X \rightarrow Z \leftarrow Y$, then Z is called a *collider* on this path; the path is called a *collider path* in that case. X is called an *ancestor* of Y if and only if there is a directed path from X to Y; Y is called a descendant of X in that case. The set of all ancestors of a variable X is denoted by "$Anc(X)$," while the set of all descendants of X is indicated by "$Des(X)$." All X for which $X \rightarrow Y$ holds are called *parents* of Y; the set of all parents of Y is referred to via "$Pa(Y)$." All Y for which $X \rightarrow \ldots \rightarrow Y$ holds are called *children* of X; the set of all children of X is referred to via "$Chi(X)$." Variables to which no arrowhead is pointing are called *exogenous* variables. Non-exogenous variables are called *endogenous* variables. A graph $G = \langle V,E\rangle$ containing a path of the form $X \rightarrow \ldots \rightarrow X$ (with $X \in$ V) is called a *cyclic graph*; an *acyclic graph* is a graph that is not a cyclic graph. A graph $G = \langle V,E\rangle$ is called a *directed graph* if E contains only directed edges.

A graph becomes a *causal graph* as soon as its edges are interpreted causally. We will interpret "$X \rightarrow Y$" as "X is a *direct cause* of Y in causal graph G." X is a *cause* (i.e., a direct/indirect cause) of Y in G if and only if there is a causal chain $X \rightarrow \ldots \rightarrow Y$ in G.

3.2.4 Bayesian Networks and Causal Models

A *directed acyclic graph* (DAG) $G = \langle V,E\rangle$ and a probability distribution Pr over G's variable set V together become a so-called Bayesian network (BN) $\langle G,Pr\rangle$ if and only if G and Pr satisfy the *Markov condition*[4] (MC). If G is an acyclic causal graph, then G and Pr become an *acyclic causal model* (CM) if and only if G and Pr satisfy the *causal Markov condition*[5] (CMC) or d-separation[6]:

[4]Cf. Glymour et al. 1991, p. 156.

[5]Cf. Spirtes et al. 2000, p. 29.

[6]For a definition of d-separation see Spirtes et al. (2000, pp. 43f.). d-separation is equivalent with CMC for acyclic causal models. For a proof see Verma (1987).

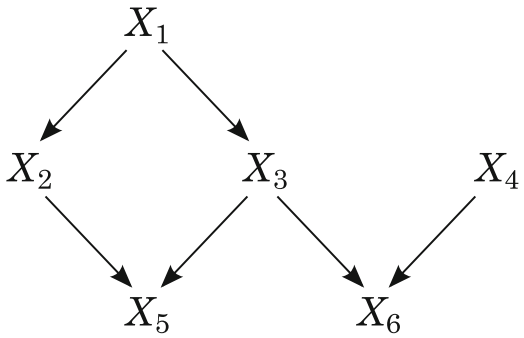


Fig. 3.1 A simple exemplary causal graph

(MC/CMC): $G = \langle V,E \rangle$ and Pr satisfy the (causal) Markov condition if and only if for all $X \in V$, $INDEP_{Pr}(X, V \backslash Des(X) | Pa(X))$.[7]

$V \backslash Des(X)$ is the set of all non-descendants of X. Note that "$Des(X)$" and "$Pa(X)$" in CMC refer to X's effects and X's direct causes, respectively, while "$Des(X)$" and "$Pa(X)$" are not causally interpreted at all in MC. The main idea behind CMC can be traced back to Reichenbach's *The Direction of Time* (1956).[8] It captures the strong intuition that conditioning on all common causes as well as conditioning on intermediate causes breaks down the probabilistic influence between two formerly correlated variables X and Y. Or in other words, the direct causes of a variable X contain all the probabilistic information which can be found under the causes of event types $X = x$; knowing the values of X's parents screens X off from all of its indirect causes.

We illustrate how CMC works by providing some examples. CMC implies for the DAG in Fig. 3.1, for instance, the following independence relations (as well as all probabilistic independence relations implied by them). These independence relations can directly be read off CMC applied to this DAG: $INDEP_{Pr}(X_1,X_4)$, $INDEP_{Pr}(X_2,\{X_3,X_4,X_6\}|X_1)$, $INDEP_{Pr}(X_3,\{X_2,X_4\}|X_1)$, $INDEP_{Pr}(X_4,\{X_1,X_2,X_3,X_5\})$, $INDEP_{Pr}(X_5,\{X_1,X_4,X_6\}|\{X_2,X_3\})$, and $INDEP_{Pr}(X_6,\{X_1,X_2,X_5\}|\{X_3,X_4\})$.

It follows from MC/CMC that the equation $Pr(X_1,\ldots,X_n) = \Pi_i\ Pr(X_i|pa(X_i))$[9] holds in every BN/acyclic CM $\langle V,E,Pr \rangle$ and, thus, that every BN/acyclic CM determines a fully defined probability distribution $Pr(X_1,\ldots,X_n)$ over the variable set V of this BN/acyclic CM. Hence, BNs/acyclic CMs allow for probabilistic

[7] In addition to MC/CMC, there are further principles of special interest when it comes to causal inference on the basis of empirical data (e.g., *causal sufficiency*, the *minimality condition*, and the *faithfulness condition*). For further details on these principles, see, for example, Spirtes et al. (2000) or Williamson (2005).

[8] See also Williamson (2010).

[9] Note that "$pa(X_i)$" stands for X_i's parents taking certain values, while "$Pa(X_i)$" stands for X_i's parents, that is, the variables which are X_i's direct predecessors in the corresponding graph.

reasoning about events which can be described in terms of the variables in V. Because $Pr(X_1, \ldots, X_n) = \Pi_i \; Pr(X_i|pa(X_i))$ holds in acyclic CMs, the conditional probabilities $Pr(X_i|pa(X_i))$ – which are called X_i's *parameters* – can represent the causal strengths of a variable X_i's direct causes. Note that $Pr(X_1, \ldots, X_n) = \Pi_i \; Pr(X_i|pa(X_i))$ and thus MC/CMC do not hold in cyclic CMs, either. It is because of this that in cyclic CMs there are always some variables whose parameters are undefined (these are the variables lying on a cyclic directed path) and, thus, that also the causal strengths of their direct causes are undefined in such models.

3.3 Biological Mechanisms

Before we can assess the strengths and shortcomings of causal graph theoretical models of biological mechanisms, we need to know what the main features of biological mechanisms are. In the last 15 years, philosophical interest in mechanisms has significantly increased. Those who endorse the mechanistic account place the concept of a mechanism at the heart of their philosophical analysis of scientific practice. They regard models of mechanisms as being involved in almost all scientific activities, let it be explanation, discovery, prediction, generalization, or intervention. There are still controversies in the debate with regard to how the notion of a mechanism should be specified, for instance, to which ontological category the components of a mechanism belong (Machamer et al. 2000; Tabery 2004; Torres 2008), whether the regular occurrence of the mechanism's behavior is a necessary condition (Bogen 2005; Craver and Kaiser 2013), or whether the concept of a mechanism can be extended such that it also accounts for the behavior of complex systems (Bechtel and Abrahamsen 2010, 2011) or for historical processes (Glennan 2010; see also Glennan's chapter in this volume). Despite these differences there are also many points of accordance. In what follows we will briefly present what are regarded as the major characteristics of biological mechanisms in the debate.

To begin with, a mechanism is always a mechanism *for* a certain behavior (Glennan 2002), for instance, the mechanism *for* protein synthesis or the mechanism *for* cell division. This is crucial because only those factors (i.e., entities and activities/interactions) that contribute to producing the specific behavior of the mechanism are said to be *components* of this mechanism.[10] An important consequence is that, although, for example, protein synthesis is the behavior of a cell, not all parts of the cell are also components of the mechanism for protein synthesis. Some parts of the cell (e.g., the centrosome and the cytoskeleton) are causally irrelevant for synthesizing proteins and thus do not count as components

[10]Craver calls these factors "constitutively relevant" and specifies this notion by his criterion of "mutual manipulability" (2007, pp. 139–159).

of the mechanism for protein synthesis.[11] In other words, the decomposition of a mechanism into its components depends on how the behavior of the mechanism is characterized (Kauffmann 1970; Craver and Darden 2001).

A second major characteristic of mechanisms is their *multilevel character*. The notion "multilevel character" refers to two distinct but related features of mechanisms: first, it appeals to the part-whole relation that exists between a mechanism and its components. This part-whole relation gives rise to the ontological claim that the mechanism as a whole is located on a higher level of organization[12] than the entities and activities/interactions that compose the mechanism. For instance, the mechanism for muscle contraction is said to be located on a higher level than the calcium ions, the sarcoplasmic reticulum, the myosin and actin molecules, etc., that interact with each other in a certain way (or that perform certain activities) in order to bring about the behavior of the mechanism as a whole (i.e., the contraction of the muscle fiber). Second, what is also meant by "multilevel character" is the fact that many mechanisms (in particular, in the biological realm) occur in nested hierarchies. Many mechanisms have components that are themselves (lower-level) mechanisms; and many mechanisms themselves constitute a component in a higher-level mechanism. For instance, the calcium pump that actively transports the calcium ions from the cytosol back into the sarcoplasmic reticulum is a part of the mechanism for muscle contraction. However, the calcium pump is also a mechanism on its own, namely, a mechanism for active transport of calcium ions. As such, it has its own components (e.g., A-, N-, and P-domain, transmembrane domain, calcium ions, ATP) with their own organization. Furthermore, the mechanism for muscle contraction constitutes itself a part in a higher-level mechanism, for instance, in the mechanism for crawling by peristalsis, a behavior that is exhibited, for example, by earthworms.

The third feature of mechanisms concerns their components. It is the one with respect to which there exists least conformity. The proponents of the mechanistic view concur that mechanisms consist of components, but they use different terminologies to classify the components, and some of them assign the components to different ontological kinds (whereas others are just not interested in metaphysical issues). For instance, Machamer et al. (2000) endorse the dualistic thesis that mechanisms are composed of entities *and* activities, which they conceive as two distinct ontological kinds. By contrast, Glennan (1996, 2002) characterizes mechanisms in a monist fashion, that is, as being constituted exclusively by entities that interact with each other and thereby change their properties. Other mechanists do not take a stand on this ontological dispute, but nevertheless draw the distinction between the

[11] However, the parts of the cell that are not components of the mechanism for protein synthesis may be components of other mechanisms. For instance, the centrosome and the cytoskeleton are components of the mechanism for cell division.

[12] We leave it open whether the notion of a level of organization must be spelled out in a mechanistic way, as, for example, Craver claims (2007, pp. 184–195). Alternatively, one could try to offer an account of levels, according to which levels are defined in not only local explanatory contexts but rather globally. In this spirit, for instance, Wimsatt takes levels to be local maxima of regularity and predictability (1976, 1994, and 2007).

spatial components of a mechanism and "what the spatial components are doing" or "the changes in which the spatial components are involved." Moreover, these authors adopt a different terminology to describe this difference. Bechtel (2006, 2008), for example, speaks of component parts and component operations (or functions). We think that it is not necessary (although legitimate) to become engaged in the ontological dispute about whether mechanisms consist of components that belong to one or to two distinct ontological kinds. One can avoid this dispute and yet argue that the two concepts – let it be entities and activities, entities and interactions, component parts and component operations, or whatever one likes – are *descriptively adequate*, that is, useful for representational purposes. When biologists represent mechanisms, they typically distinguish between the object itself (e.g., ribosome) and what the object is doing or the interactions in which the object is involved (e.g., binding, moving along the mRNA, releasing polypeptide). Thus, one should account for this difference when one models biological mechanisms. This, however, leaves open the ontological question of whether activities can be reduced to property changes of entities[13] or not. In sum, the third feature of mechanisms is that they are represented as having two kinds of components, entities and activities (or operations or interactions).

A fourth major characteristic of mechanisms is the importance of the *spatial and temporal organization* of their components for the functioning of the mechanism. Only if the components of a mechanism are organized in a specific way, the mechanism as a whole brings about the behavior in question. It is important to note that mechanisms are organized in a spatial as well as in a temporal manner. The spatial organization refers to the fact that certain entities are localized in certain regions of the mechanism, move from one region to another, and perform different activities in different regions. For instance, it is significant to the functioning of the mechanism of photosynthesis that the transport of electrons through the thylakoid membrane causes the transport of protons from the chloroplast stroma into the thylakoid lumen and that the resulting chemiosmotic potential is used for ATP synthesis by transporting the protons back into the stroma again. The temporal organization means that a mechanism is temporally divided into certain stages which have characteristic rates and durations as well as a particular order. Earlier stages give rise to latter stages so that there exists a "productive continuity" (Machamer et al. 2000, p. 3) between the stages of a mechanism. In other words, the activities or interactions are "orchestrated" (Bechtel 2006, p. 33) such that they produce the phenomenon of interest. Consider the mechanism of photosynthesis again. This mechanism is also characterized by a specific sequence of activities. The first step is the absorption of a photon (by the photosystem II). This causes the excitation of an electron, which is followed by the transport of this electron down the electron transport chain. This transport brings about the transport of protons and so on.

[13]Or, in the case of an activity that involves two entities, two events in which the change of one property of one object causes the property change of another entity.

At this point one could discuss further features of mechanisms, like the fact that most mechanisms produce a certain behavior in a regular way (given certain conditions) or that the components of mechanisms might be connected by a special kind of causal relations, namely, "productive causal relations" (Bogen 2008). However, these characteristics of mechanisms are far more controversial than the ones we have mentioned so far. This is why we do not take them for granted here. In what follows we examine the question of whether causal graph theoretical models of biological mechanisms are able to capture the major characteristics of mechanisms that we have presented in this section, namely, the multilevel character of mechanisms, their two kinds of components, and the spatial and temporal organization of their components. We do this by means of an extended analysis of an example from recent biological research. As announced before, the result of our analysis will be that causal graph theory succeeds with regard to some respects while failing with regard to others (Sect. 3.5). But before, we give a short introduction to the case study that we are concerned with (Sect. 3.4).

3.4 Feedback Inhibition of ACCase by 18:1-ACP in *Brassica napus*

Feedback inhibition is a common mode of metabolic control. Generally speaking, in feedback inhibition a product P produced late in a reaction pathway inhibits an enzyme E that acts earlier in the pathway and that transforms the substrate S into an intermediate product IP_1. Figure 3.2 illustrates this general connection.

Figure 3.2 shows that the substrate S is transformed in several steps into the product P (via the intermediate products $IP_1, \ldots, IP_n$). As P accumulates, it slows down and finally switches off its own synthesis by inhibiting the regulatory enzyme E that often catalyzes the first committed step of the pathway. That way, feedback inhibition prevents the cell from wasting resources by synthesizing more P than necessary. Because enzyme activity can be rapidly changed by allosteric modulators, feedback inhibition of regulatory enzymes provides almost instantaneous control of the flux through the pathway.

Many instances of this general mechanism of feedback inhibition can be found in nature. In this chapter, we focus on an example from contemporary botanical research, namely, on the feedback regulation of fatty acid biosynthesis in canola

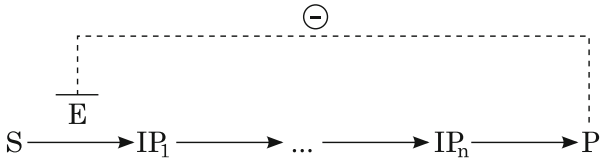


Fig. 3.2 The general mechanism for feedback inhibition

(*Brassica napus*), which has only recently been identified by Andre et al. (2012).[14] Fatty acid biosynthesis is a crucial process for both plants and animals, providing the cell with components for membrane biogenesis and repair and with energy reserves in specialized cells (such as epidermal cells or the cells of oilseeds). Since the need for fatty acids not only varies with the cell type but also depends on the stage of development, time of the day, or rate of growth, fatty acid biosynthesis must be closely regulated to meet these changes. Although the biochemistry of plant acid biosynthesis has been extensively studied,[15] comparatively little is known about its regulation and control (Ohlrogge and Jaworski 1997). However, knowing the mechanism of how fatty acid biosynthesis in plants is regulated is important, not least because it may give rise to the design of strategies for increasing fatty acid synthesis in plants (cf. Tan et al. 2011). This is particularly significant in light of the economic potential of genetically manipulated oil crops for improved nutritional quality or as renewable sources of petrochemical substitutes.[16]

The main aim of the experimental studies conducted by Andre et al. (2012) was to discover the feedback system that regulates the biosynthesis of fatty acids in the plastids of *Brassica napus*. The major results of their studies are twofold: first, they provide evidence for the hypothesis that plastidic acetyl-CoA carboxylase (in short, ACCase) is the enzymatic target of the feedback inhibition (i.e., the enzyme E that is inhibited). ACCase catalyzes the transformation of acetyl-CoA into malonyl-CoA. Second, their experiments indicate that the 18:1-acyl carrier protein (in short, 18:1-ACP) is the feedback signal, that is, the inhibitor of ACCase. On the basis of these findings, they proposed the mechanism for feedback inhibition of fatty acid synthesis in *Brassica napus* that is illustrated in Fig. 3.3.

The mechanism for feedback inhibition that takes place in the plastid (depicted in the upper, inner box) can be characterized as an instance of the general mechanism presented in Fig. 3.2. The enzyme ACCase (E) converts the substrate acetyl-CoA (S) into the intermediate product malonyl-CoA (IP_1), which is then transformed into the product 18:1-ACP (P). If the concentration of 18:1-ACP increases, more and more 18:1-ACP molecules bind to ACCase molecules and inhibit them. This, in turn, slows down and finally switches off the synthesis of further 18:1-ACP.

[14] Empirical work on similar regulation mechanisms, for instance, in tobacco suspension cells (Shintani and Ohlrogge 1995) and in *Escherichia coli* (Heath and Rock 1995; Davis and Cronan 2001), has been carried out before.

[15] For an overview about lipid biosynthesis, see, for instance, Ohlrogge and Browse (1995).

[16] Canola (*Brassica napus*) is the third largest source of vegetable oil supply. It is of high nutritional value (because of its high concentrations of unsaturated C18 fatty acids and a low level of erucic acid) and a suitable source for biodiesel fuels as well as for raw materials in industry (Tan et al. 2011).

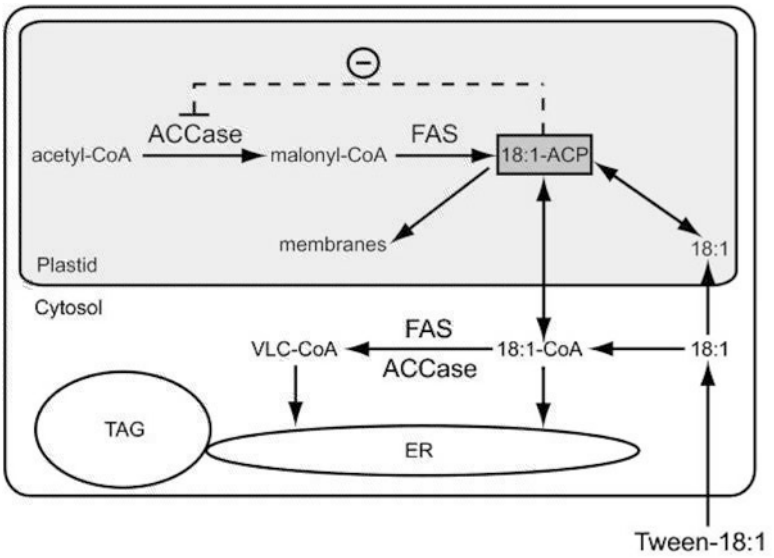


Fig. 3.3 Mechanism for feedback inhibition of fatty acid synthesis in *Brassica napus* (Reproduced from Andre et al. 2012)

3.5 Modeling the Mechanism for Feedback Inhibition

The mechanism presented in the previous section can be characterized as bringing about *the regulation of the synthesis of 18:1-ACP* (which is a fatty acid). One way to characterize this phenomenon in more detail is to specify it quantitatively: the concentration of 18:1-ACP is regulated such that it very likely does not reach a certain upper bound b (i.e., the probability for a concentration of 18:1-ACP lower than b is greater than a certain defined probability threshold r). Figure 3.4 shows an illustration.

3.5.1 A Causal Graph Theoretical Model of the Mechanism for Feedback Inhibition

How can the mechanism that brings about the regulation of fatty acid synthesis (more precisely, the regulation of the synthesis of 18:1-ACP) be represented within a causal graph framework? At first, we need to introduce a variable P, standing

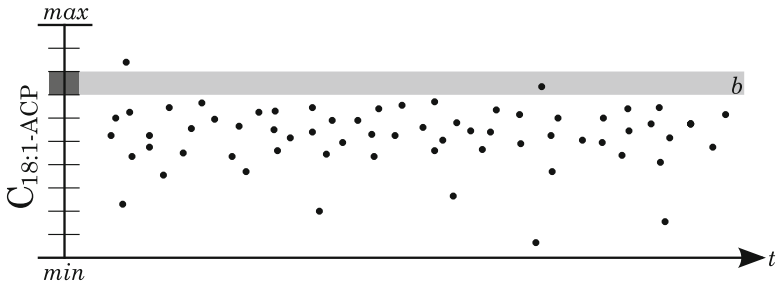


Fig. 3.4 The explanandum phenomenon of 18:1-ACP regulation [The dots stand for the 18:1-ACP concentrations ($C_{18:1\text{-}ACP}$) measured over time (t) (To be precise, the empirical data that biologists actually gather are not concentrations. Rather, they measure, for instance, optical densities (in spectrophotometric studies) and then draw inferences from the density values about the concentrations). More than r (95 % in this example) of 18:1-ACP concentrations measured so far do not exceed b]

for the concentration of the product 18:1-ACP.[17] P shall be a discrete variable fine-grained enough to correspond to the given measurement accuracy. The phenomenon may then be described as $Pr(p \leq b) > r$.

Furthermore, the concentration of the substrate acetyl-CoA (represented by variable S) is causally relevant for the 18:1-ACP concentration P: the higher the concentration of acetyl-CoA is, the *higher* will be the probability for higher 18:1-ACP concentrations. Another factor that is causally relevant for the 18:1-ACP concentration is the concentration of the regulatory enzyme ACCase. Here we have to distinguish between active enzymes and enzymes which bind the product 18:1-ACP (at the effector interaction site). We represent the former by the variable E_{active} and the latter by the variable $E_{P\text{-}bound}$. While the concentration of active enzymes is causally relevant to the concentration of the product 18:1-ACP (the higher E_{active}'s value, the *higher* the 18:1-ACP concentration), the 18:1-ACP concentration is causally relevant to the concentration of P-bound enzymes (the higher P's value, the *higher* $E_{P\text{-}bound}$'s value) which is, again, causally relevant to the concentration of active enzymes (the higher $E_{P\text{-}bound}$'s value, the *lower* E_{active}'s value), etc. The negative causal influence of $E_{P\text{-}bound}$ on E_{active} represents the fact that the binding of 18:1-ACP molecules to active ACCases causes the inhibition of the ACCases (i.e., the ACCases becoming inactive), and the negative causal influence of E_{active} on S stands for the fact that many active enzymes decrease the amount of the ACCases. According to these considerations, we may illustrate the mechanism by the causal graph depicted in Fig. 3.5.

To get a causal model, we have to supplement the causal graph depicted in Fig. 3.5 with a probability distribution Pr over variable set $V = \{S,P,E_{active},E_{P\text{-}bound}\}$.

[17] Note that variables are always represented by italic letters. The italic "P," for example, stands for a variable describing the concentration of the product 18:1-ACP, while the non-italic "P" stands for the concentration of the product 18:1-ACP itself.

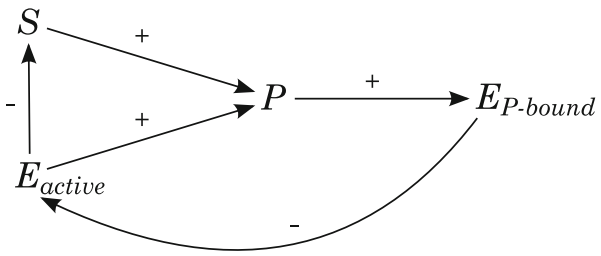


Fig. 3.5 Static cyclic CM of the mechanism for feedback inhibition [S and E_{active} are direct causes of P. P is a direct cause of $E_{P\text{-}bound}$ which is a direct cause of E_{active} which is, again, a direct cause of S and P, etc. Direct causal influences are represented by arrows. A plus ("+") above an arrow stands for a positive causal influence (i.e., high cause values lead to *high* effect values), and a minus ("−") stands for a negative causal influence (i.e., high cause values lead to *low* effect values)] (One might object that this causal graph is inadequate because it contains two variables that are analytically dependent, namely, $E_{P\text{-}bound}$ and E_{active}. We do not think that this is the case. $E_{P\text{-}bound}$ and E_{active} are *analytically independent* variables because there is a temporal distance between the binding of P to E and the inactivation of E (i.e., the conformational change of the substrate binding site). In other words, the binding of P to E and the inactivation of E are not the same processes occurring at the same time, but rather the former causes the latter. This is also why there exists a submechanism that specifies this causal relation)

Pr will imply that the probability of $p \leq b$ is greater than r (this is the phenomenon the mechanism brings about). The probabilities Pr will correspond to the positive/negative causal influences as described above. So the probability for high P-values, for example, will be high given high S- and E_{active}-values, and low given low S- or E_{active}-values.

The probabilities Pr are interpreted as inductively inferred limit tendencies of the observed frequencies of the diverse concentrations, as they are found under *normal* conditions. These *normal conditions* can be captured by adding a context $C = c$. This context is simply an instantiation of a variable or a set of variables which stand for the typical experimental setup and are not (or only slightly) changed during measuring or manipulating S, P, E_{active}, or $E_{P\text{-}bound}$. With regard to our case study, the context $C = c$ will include a certain temperature (or range of tolerable temperatures), a particular level (or tolerable range) of salinity, and a certain pH value (or range of tolerable pH values). The conditional probabilities along the causal arrows should correspond to the causal strengths of the variables' direct causes in context $C = c$.

Here we can observe the *first* problem of our causal model: while the parameters of a causal model are uniquely defined in an acyclic CM, this is not the case in cyclic CMs. This is a problem when it comes to explaining or predicting certain phenomena. We typically explain or predict a variable X's taking value x by means of this variable's direct or indirect causes and its parameters or the parameters of the variables lying between X and its indirect causes. So we explain or predict $X = x$ by reference to X's causes and only to X's causes and not to X's effects. But in our cyclic CM, some variable's causes are also their effects. P, for example, is a cause *and* an effect of $E_{P\text{-}bound}$. So conditioning on P does not correspond to the probabilistic influence of $E_{P\text{-}bound}$'s direct causes alone, but rather to a mixture

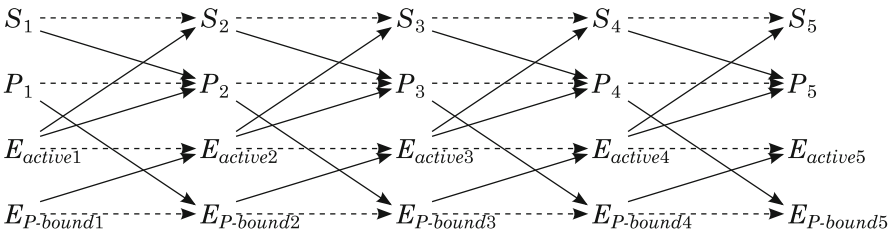


Fig. 3.6 The causal graph of a five-stage dynamic CM representing the mechanism for feedback inhibition in *Brassica napus*

of the probabilistic influences one gets from $E_{P\text{-}bound}$'s direct causes and some of its effects. In other words, conditioning on P does not give us the probabilistic influence of P on $E_{P\text{-}bound}$ transported only over path $P \rightarrow E_{P\text{-}bound}$, but the mixed probabilistic influence of P transported over $P \rightarrow E_{P\text{-}bound}$ and $E_{P\text{-}bound} \rightarrow E_{active} \rightarrow P$. A *second* problem of our causal model is that it does not capture the dynamic aspect of mechanisms – it does not show how the parts of the mechanism described influence each other over a period of time. A *third* deficit of our causal model is that it does not represent any hierarchic organization, that is, it does not account for the fact that mechanisms are often embedded in higher-level mechanisms and have parts that are (sub)mechanisms themselves (see Sect. 3.3). The above model just describes the causal relations that are responsible for bringing about the behavior of the mechanism, that is, it refers only to causes at one and the same ontological level and therefore (even if the first problem would not exist) does not, strictly speaking, allow for *interlevel* explanation/prediction. In order to cope with these three problems, in the next two subsections, we expand our causal model that represents the mechanism for feedback inhibition of fatty acid synthesis in *Brassica napus*.

3.5.2 Dynamic Causal Models

The first two problems discussed in the last section can be solved by unrolling the causal model over a period of time and thereby constructing a dynamic CM.[18] In doing so, we quite plausibly presuppose that causal influences need some time to spread and do not occur instantaneously. We get a dynamic CM if we add time indices to the variables of our system $V = \{S, P, E_{active}, E_{P\text{-}bound}\}$, representing the mechanism's diverse stages. By presupposing that causal influences need some time to take place, we can generate the dynamic CM whose causal graph is depicted in Fig. 3.6 (for five stages) on the basis of our static CM in Sect. 3.5.1.

The dashed arrows transport probabilistic influences (the substrate concentration S_i, for instance, is always probabilistically relevant to the substrate concentration

[18] Similar considerations can already be found in the first (but not in the second) edition of Spirtes et al. (2000).

at the next stage) in exactly the same way as their non-dashed counterparts. The only difference is that we interpret continuous arrows as direct causal connections while we want to leave it open whether the dashed arrows represent such causal connections. Dashed arrows could, for example, also be interpreted as analytic dependencies.[19] The variables of the five stages together with the continuous and the dashed arrows constitute the dynamic CM's causal graph.

The corresponding static CM's topological structure can be read off from the dynamic CM. One just has to abstract from the diverse stages of the dynamic CM and look at the continuous arrows: there has to be an arrow from S to P, from P to $E_{P\text{-}bound}$, from $E_{P\text{-}bound}$ to E_{active}, and from E_{active} to S and to P in the corresponding static CM, and these all have to be causal arrows in this static CM.

Note that the time intervals between two stages of a dynamic CM should be suitably chosen. On the one hand, if they are too small, then the causal influence may not have enough time to spread from the cause to the effect variable and correlations between causes and effects will get lost. On the other hand, these intervals should not be too large, either. This may lead to violations of very basic causal intuitions. To give an example, suppose the causal model in Fig. 3.6 shows the correct causal structure of the mechanism for feedback inhibition of fatty acid synthesis in *Brassica napus*. Then S is an indirect but not a direct cause of $E_{P\text{-}bound}$. S's causal influence on $E_{P\text{-}bound}$ is mediated via P. But if the interval between two stages were too large, say, for example, it were chosen such that stage 3 in the dynamic CM in Fig. 3.6 would be the next stage after stage 1, then S and $E_{P\text{-}bound}$ would be correlated and this correlation would not break down under conditionalization on the intermediate cause P. Thus, conditioning on an effect's direct causes would not screen it off from its indirect causes.

Dynamic CMs have some advantages over static CMs. First of all, they are acyclic CMs, and, thus, we can use the same methods as in BNs to compute the probabilities we are interested in. Furthermore, CMC holds and the causal model's parameters are defined. So we know the causal strengths of a variable's causes, and we can thus use dynamic CMs to explain certain phenomena which can be described by means of endogenous variables. So the *first* problem discussed in Sect. 3.5.1 can be solved: we can generate explanations and predictions by referring to the causes of the event of interest and to the probabilistic influence of these causes on this event. In addition, we can predict the probabilities of certain effects of interventions. We can, for example, predict the probability of certain P-concentrations at stage 5 given certain S- and E_{active}-concentrations at stage 1 when we change the concentration of S in a certain way at stage 3 via manipulation. The *second* problem can also be solved: the dynamic CM tells us how the parts of the mechanism described influence each other over a period of time, and we can thus also make predictions about what will (most likely) happen at later stages of the mechanism when we manipulate

[19] "Analytic dependence" is a notion that captures a wide range of noncausal dependences, for example, conceptual dependence, definitional dependence, and dependence which is due to a part-whole relation.

certain variables at earlier stages of the mechanism. Another nice feature of dynamic CMs is, provided the time intervals between the diverse stages of the mechanism are suitably chosen, that standard methods can be used for causal discovery because CMC holds for dynamic CMs. Causal discovery is still a serious problem for cyclic CMs, and there are only a few algorithms which, in general, do not lead to very detailed causal information (cf. Richardson 1996; Spirtes 1995). The third problem, however, still remains: our dynamic CM captures only causal information at one and the same ontological level and thus does not allow for interlevel mechanistic explanation, manipulation, and prediction.

3.5.3 Hierarchically Ordered Causal Models

There are at least two possibilities to represent the hierarchic organization of mechanisms within causal graph theory, that is, to solve the *third* problem that we mentioned at the end of Sect. 3.5.1. Each of these approaches has its own merits and deficits. One of these possibilities is developed in detail in Casini et al. (2011). Casini et al. provide a quite powerful formalism. They propose to start to represent a mechanism's top level by a causally interpreted BN. Such a BN's variable set V may then contain some so-called network variables. These are variables whose values are BNs themselves. Network variables (or, more precisely, the BNs which are their possible values) are intended to represent the possible states (e.g., "functioning" and "malfunctioning") of a mechanism's submechanisms. These BNs' variable sets may then themselves contain network variables which stand for the possible states of a submechanism's submechanisms and so on. To connect the diverse levels of the mechanism represented by such BNs, Casini et al. suggest an additional modeling assumption: the *recursive causal Markov condition* (RCMC). Whenever this condition holds, then Casini et al.'s formalism allows for probabilistic reasoning across the diverse levels of the represented mechanism.

In this chapter, we can discuss Casini et al.'s (2011) approach only very briefly. For a detailed discussion of their formalism see Gebharter (forthcoming). Though their formalism is definitely powerful, their crucial modeling assumption RCMC is quite controversial. First of all, it is neither obvious that RCMC holds in general, nor is it clear how one could distinguish cases in which it holds from cases in which it does not. Secondly, RCMC leads to contra-intuitive consequences. We have the strong intuition that learning information about a mechanism's microstructure should at least sometimes lead to better (or at least different) predictions of the phenomena this mechanism will bring about. This should be the case, for example, when the macro-variable describing the possible states of the mechanism is described in a quite coarse-grained way, while more and more knowledge about the mechanism's microstructure is collected. But, according to RCMC, a mechanism's micro-variables are probabilistically screened off from its macro-variables whenever the state of the submechanism represented by a network variable is known. A third deficit of Casini et al.'s approach is that it does not provide

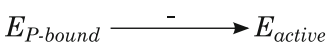


Fig. 3.7 Static CM of the phenomenon that is brought about by the submechanism for allosteric inhibition

any information about how a submechanism's microstructure is connected to the macrostructure of the overlying mechanism, that is, how exactly changes of some of the submechanism's micro-variables' values influence the mechanism's macro-variables due to probabilistic influences transported over its causal microstructure. Such information is crucial when it comes to the question of how macro-phenomena can be controlled by manipulating some of their underlying mechanisms' micro-variables.

In what follows we sketch an alternative approach for representing the hierarchic structure of mechanisms which avoids these problems. According to our approach, the submechanisms that a particular mechanism contains are, at least in most cases, adequately represented not via network variables, as Casini et al. (2011) propose, but via *causal arrows*. We will illustrate this claim on the basis of the case study that we have already introduced, namely, the mechanism for feedback inhibition of fatty acid synthesis in *Brassica napus*. This mechanism can be modeled within a causal graph framework as described in Sect. 3.5.1. An example for a submechanism of this mechanism is the mechanism for allosteric inhibition. This submechanism specifies the causal arrow between the variables $E_{P\text{-}bound}$ and E_{active} (see Fig. 3.7). That is, it describes how exactly the binding of the product 18:1-ACP (i.e., P) to the regulatory enzyme ACCase (i.e., E_{active}) causes the inhibition or inactivation of ACCase with the effect that ACCase cannot bind the substrate acetyl-CoA (i.e., S) and convert it into 18:1-ACP anymore. In other words, this submechanism discloses why it is the case that the higher the concentration of 18:1-ACP, the lower the concentration of active ACCase.

But how can such a submechanism be modeled within a causal graph framework, and how can it be related to the mechanism for feedback inhibition of which it is a part? In order to assess these questions, we need to go into more scientific details. Unfortunately, the biochemical submechanism that explains how the binding of 18:1-ACP to the enzyme ACCase ($E_{P\text{-}bound}$) causes the inhibition of ACCase (E_{active}) in *Brassica napus* has not been discovered yet (Andre et al. 2012). The same is true for the biochemical inhibition mechanisms in other species, for instance, in *Escherichia coli* (Heath and Rock 1995; Davis and Cronan 2001). However, in order to get an idea of how the model of the submechanism might look like, we will consider a different but analogous example, in which extensive molecular and structural studies have been carried out to unravel the biochemical mechanism of inhibition. In their recent work, Ganesan et al. (2009) investigated a different feedback system, namely, the allosteric inhibition of the enzyme serine protease (more precisely, of hepatocyte growth factor activator, in short "HGFA") by an antibody (Ab40). Their goal was to unravel the molecular details of this inhibition mechanism. That is, they aimed at characterizing the molecular interactions and

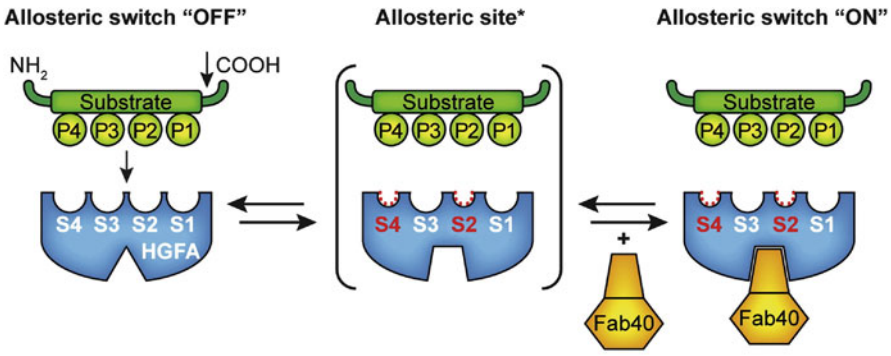


Fig. 3.8 Qualitative model of the mechanism for allosteric inhibition of HGFA by Fab40 (Fab40 is a special type of Ab40) (Adapted from Ganesan et al. 2009. With permission from Elsevier)

conformational changes that are caused by the binding of Ab40 (in general terms, of product P) to the effector interaction site of the enzyme HGFA (in general terms, to enzyme E) and that bring about the inhibition or deactivation of HGFA. Their work is very useful for our analysis because, on an abstract level, Ganesan et al. (2009) were interested in discovering the same submechanism as the one we singled out above, namely, the submechanism that explains how the binding of P to E causes the inhibition of E, in other words, why it is the case that the higher $E_{P\text{-}bound}$'s value, the *lower* E_{active}'s value.

The exact route by which the amino acids that compose E transmit the allosteric effect, that is, by which intermediate steps the binding of P to the remote effector interaction site of E causes the altered catalytic activity of E, is in general very poorly known (Sot et al. 2009). However, the structural and kinetic studies that Ganesan et al. (2009) performed produce some relief. One of their main results is that the binding of Ab40 (i.e., P) to the effector interaction site of HGFA (i.e., E) is accompanied by a major structural change (called the "allosteric switch"; Ganesan et al. 2009, p. 1620), namely, the movement of a certain part of the enzyme, the 99-loop, from the competent into the noncompetent conformation. This, in turn, obstructs the binding of the substrate to the enzyme E; more precisely, it causes a steric clash between the P2-Leu and the S2 subsite of E and the loss of stabilizing interactions between P4-Lys and the S4 subsite of E. The diagram in Fig. 3.8 provides a general illustration of these changes (while leaving out most of the molecular details).

The molecular interactions could be described in far more details. However, the foregoing description suffices for our purposes. How can this submechanism for allosteric inhibition of HGFA by Ab40 be modeled in a causal graph framework? We propose to model the submechanism with a static CM containing the variables and causal topology depicted in Fig. 3.9.

The first thing to note is that *B*, *99-loop*, *S2*, and *S4* are binary (and, thus, qualitative) variables. *B* can take one of the two values "bindings between functional groups of Ab40 and the effector interaction site of HGFA are established" and

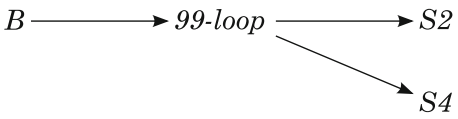


Fig. 3.9 Static CM of the submechanism for allosteric inhibition of HGFA by Fab40

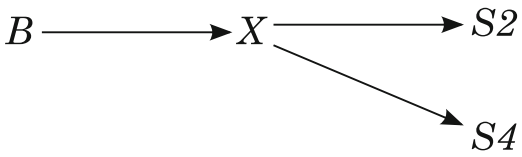


Fig. 3.10 Static CM of the hypothetical submechanism for allosteric inhibition of ACCase by 18:1-ACP (The corresponding possible values of the variables are the following: *B* can take one of the two values "bindings between functional groups of 18:1-ACP and the effector interaction site of ACCase are established" and "bindings between functional groups of 18:1-ACP and the effector interaction site of ACCase are not established." *X* can take one of the two values "being in the competent state" and "being in the noncompetent state." S2 and S4 can take one of the two values "having an ideal conformation that allows its binding to a certain part of 18:1-ACP" and "having a deformed conformation that inhibits its binding to a certain part of 18:1-ACP.")

"bindings between functional groups of Ab40 and the effector interaction site of HGFA are not established." *99-loop* can take one of the two values "being in the competent state" and "being in the noncompetent state." *S2* can take one of the two values "having an ideally shaped hydrophobic pocket to recognize P2-Leu" and "having a deformed pocket so that P2-Leu cannot be recognized." *S4* can take one of the two values "being able to perform stabilizing interactions to P4-Lys" and "being unable to perform stabilizing interactions to P4-Lys." This model describes that if bindings between functional groups of Ab40 and the effector interaction site of HGFA are established, then the probability is high that 99-loop is in its competent state, which is why the probability is high that S2 has an ideally shaped hydrophobic pocket to recognize Leu and S4 is able to perform stabilizing interactions to P4-Lys. On the higher level, we would say that if P (Ab40) binds to E (HGFA), this submechanism brings about the behavior that E (HGFA) is inactive (which means, on the lower level, that the two amino acids P2-Leu and P4-Lys of the substrate cannot bind to the substrate binding sites S2 and S4 of the enzyme (HGFA)).

We are aware of the fact that it is very unlikely that the biochemical submechanism for the inhibition of ACCase by 18:1-ACP in *Brassica napus* looks exactly like the submechanism for the inhibition of HGFA by Ab40, which we just described. There are too many molecular differences between the two enzymes and the two inhibitory products. However, for the sake of the argument, suppose that also in the case of the inhibition of ACCase, the binding of 18:1-ACP causes the movement of some part of the enzyme X from a competent state into a noncompetent state. Suppose further that this allosteric switch brings about certain molecular and conformational changes in two substrate binding sites S2 and S4 of the enzyme ACCase, which prevent the substrate to bind to the enzyme. A static CM of this hypothetical submechanism would look like the one in Fig. 3.10.

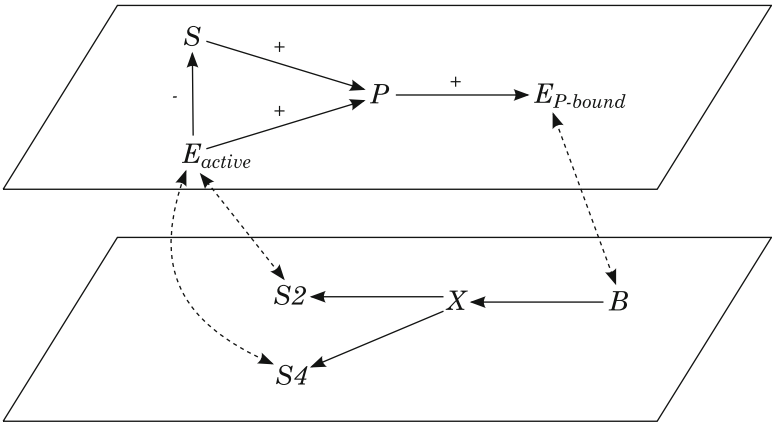


Fig. 3.11 Hierarchic static CM of the mechanism for feedback inhibition and of one of its submechanisms, namely, the biochemical mechanism for allosteric inhibition

On this basis we can now tackle the crucial question of how the model of the mechanism for feedback inhibition, which we developed in Sect. 3.5.1, and the model of one of its submechanisms, namely, of the biochemical mechanism of allosteric inhibition, can be related within a causal graph framework. We propose to model the hierarchic order of this multilevel mechanism by means of a *hierarchic static causal model* with the topological structure depicted in Fig. 3.11.

The two-headed arrows between $E_{P\text{-}bound}$ and B as well as between $S2$ and E_{active} and $S4$ and E_{active} which connect the two levels of the two mechanisms do not stand for causal, but rather for constitutive relevance relations, for instance, in the sense of Craver (2007). Hence, they transport probabilistic dependencies and the effects of manipulations in the same way as direct causal loops in static CMs. Note that the causal arrow $E_{P\text{-}bound} \rightarrow E_{active}$ in our original static CM disappeared in the hierarchic causal model. It is replaced by the underlying mechanism of this causal arrow, that is, by a causal structure whose input and output variables are connected to $E_{P\text{-}bound}$ and E_{active}, respectively, via constitutive relevance relations in Fig. 3.5. Also note that it is not clear how the submechanism represented by $E_{P\text{-}bound} \rightarrow E_{active}$ could be analyzed in Casini et al.'s (2011) approach. They would need to add a network variable N between $E_{P\text{-}bound}$ and E_{active} ($E_{P\text{-}bound} \rightarrow N \rightarrow E_{active}$). But then and because there is no intermediate (macro-level) cause N between $E_{P\text{-}bound}$ and E_{active}, it is unclear what this network variable N should represent at the mechanism's macro-level.

Our hierarchic static CM can be used for mechanistic reasoning[20] across diverse levels. In contrast to Casini et al.'s (2011) models, our model also tells us how

[20]The main difference between mechanistic reasoning and causal reasoning is that mechanistic reasoning makes use not only of causal but also of constitutive relevance relations. In other words, mechanistic reasoning contains not only *intra*level reasoning but also *inter*level reasoning.

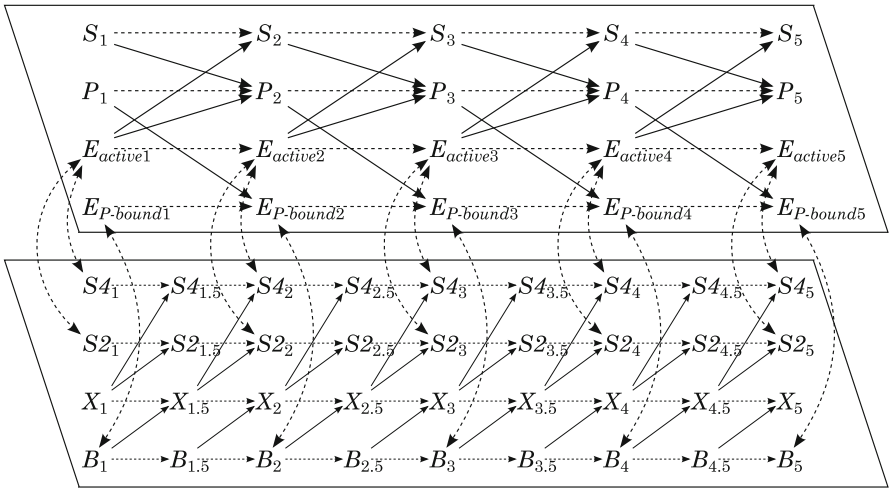


Fig. 3.12 Hierarchic dynamic causal model of the mechanism for feedback inhibition and the biochemical mechanism for allosteric inhibition

exactly probabilistic influence between macro-variables is transported over the underlying mechanism's causal microstructure and how exactly (i.e., over which causal and/or constitutive relevance paths) manipulations of micro-variables influence certain macro-variables. For example, if we manipulate *S4*, this will change E_{active} and *S2* because *S4* and *S2* are constitutively relevant for E_{active}. Since *X* is a direct cause of *S4*, changing *S4* will, on the other hand, not have a direct influence on *X*'s value. But changing *S4* will nevertheless have a quite indirect influence on *X*: a change of *S4*'s value will have an influence on E_{active}'s value at the macro-level, which influences its macro-level effect $E_{P\text{-}bound}$. Since *B* is constitutively relevant for $E_{P\text{-}bound}$, $E_{P\text{-}bound}$-changes will lead to *B*-changes which will, since *B* is a direct cause of *X* at the micro-level, lead to certain *X*-changes.

Though such hierarchic models as the one depicted in Fig. 3.11 can be used for probabilistic reasoning across a mechanism's diverse levels, they cannot generally be used for explanation and prediction. The reason is the same as in the case of static CMs, as illustrated in Sect. 3.5.1: a certain $E_{P\text{-}bound}$-value, for example, can be explained or predicted only via reference to $E_{P\text{-}bound}$'s causes, for example, *P*. But in our hierarchic static CM, *P* does influence $E_{P\text{-}bound}$ not only as a cause but also as an effect: *P* influences $E_{P\text{-}bound}$ not only over $P \rightarrow E_{P\text{-}bound}$ but also over $P \leftarrow E_{active} <\cdots> S2 \leftarrow X \leftarrow B <\cdots> E_{P\text{-}bound}$ and $P \leftarrow E_{active} <\cdots> S4 \leftarrow X \leftarrow B <\cdots> E_{P\text{-}bound}$. So the probabilistic influence of *P* on $E_{P\text{-}bound}$ does not correspond to *P*'s causal influence on $E_{P\text{-}bound}$ alone. We can solve this problem by rolling out our hierarchic model over time as we have already done for our original static CM in Sect. 3.5.2. Figure 3.12 is an illustration of the result of this procedure.

Note that, while causal influences need some time to spread, value changes produced by constitutive relevance relations occur instantaneously. Because of this,

the two-headed dashed arrows representing such constitutive relevance relations only connect variables at one and the same stage. This also corresponds to the fact that one cannot change one of two constitutively dependent variables without changing the other. Note also that the causal arrows from $E_{P\text{-}bound}$ to E_{active} disappeared in the hierarchic dynamic CM. This is because these arrows represented a submechanism at work which is explicated in more detail in the hierarchic dynamic CM – the hierarchic dynamic CM tells us exactly (and, in contrast to our original dynamic CM developed in Sect. 3.5.2, in a mechanistic way)[21] how $E_{P\text{-}bound}$ influences E_{active} and thus finally solves problem three, too: hierarchic dynamic CMs allow for probabilistic interlevel explanation and prediction of certain E_{active}-values. Certain E_{active}-values, for instance, can be mechanistically explained or predicted by certain $E_{P\text{-}bound}$-values: $E_{P\text{-}bound}$ at stage 1 has some influence on its constitutive part B at stage 1. B at stage 1 causes X at the micro-level at stage 1.5 which causes *S2* and *S4* at the micro-level at stage 2, and, since *S2* and *S4* are constitutively relevant for E_{active}, they have a direct probabilistic influence on E_{active} at stage 2.

One could object that, since the two-headed dashed arrows in our hierarchic dynamic CM transport the influences of interventions in both directions, CMC does not hold in such models and, hence, they should have the same problems as static CMs when it comes to explanation and prediction. The first point of such an objection is definitely true: CMC does not hold for hierarchic dynamic CMs.[22] However, this does not lead to the suspected consequence. The problem for explanation and prediction in static CMs was that the probabilities one gets when conditioning on some variables also provide some information which can only be achieved if one also knew these variables effects (in other words, probabilistic information is transported not only over cause paths but also over effect paths). But the events that we want to explain do not occur *because* some of their effects occurred (i.e., because they had a probabilistic influence on them), and events we want to predict cannot be predicted via reference to some of their effects (which have not occurred yet). However, this problem does not arise for hierarchic dynamic CMs. In a hierarchic dynamic CM, cycles appear only due to constitutive relevance relations within certain stages, and, thus, conditioning on a variable's causes does *only* provide probabilistic information about this variable's values transported over cause or constitutive relevance paths. It *never* provides probabilistic information transported over an effect path.

[21] Note also that Casini et al.'s (2011) approach does not allow for mechanistic reasoning in this sense. In their approach, the question of how two or more macro-variables (e.g., $E_{P\text{-}bound}$ and E_{active} in our example) influence each other can only be answered by causal connections at the macro-level. In our approach, on the other hand, we can explain such an influence by reference to the underlying mechanism – we can tell a story about how $E_{P\text{-}bound}$ influences E_{active} by demonstrating how $E_{P\text{-}bound}$'s constitutively relevant parts causally influence E_{active}'s constitutively relevant parts at the micro-level.

[22] Note that d-separation may still be assumed to hold.

3.6 Merits and Limits of Causal Graph Theoretical Models

On the basis of the preceding analysis, we can now approach the question of whether causal graph theory is suited for modeling biological mechanisms and what the advantages and shortcomings of representing mechanisms within a causal graph framework are. In the previous literature the concern has been raised that, even if it is possible to provide causal graph theoretical models of biological mechanisms, they are deficient because they fail to comprise some important kinds of information. In this line, for instance, Weber (2012) argues that because causal graph theoretical models only encompass sets of variables and relations of causal dependence, they fail to include information about the structure of biological entities (such as information about the DNA double helix topology and the movements undergone by a replicating DNA molecule) and about their spatiotemporal organization. However, claims like these remain on a quite general level. Our goal in this section is to use the results of our analysis of the case study in the previous section in order to assess and to specify these claims. We do so by pointing out which kinds of information about biological mechanisms cannot or can only insufficiently be represented within a causal graph framework and what are the reasons for these failures. In addition to revealing the *limitations* of causal graph theoretical models of mechanisms, we also highlight the *virtues* they have with respect to certain scientific purposes.

To begin with, recall the major characteristics of biological mechanisms that we identified in Sect. 3.3. First, mechanisms possess a multilevel character, which means, on the one hand, that there exists a part-whole relation between the mechanism and its components and, on the other hand, that mechanisms frequently occur in nested hierarchies. Second, mechanisms are represented as having two different kinds of components: entities (having particular properties) and activities (or interactions, operations, etc.). Finally, a mechanism brings about a specific behavior only if its components are spatially and temporally organized in a certain way. Can all these three features of biological mechanisms adequately be represented by causal graph theoretical models?

Consider first the *multilevel character* of mechanisms. As we have shown in the previous section, the fact that many mechanisms occur in nested hierarchies (i.e., that they are embedded in higher-level mechanisms and have components that are themselves submechanisms) can be represented in at least two ways. On the one hand, one can represent a mechanism's submechanisms by so-called network variables, as, for instance, Casini et al. (2011) do. We, on the other hand, think that there are good reasons for representing such submechanisms by causal arrows between variables X and Y. In our approach one can generate a hierarchic causal model by replacing such a causal arrow by another causal structure. This causal structure should be on a lower ontological level than X and Y, it should contain at least one constitutively relevant part of X and at least one of Y, and there should be at least one causal path going from the former to the latter at the micro-level. Such hierarchic models allow, in contrast to purely qualitative models, for probabilistic mechanistic reasoning *across different levels*. Hierarchic dynamic CMs

do even allow for probabilistic mechanistic interlevel explanation and prediction. Contrary to Casini et al.'s models, they can also provide detailed information about how certain causal influences at the macro-level are realized by their underlying causal influences propagated at the micro-level. This is important when it comes to questions about how certain manipulations of macro- or micro-variables influence certain other macro- or micro-variables of interest and how a mechanism's causal microstructure is connected to its macrostructure.

Let us now turn to the second feature of mechanisms. Do causal graph theoretical models succeed in representing mechanisms as being composed of two different kinds of components, namely, *entities and activities* (or operations, interactions, etc.)? It is quite clear that causal models represent entities. Precisely speaking, the individuals in the domains $D_{X1},\ldots,D_{Xn}$ of the causal model's variables $X_1,\ldots,X_n$ represent the entities that are components of the mechanism. Furthermore, the variables $X_1,\ldots,X_n$ taking certain values represent different properties or different behaviors of these entities. But can causal graph theoretical models represent activities, too?

A convenient first step towards an answer to this question seems to be to scrutinize the activities that are involved in our case study. Examples of activities that are part of the mechanism for feedback inhibition of fatty acid synthesis in *Brassica napus* are the *binding* of 18:1-ACP (P) to ACCase (E), the *transformation* of acetyl-CoA (S) into 18:1-ACP (P) (via the intermediate product malonyl-CoA), and the *inhibition* of ACCase (E) by 18:1-ACP (P) (see description of Fig. 3.5). The submechanism that brings about the activity of the inhibition of ACCase by 18:1-ACP is, in turn, composed of the following micro-activities: the *establishment* of a certain kind of binding between a functional group of 18:1-ACP and the effector interaction site of ACCase, the *shifting* the conformation of a particular part of ACCase, the *deformation* of the conformation of the S2 part of the substrate binding site of ACCase, etc. (see description of Fig. 3.9). What all these activities have in common is that they are temporally extended *processes* that involve some kind of *change*. Correspondingly, Machamer et al. have characterized activities as being "the producers of change" (2000, p. 3). It should be noted that not all activities must involve interactions between two or more distinct entities.[23] There might also be activities (so-called noninteractive activities (Tabery 2004, p. 9; Torres 2008, p. 246), like the shifting of the conformation of a particular part of ACCase) that involve only one entity (i.e., the particular part of ACCase) and a change of its properties (i.e., from the property "being in a competent state" to "being in a noncompetent state").[24] In any case, activities involve the change of properties. In

[23] According to Glennan (2002, p. 344), an interaction is an occasion on which a change in a property of one component of the mechanism brings about a change in a property of another component.

[24] As mentioned in Sect. 3.3, we leave it open whether activities can be reduced to state transformations via property changes or whether there is something lost by this reduction (such as the *productive nature* of activities).

principle, the variables of a causal graph theoretical model could just be chosen in such a way that the different values they can take represent different processes or changes of properties. However, such a choice of variables would completely be at odds with experimental practice in biology. In most cases it is difficult or even impossible to measure entire processes by just measuring once. Rather, what biologists do, for instance, to collect empirical data about the inhibition of ACCase by 18:1-ACP, is that they measure the concentration of the product (which is an indicator of ACCase's activity and, thus, also of its inhibition) to different times. Against this background it would be inadequate to choose the variable in such a way that one of its values represents the entire process/activity of inhibition of ACCase by 18:1-ACP. The option of representing activities simply by variables taking certain values can also be ruled out by the following argumentation: if activities were represented by variables taking certain values, then activities would neither involve changes nor be productive – they would rather occur due to other productive causal relations. Since activities are productive and involve changes, they must be represented differently.

We think that there are two ways in a causal graph theoretical model by which the activities that compose a mechanism can be captured: they can either be represented by *causal arrows* between variables. For instance, the causal arrow between *S* and *P* in Fig. 3.5 represents the activity "transformation of acetyl-CoA into 18:1-ACP." This is the option that matches the neat picture that several authors seem to have in mind: in a causal model the variables represent the entities (and their possible properties), and the arrows represent the activities. However, our analysis shows that things are not that neat. There is a second, equally adequate way to represent activities in causal graph theoretical models, namely, representing them by the *change of the value of a variable*. For instance, the activity "shifting the conformation of a particular part of ACCase" is represented in Fig. 3.9 by the variable *X*, changing its value from "being in a competent state" to "being in a noncompetent state."

A related view of static CMs, which we have to give up, is the neat view that the different variables in static CMs always represent the possible properties and activities of *distinct* entities. The flexibility of the choice of variables allows that one static CM contains variables that represent different possible properties (and activities) of *the same* entity. For instance, in our static CM depicted in Fig. 3.5, the variables $E_{P\text{-}bound}$ and E_{active} both refer to the concentrations of enzymes but describe different properties of these enzymes, namely, "being bound to P" and "being active." In other words, in causal graph theoretical models, the boundaries between different entities and between entities and activities often become *fuzzier* than in qualitative models. This fuzziness may have the disadvantage of impeding the understanding of how a mechanism brings about a certain phenomenon – when one looks at a static CM or at a dynamic CM, one does not recognize at first sight what the entities are and which activities they perform.

To conclude, we think that it is possible to represent mechanisms as being composed of entities and activities in a causal graph framework. However, what

one does not get are neat static CMs in which each variable represents a distinct entity and the arrows represent activities. This might be disadvantageous for some purposes, but not for others.

Finally, how do things stand with the third main feature of mechanisms, namely, with the *spatial and temporal organization* of their components? How much and which structural and spatial information one actually represents simply depends on one's choice of variables. In our case study, for instance, the causal graph theoretical model depicted in Fig. 3.11 contains structural as well as spatial information: the variable *S2*, for example, refers to a particular entity, namely, the S2 part of the substrate binding site of ACCase, and to the two possible *structural* properties that this entity can exhibit, namely, "having an ideal conformation that allows its binding to a certain part of 18:1-ACP" and "having a deformed conformation that inhibits its binding to a certain part of 18:1-ACP."[25] A different example is the variable $E_{P\text{-}bound}$ which represents the concentration of those regulatory enzymes (ACCases) that are bound to, that is, *spatially* connected to, the product 18:1-ACP. Hence, it is possible to include certain crucial structural and spatial information about the components of a mechanism into a causal graph theoretical model – one just has to choose variables that refer to structural and spatial properties.

Information about the temporal organization can be captured by and read off from the causal arrows of dynamic CMs: in the example we discussed in Sect. 3.5.2, for instance, *S* at stage 1 causes *P* at stage 2, which causes $E_{P\text{-}bound}$ at stage 3. So *at first* S interacts with P, *then* P interacts with $E_{P\text{-}bound}$, etc. However, even if there are no in-principle reasons for why it is impossible to include all the details of the spatial and temporal organization of a mechanism's components into a causal graph theoretical model, this does not preclude that there may be heuristic reasons for doing so. For instance, including all the relevant spatial, structural, and dynamic information might give rise to a causal model that includes too many different variables, so that it is unmanageable and thus not useful.

In sum, causal graph theoretical models *can* account for the three main features of mechanisms. However, they do so in a quite abstract way, which is why they are far worse than purely qualitative models with respect to the purpose of providing understanding. Qualitative models tell us in a very intelligible way how the components of a mechanism interact to bring about the phenomenon of interest. They make, contrary to probabilistic causal models, clear distinctions between the macro- and the micro-level (i.e., between mechanisms and their submechanisms) and between distinct entities and activities (or operations, interactions, etc.). Purely qualitative models of mechanisms can also be used to explain certain behaviors of systems by revealing how the components of a mechanism bring about the behavior in question. These qualitative models are, however, limited. They fail when it comes to explain why certain systems frequently (but not always) bring about certain behaviors. In other words, they fail when it comes to explaining probabilistic

[25] Of course, these two properties could be and, in fact, are specified in more detail in biological practice. We give this general and brief characterization just for heuristic reasons.

phenomena like the phenomenon described in Sect. 3.5. Moreover, they do not allow for probabilistic prediction and (interlevel) manipulation. But knowing how we can bring about a particular phenomenon with high probability is a crucial investigative strategy in the biological sciences. Finally, purely qualitative models fail to integrate qualitative information with quantitative, probabilistic information. The latter is an important task in certain research areas like epigenetics where laboratory molecular experiments need to be brought together with ecological or evolutionary observational studies and computer simulations.

3.7 Conclusion

In this chapter, we have shown how the formal framework of causal graph theory can be used to model biological mechanisms in a probabilistic and quantitative way. Our analysis of the mechanism for feedback regulation of fatty acid biosynthesis in *Brassica napus* revealed that causal graph theoretical models can be extended such that they can also account for more complex forms of organization of the components of a mechanism (like feedback) as well as for the fact that mechanisms are frequently organized into nested hierarchies. We argued that, because causal graph theoretical models are not purely qualitative, but rather include probabilistic and quantitative information, they are useful in the context of causal discovery – in particular if one wants to make quantitative, probabilistic predictions or conduct manipulations. What is more, since causal graph theoretical models allow us to represent different levels of mechanisms in the same model (e.g., a mechanism, one of its submechanisms, and the relations between them), they enable us to carry out *interlevel* mechanistic manipulation and prediction, too.

However, our analysis of the case study did not only disclose advantages of representing biological mechanisms within a causal graph framework. Rather, it gave rise to the more *balanced view* that probabilistic, quantitative models of mechanisms – although there are clear merits with respect to some purposes – also have shortcomings with respect to other purposes. Accordingly, our analysis revealed that causal graph theoretical models have the resources to represent the three main features of biological mechanisms, namely, their multilevel character, their two kinds of components, and the spatial and temporal organization of their components. However, it also became clear that in some respects probabilistic, quantitative models of mechanisms are insufficient (e.g., because the boundaries of entities and between entities and activities become fuzzy and because the amount of structural/spatial and dynamical information that can be represented is limited) which makes them inadequate for some purposes (in particular for providing understanding). With this analysis we hope to have shed some light on the merits and limitations of modeling biological mechanisms within a causal graph framework and to have provided some interesting prospect for future philosophical work.

Acknowledgments We would like to thank the members of the research group "Causation, Laws, Dispositions, and Explanation at the Intersection of Science and Metaphysics" (FOR 1063), the participants of the colloquia at the University of Cologne and at the University of Düsseldorf, and the members of the Lake Geneva Biological Interest Group at the University of Geneva for their helpful comments on earlier drafts. This project was made possible by the funding provided by the Deutsche Forschungsgemeinschaft (DFG).

References

Andre, C., Haslam, R. P., & Shanklin, J. (2012). Feedback regulation of plastidic acetyl-CoA carboxylase by 18.1-acyl carrier protein in *Brassica napus*. *PNAS, 109*(25), 10107–10112.

Baedke, J. (2012). Causal explanation beyond the gene: Manipulation and causality in epigenetics. *Theoria, 74*, 153–174.

Bechtel, W. (2006). *Discovering cell mechanisms. The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (2008). *Mental mechanisms. Philosophical perspectives on cognitive neuroscience*. New York/London: Taylor and Francis Group.

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A, 41*(3), 321–333.

Bechtel, W., & Abrahamsen, A. (2011). Complex biological mechanisms: Cyclic, oscillatory, and autonomous. In C. A. Hooker (Ed.), *Philosophy of complex systems* (Handbook of the philosophy of science, Vol. 10, pp. 257–285). New York: Elsevier.

Bogen, J. (2005). Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 397–420.

Bogen, J. (2008). Causally productive activities. *Studies in History and Philosophy of Science Part A, 39*(1), 112–123.

Casini, L., Illari, M. P., Russo, F., & Williamson, J. (2011). Models for prediction, explanation, and control: Recursive Bayesian networks. *Theoria, 70*, 5–33.

Craver, C. F. (2007). *Explaining the brain. Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.

Craver, C., & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in neuroscience* (pp. 112–137). Pittsburgh: University of Pittsburgh Press.

Craver, C., & Kaiser, M. I. (2013). Mechanisms and laws: Clarifying the debate. In H.-K. Chao, S.-T. Chen, & R. L. Millstein (Eds.), *Mechanism and causality in biology and economics* (pp. 125–146). Dordrecht: Springer.

Darden, L. (2006). *Reasoning in biological discoveries*. New York: Cambridge University Press.

Darden, L. (2008). Thinking again about mechanisms. *Philosophy of Science, 75*(5), 958–969.

Davis, M. S., & Cronan, J. E., Jr. (2001). Inhibition of *Escherichia coli* acetyl coenzyme a carboxylase by acyl-acyl carrier protein. *Journal of Bacteriology, 183*(4), 1499–1503.

de Sol, A., Tsai, C.-J., Ma, B., & Nussinov, R. (2009). The origin of allosteric functional modulation: Multiple pre-existing pathways. *Structure, 17*, 1042–1050.

Ganesan, R., Eigenbrot, C., Wu, Y., Liang, W.-C., Shia, S., Lipari, M. T., & Kirchhofer, D. (2009). Unraveling the allosteric mechanism of serine protease inhibition by an antibody. *Structure, 17*, 1614–1624.

Gebharter, A. (forthcoming). A formal framework for representing mechanisms?. Philosophy of Science.

Glennan, S. S. (1996). Mechanism and the nature of causation. *Erkenntnis, 44*, 49–71.

Glennan, S. S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, 342–353.

Glennan, S. S. (2005). Modeling mechanisms. *Studies in the History and Philosophy of Biological and Biomedical Sciences, 36*(2), 443–464.
Glennan, S. S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*(2), 251–266.
Glymour, C., Spirtes, P., & Scheines, R. (1991). Causal inference. *Erkenntnis, 35*, 151–189.
Heath, R. J., & Rock, C. O. (1995). Regulation of malonyl-CoA metabolism by acyl-acyl carrier protein and β-ketoacyl-acyl carrier protein synthases in *Escherichia coli*. *The Journal of Biological Chemistry, 270*(26), 15531–15538.
Jon, W. (2005). *Bayesian nets and causality*. Oxford: Oxford University Press.
Kauffman, S. A. (1970). Articulation of parts explanation in biology and the rational search for them. *Boston Studies in the Philosophy of Science, 8*, 257–272.
Machamer, P., Darden, L., & Carver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Ohlrogge, J. B., & Browse, J. G. (1995). Lipid biosynthesis. *The Plant Cell, 7*, 957–970.
Ohlrogge, J. B., & Browse, J. G. (1997). Regulation of fatty acid synthesis. *Annual Review of Plant Physiology and Plant Molecular Biology, 48*, 109–136.
Perini, L. (2005). Explanation in two dimensions: Diagrams and biological explanation. *Biology and Philosophy, 20*, 257–269.
Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.
Richardson, T. (1996). A discovery algorithm for directed cyclic graphs. In E. Horvitz & F. Jensen (Eds.), *Proceedings of the 12th conference on uncertainty in artificial intelligence* (pp. 454–461). Portland: Morgan Kaufmann.
Shintani, D. K., & Ohlrogge, J. (1995). Feedback inhibition of fatty acid synthesis in tobacco suspension cells. *The Plant Journal, 7*(4), 577–587.
Skipper, R. A., & Millstein, R. L. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in the History and Philosophy of Biological and Biomedical Sciences, 36*(2), 327–347.
Spirtes, P. (1995). Directed cyclic graphical representations of feedback models. Proceedings of the 11th conference on uncertainty in artificial intelligence (pp. 491–498). San Francisco: Morgan Kaufmann.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge: MIT Press.
Tabery, J. G. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of Science, 71*, 1–15.
Tan, H., Yang, X., Zhang, F., Zheng, X., Qu, C., Mu, J., Fu, F., Li, J., Guan, R., Zhang, H., Wang, G., & Zou, J. (2011). Enhanced seed oil production in Canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiology, 156*, 1577–1588.
Torres, P. J. (2008). A modified conception of mechanisms. *Erkenntnis, 71*(2), 233–251.
Verma, T. (1987). Causal networks: semantics and expressiveness (Tech. Rep.). Cognitive Systems Laboratory, University of California.
Weber, M. (2012). Experiment in biology. In E. N. Zalta (Ed.). The Stanford encyclopedia of philosophy (Spring 2012 Edition), URL = http://plato.stanford.edu/archives/spr2012/entries/biology-experiment/
Williamson, J. (2010). Probabilistic theories. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 185–212). Oxford: Oxford University Press.
Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind-body problem. In G. G. Globus (Ed.), *Consciousness and the brain* (pp. 205–267). New York/London: Plenum Press.
Wimsatt, W. C. (1994). The ontology of complex systems: Levels, perspectives, and causal thickets. *Canadian Journal of Philosophy, 20*, 207–274.
Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge: Harvard University Press.

Chapter 4
Semiotic Explanation in the Biological Sciences

Ulrich Krohs

Abstract Many biological explanations are given in terms of transduced signals and of stored and transferred information. In the following, I call such information-theoretical explanations "semiotic explanations." Semiotic explanation was hardly ever discussed as a distinct type of explanation. Instead, philosophers looked at information transfer as a somewhat unusual subject of mechanistic explanation and consequently attempted to frame biological information as being observable within physicochemical mechanisms. However, information-theoretical terms never occur in isolation or as a plug-in in mechanistic models but always in the context of information-theoretical models like the semiotic model of protein biosynthesis. This chapter proposes that "information" enters the game as a theoretical term of semiotic models rather than as an observable and that semiotic models have explanatory value by explaining molecular mechanisms in functional rather than in mechanistic terms.

Keywords Biological information • Conserved quantity • Model structure • Nonconservative model • Signal

4.1 Introduction

Biology uses several different kinds of explanation. Among those are causal-mechanistic, constitutive, evolutionary, and deductive-nomological explanations, all of which are well studied in philosophy of science. Giving a causal-mechanistic account is the right way to explain glycolysis or fatty acid synthesis. Constitutive accounts are used in explaining the locomotion system of vertebrates as being made up of bones, muscles, and tendons or in explaining

U. Krohs (✉)
Westfälische Wilhelms Universität Münster, Philosophisches Seminar, Domplatz 6, 48143 Münster, Germany
e-mail: ulrich.krohs@uni-muenster.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_4,

cell respiration as being constituted by the respiratory chain, the NADPH/$NADP^+$ system, the TCA cycle, etc. To explain the presence of particular organismic traits in an organism, evolutionary explanations, which refer to an iterated sequence of variation and selection events, seem to be the adequate kind of explanation to give. Deductive-nomological explanations, finally, though they might be less often applied in biology than in physics, are used whenever a phenomenon is found to be governed by a general law.

Some other types of biological explanation, however, are less well understood and raise severe philosophical concerns. Those are functional explanation, which is regarded as teleology laden and was discussed continuously for half a century in philosophy of science (and by Kant anyway), and explanation in terms of transduced signals and of stored and transduced information. This chapter concentrates on the latter.[1] In the following, I shall call such information-theoretical models "semiotic explanations."

A common account of protein biosynthesis may serve as an example for a semiotic explanation. It represents one of two different ways in which molecular biologists describe the DNA-dependent biosynthesis of nucleic acids and of proteins. This first account explains protein biosynthesis in terms of information transfer and decoding, where the protein sequence is regarded as being coded in the base sequence of DNA. Protein biosynthesis is, thus, explained as a sign process or semiotic process. There is of course also an explanation of another kind that explains protein biosynthesis. It is given in terms of the structures of the molecules involved, the chemical reactions the molecules undergo, and the kinetics and thermodynamics of reactions and biosynthetic pathways. Both models explain the very same process but frame it differently. The first model is a *semiotic model* that gives a semiotic explanation in the sense introduced above; the second one is a *physicalistic model* that explains the same process of protein biosynthesis on the basis of the biochemical processes involved, without referring to any coding function of the involved biochemical components.

While the physicalistic model is generally accepted as giving a proper scientific explanation, the intriguing semiotic model is often challenged because it applies seemingly intentionalist concepts in the non-intentional realm of molecules. It refers to information coded in the DNA and describes the different ways in which information is processed within the cell. It states that information is being copied when a structurally identical molecule of DNA is synthesized, that it is transcribed to RNA, that RNA may be further processed, and that the information of some particular kind of RNA is translated into the sequence of a protein. The whole model is based on semiotic – or sign-theoretical – terminology, using not only the terms "information," "coding," "copying," "transcribing," and "translating" but also "proofreading," "correcting," "recognizing," and many other terms from the field of text processing (see, e.g., Alberts et al. 2002). A vivid discussion was going on among philosophers of biology about whether the term "information" is

[1] In Krohs (2009a, 2011), I deal with the first kind of explanation.

used merely metaphorically in this context (Kay 2000; Griffiths 2001), whether it should be regarded as completely discredited (Sarkar 1998, 2005; Moss 2003), or whether the concept can be naturalized – and if so, in which way this might be done (Sterelny et al. 1996; Godfrey-Smith 1999, 2000; Maynard Smith 2000; Griffiths 2001; Jablonka 2002; Stegmann 2005).[2]

The philosophically less problematic physicalistic model is rather detailed and can be sketched here only superficially. The following account shall merely give an idea of the way this model refers to the processes in question: The structure of the DNA is a sequence of the four bases thymine, adenine, guanine, and cytosine; the molecule is replicated by polymerization of deoxyribonucleotides, the process being catalyzed by the DNA-dependent DNA polymerase and by a strand of DNA; this reaction is thermodynamically driven by the hydrolysis of a pyrophosphate bond in the nucleotides. The description will, of course, add more steps and more details. In the analogous case of RNA biosynthesis, the DNA-dependent RNA polymerase is involved as a catalyst instead, and ribonucleotides are the reactants instead of deoxyribonucleotides. The model also includes the kinetic data of the reactions (Alberts et al. 2002). Overall, the model describes the mechanism of DNA replication, of RNA biosynthesis, and of protein biosynthesis in terms of the components involved and of their interactions. It can therefore be regarded as a mechanistic explanation in the sense of Machamer et al. (2000), Craver (2001), and Bechtel and Abrahamsen (2005).[3] Protein biosynthesis was even made a paradigm case of biological mechanistic explanation (Darden and Craver 2002).

The semiotic and the physicalistic model are, of course, related to each other. Biologists and many philosophers therefore claim that the semiotic model is only a shorthand version of the physicalistic model to which it may be reduced. However, the reducibility claim runs into problems because information is being regarded as multiply realizable. That the very same piece of information may occur in different realizations during processes of information transduction forms a major obstacle to reduction, because identity through different realizations cannot be captured in physicochemical terms, which refer to the realizations only. I shall therefore treat semiotic and physicalistic models separate and inquire into their respective explanatory values.

There is little doubt about the explanatory value of the physicalistic model: it explains the physicochemical processes going on in protein biosynthesis, i.e., it states the mechanism of protein biosynthesis. The case of the semiotic model is not so clear and demands further philosophical analysis. As will become clear in Sect. 4.4, I do not attribute the explanatory success of semiotic models to the concept of information, though it is obviously crucial to these models. No concept has explanatory power in itself. The basic unit of scientific explanation, as Morgan and Morrison (1999) and Giere (2004) plausibly argue, is the model

[2]For a detailed outline of the debate, see Godfrey-Smith (2007).

[3]Glennan's (1996, 2002) approach is similar, except not counting the interactions among the constituents of a mechanism.

rather than a concept or an isolated general statement that makes use of the concept. Consequently, in trying to understand the explanatory role of semiotic models, I do not start from the very concept of information but from the semiotic model as a whole.

In the following, I first introduce a general distinction between two different kinds of models (Sect. 4.2). Using this distinction as a tool for discerning the epistemic virtues of different models, I then discuss the question of whether or not the semiotic model may be reduced to the physicalistic model (Sect. 4.3). Next, the epistemic role of the semiotic model is discussed (Sect. 4.4). Finally (Sect. 4.5), I propose an altered view on the very concept of genetic information.

4.2 Conservative and Nonconservative Models

Models making use of semiotic terms are in fact of a special kind, different from the kind of models used in physics – and different as well from the physicalistic models that are used in biology. In order to conceptualize this difference, I am introducing a distinction between *conservative* and *nonconservative* models. The distinction is such that it singles out physicalistic models as one of the two kinds. It will be shown that semiotic models belong to the other kind.

In order to find a criterion that singles out physicalistic models from other models in biology, we must use physics as our reference. However, any criterion that is supposedly valid may be falsified by the further development of science. There is neither a stable content of physical theories through the centuries nor a stable language of physics (e.g., Hempel 1980). Therefore, we should not look for an a priori valid criterion but for a demarcation criterion that holds with respect to present-day physics. Causality might count as the first candidate for such a criterion. However, although it may be regarded as one of the central notions of physics, there are also noncausal processes or at least processes that cannot be described as causal ones, such as radioactive decay. Causality also fails to hold in the realm of (relativistic) quantum mechanics. So it does not seem to demarcate physical theories properly. Instead, nowadays the minimal requirement for any physical theory or model seems to be that certain variables obey conservation laws: the laws of the conservation of energy (including mass), of net charge, of momentum, and of angular momentum. This holds for the whole range of accepted physical theories, from the classical harmonic oscillator to quantum electrodynamics (Tipler and Mosca 2007).[4] It holds also for the theory of dissipative structures. A dissipative system loses energy through time, but the energy is not annihilated. Any adequate physical model of such a system must postulate a reservoir outside the considered system that takes up dissipated energy. Constancy of energy of the higher system,

[4]We must abstain here from phenomena such as symmetry breaking at the level of elementary particles that are not yet understood satisfactorily.

then, which is made up of the system under investigation plus its environment, is presupposed. This is indeed generalizable: each model in any field of contemporary physics must observe the conservation laws.

Conservativity therefore may be used as a criterion to demarcate physicalistic models, i.e., models of the physical and physicochemical perspective on a phenomenon, in present-day science.[5] Physicalistic models are conservative models. An example is the physicalistic model of cell biology, which describes the reaction pathway from DNA to protein by reference to molecules, reaction kinetics, binding energies, etc. Though it may not usually be spelled out fully in terms of energy conservation but is given as a partial model only, scientific research aims at describing every single step in accordance to the requirements of physical theories, in particular to the conservation of masses and energies (cf. the references given by Alberts et al. 2002 and by Darden and Craver 2002). A premise of stating the model is that the full model meets the requirements of the conservation laws, so that all calculations performed on the basis of this model rely on those laws.

Models which focus on quantities that are not derivable from conserved quantities I call nonconservative models. The nonconserved variables of such models do not represent physical quantities. Examples of nonconservative quantities are cellular signals that are related to hormone action or to external stimuli. A signal in the sense of biological information transduction can simply disappear, without being transformed into anything else. There is no law of signal conservation, nor can signals be deduced from conserved quantities – the same amount of energy and the same configuration of masses may or may not be a signal, and in case of being a signal, it may signal completely different things.

Nonconservative quantities can also be found in the realm of technology (Krohs 2009b), where truth-values in models of logic circuits and other symbolic variables[6] may serve as examples. Besides semiotic or symbolic variables, other functions and functional variables are nonconservative as well. This, again, holds for the biological as well as for the technical realm.[7]

[5]Dowe (1992) and Salmon (1994) correctly identified conservation laws as being at the core of modern physics. The link that these authors draw between conservativity and causality, however, can hardly be justified. In contrast to their proposal, conservativity may neither count as a necessary, nor as a sufficient condition for causality: neither is each conservative process causal (e.g., radioactive decay, tunneling, quantum transitions), nor are all causal processes conservative (e.g., semiotic processes; see Sect. 4.3).

[6]Herbert Simon calls any technical information processing system a physical symbol system (Simon 1996, p. 21, pp. 187–188).

[7]Here the concept of function is taken in the sense of a causal role function (Cummins 1975), which nevertheless allows for judgment about malfunction. To allow for this normativity of the concept, a modification needs to be introduced into Cummins's account, e.g., by reference to fixed types of function bearers (Krohs 2009a, 2011).

4.3 Semiotic Models as Nonconservative Models and the Question of Reduction

To further work out the difference between semiotic and physicalistic models, the distinction between information and its carrier is crucial. The carrier of information may be an electric potential or electric current, be it in a computer or in a nerve cell; it may be the ink on a piece of paper or compressional waves in the air; or it may be, according to semiotic models of molecular biology, the structure of a nucleic acid molecule. The carriers are physical entities, and all transformations they undergo obey the conservation laws. Consequently, they and their relations and interactions may be described by an appropriate physicalistic model.

The case is different with the information that they carry. As already mentioned, information may disappear without residue. It may also appear without being governed or restricted by conservation conditions. For the first case, consider a technical device for information processing like a logical gate, say, the NOR gate. Its output is "one" if and only if both inputs are "zero." In all three other cases of defined input, the output will be "zero." If the output is stored and the gate is then switched off, the stored information is "one" or "zero." In case it is "zero," the information about the input channels, namely, which one of them was "one," is lost. The lost information is neither transformed nor dissipated; it is annihilated. Information is not conserved, and from two bits of information, only one is left. A similar case of information loss can be found in any degenerate code, like the DNA code, where in most cases more than one base triplet codes for an amino acid. The informational content of the third base is lost during translation into a protein sequence. Nonconservativity of information is, however, not restricted to cases of redundancy. Imagine a breakdown of a computer occurring before a freshly composed text or the data obtained in a series of measurements were saved. The energy balance of the breakdown may depend on the number of bits stored in the computer, i.e., on the size of its memory. But it does not depend, at least not in a systematic manner, on the symbolic content of the memory. The information is not transformed but lost when the system breaks down. Similarly, on the hydrolysis of a piece of DNA, genetic information is lost, although binding energy and molecular material of the carrier of information are conserved.

Nonconservativity holds as well for information increase. New information may be generated when a random sequence of DNA is synthesized, when a point mutation gives rise to an altered sequence, or when the insertion of a base or of some pseudogene occurs in a living cell. Thus, on the transformation of some molecule, we see an increase, alteration, or decrease of information. (This, of course, is the image drawn by the semiotic model, not by a physicalistic account, which does not support talking about information.) Information does not obey

conservation laws; only the underlying molecular processes do. Semiotic models are nonconservative ones.[8]

These considerations entail several reasons why semiotic models cannot be reduced to physicalistic ones.[9] First, the semiotic model allows to discern between informational processes running properly and processes going wrong. It allows identifying several kinds of copying errors, correction functions, etc. This means that the semiotic model is a functional model that discerns function from malfunction; it is a normative model. Function in the normative sense is and must be absent from physicalistic models Therefore, they cannot fully account for what the semiotic model explains. Next, also concerning the aspect of functionality, physicalistic models can neither account for nor explain multiple realizability of semiotic entities, i.e., for the fact that the same piece of information can have various carriers. Third, it is all but clear how the identity of a piece of information or of a signal through various realizations could be described by merely referring to its various heterogeneous carriers. And finally, though perhaps only of pragmatic relevance, the task of reduction would be much larger than envisaged by accounts concentrating on the concept of information itself: a whole set of semiotic concepts is involved, many if not all of them referring in their original context to intentional text processing. All of those needed to be reduced to physicalistic descriptions.

So the semiotic model is not only an incomplete version of a physicalistic one. It makes use of classifications, such as being a signal or coding for some component, that are alien to physicalistic models. There are two mutually nonexclusive ways to explain protein biosynthesis: by a conservative, physicalistic model and by a nonconservative, semiotic one. Neither of them alone covers all that can be known about the process that is to be explained.

4.4 The Epistemic Role of Semiotic Models in Biology

As already mentioned, the physicalistic model of protein biosynthesis explains what is going on physicochemically during this process. Its epistemic role is to give a mechanistic account of the process. A mechanism in this sense consists of the set of the entities involved and the relations that hold between them. But being conceived as a mechanism, it is also conceived as the mechanism *of something*, namely, as an instantiation of a cellular capacity which is individuated functionally and which consists of a set of functional roles. Darden and Craver (2002) describe

[8] A further question is whether nonconservativity of functional models holds in general. This seems to be the case (Krohs 2004). The function of a screw (or of any other mechanical device) of being a stop for a lever can simply be lost under certain circumstances, e.g., if the lever is bent. There is no necessity of the function being transformed into anything else according to any conservation law.

[9] Only theory reduction is at stake here. Ontological reducibility may be presupposed, be the semiotic model reducible to the physicalistic one or not.

many of the role functions in protein biosynthesis in terms of information flow. So the mechanistic explanation is also related to the semiotic model, which is given exactly in these functional terms. Inquiring the epistemic role of the semiotic model, one must consider, then, how it relates to what Darden and Craver call a mechanistic schema. The semiotic model is clearly not identical with the schema, since the latter does not refer to actual components of a mechanism but provides placeholders instead: "Mechanism schemata are abstract frameworks for mechanisms. They contain place-holders for the components of the mechanism (both entities and activities) and indicate, with variable degrees of abstraction, how the components are organized" (Darden and Craver 2002, p. 4). The semiotic model, in contrast, does refer to molecular components of the cell, namely to the same ones as the physicalistic account does, so the places are already filled. In contrast to a schema, the semiotic model has an ontology, even more or less the same ontology as the physicalistic one, with the notable exception of semiotic terms. The latter do not have correlates in the physicalistic model. In particular, they are not placeholders for physical entities. So the semiotic model is itself an instantiation of a schema rather than an un-instantiated schema. Darden and Craver conceive the schema as the schema of the mechanism, which is described by the physicalistic model (in my terminology). It now turns out to be the schema of both the physicalistic and the semiotic model.[10]

If the semiotic model is not reducible to the physicalistic one and a fortiori not simply a shorthand or laboratory slang version of the latter, it remains to clarify what precisely the epistemic role of the semiotic model is. The answer is to be found in the biologists' aim to explain both, the physicochemical processes of living entities, and their functional organization. The question about functionality, while absent from physics and chemistry, forms the very basis of physiology. A functional model, e.g., the model of the blood circuit as a distributor system, the model of the liver as a detoxifier, or the model of a mitochondrion as the power station of the cell, helps to understand a biological entity as an organized system. It embeds a particular capacity into the hierarchical structure of capacities of the organism. Such a particular capacity may contribute to the overall capacities (i.e., function) or fail to contribute properly (i.e., malfunction). The semiotic model that serves as an example throughout this chapter places the pathways involved in information processing into a hierarchy of contributions to growth, self-maintenance, and proliferation of the cell; to the regulation of cell metabolism and integration of the organism; and to degradation of cellular components and to cell aging. It does so by simplifying the physicalistic description and at the same time introducing functional

[10]Darden and Craver (2002, p. 5) ascribe work on information flow to molecular biologists and work on the flow of matter and energy to biochemists. While this might be considered a somewhat artificial attribution of different research topics to disciplines, it clearly emphasizes that physicochemical and semiotic analyses are categorically different and thus should indeed give rise to models of different kind.

entities which are absent from physicalistic models: firstly, diverse processes are unified in regarding physicochemically different steps as the processing of one identical signal or piece of information; secondly, the physical requirements of particular realizations of this organization are disregarded. Most, if not all, functions could be realized in many different ways (Carrier 2000), and each realization would underlie different physical constraints. So building nonconservative models is not sloppiness in constructing a model in order to get rid of too much detail. It is rather the prerequisite for an integrated view on biological organization. The semiotic model, while being silent on the physicochemical mechanism, gives quite accurate an explanation of the functional structure of protein biosynthesis and allows for precise and successful predictions of the behavior of the system's processing of different pieces of DNA. It also allows for judging whether or not, in a particular case, the processes are running properly.

The problem seems to remain that information talk, known from settings with intentional senders and receivers, seems to be inadequate when applied on the molecular level. What, then, justifies the use of semiotic models and their transfer into a new area of application? When we are asking for the justification of the use of a model, two candidates are available for what may be regarded as explanatory: the structure of a model or its conceptual content. With regard to the content, the semiotic model is all about information and its processing. This content does not seem to be justified by the phenomenon to be described, as the critique mentioned in Sect. 4.1 has shown. In particular we do not want to assume or presuppose intentionality on the molecular level, which makes explanation of protein biosynthesis in terms of semiotic processes somewhat dubious. I therefore propose to search for explanatory power in the structure rather than in the conceptual content of semiotic models. Structurally, the semiotic model appears to fit well to the phenomena as described by the physicalistic model. This is not affected by the somehow odd reference of the model to molecular information. In particular, features like the degeneracy of the code – i.e., the finding that different base triplets give rise to the incorporation of the same amino acid into a protein – or the different steps of transcription are captured in a straightforward way by an account that allows for multiple realizability. In the case of signal transduction, one and the same signal, i.e., a nonconserved quantity, is described as being a conformation change of a small molecule, of an enzyme, of an ion channel, etc. This is reflected by the structure of the relevant semiotic model. Its structure, consequently, must be regarded as carrying or contributing to the explanatory power of the semiotic model. The features of the system that are captured by this structure are, in contrast, not grasped by the physicalistic model – neither by its structure nor by its conceptual content. This is why biologists cannot refrain from using nonconservative, semiotic models.

Conceptual content and structure can both matter for the explanatory power of a model. This is usually taken for granted for the content side and spelled out for

the structural side by structural realism.[11] Since, in the present case, we find that the conceptual content can hardly explain the epistemic value which the semiotic model obviously has, structure alone seems to be responsible for the epistemic success of semiotic models. We may still blame the model for its misleading content, but as long as the structure of the model is required for epistemic reasons and cannot be had without this very content, the somewhat dubious content alone does not seem to be a sufficient reason for eliminating the model. So, if structural explanation is a part of scientific explanation, the semiotic model has its merits exactly in this realm. Since the conceptual content seems to be ontologically inadequate, a nonconservative model does not, or should not, even aim at a realistic description of the inventory of the physical world – otherwise, it needed to be conservative.[12] Models of both kinds have a different status. Consequently, it would be consistent to allow for metaphorical content of nonconservative models and nevertheless demand ontological adequacy for the conceptual content of conservative models.[13] (Theoretical terms, nevertheless, are notoriously posing a problem in this respect.)

4.5 How to Deal with the Concept of Genetic Information

My account of the explanatory power of semiotic models does not interpret the conceptual content of these models and therefore does not explicate the concept of information. It remains puzzling why semiotic terminology seems to be crucial for the models in question. For now, I can just present a guess why biologists describe the functional organization of protein biosynthesis in semiotic terms. The guess is that simply no other functional model could yet be found that has a comparable structure, i.e., that describes various sequential conformation changes as realizations of the same function or, more generally, of the same nonconservative quantity. No such model was found that uses a terminology that avoids semiotic concepts and nevertheless manages to hook up with our understanding of some processes we are familiar with so that it can be integrated into our system of knowledge. The use of the concept of information seems to be the price biologists have to pay for gaining a structurally adequate nonconservative model of the functionality of protein biosynthesis.

[11] Nevertheless, one needs not subscribe to structural realism, neither in its epistemic (Maxwell 1970; Worrall 1989, 1994) nor in its ontic variant (Ladyman 1998; French and Ladyman 2003), to accept the explanatory power of the structure of a model.

[12] In so far, the realist interpretation of semiotic terms in molecular biology by some biosemioticians is misguided.

[13] Since the physicalistic model carries the realist burden, the correlated semiotic model even must not be interpreted in a realist way. I see no reason for going as far as postulating an informational ontology (Floridi 2008, 2009).

Does this mean that the concept of genetic information is used metaphorically? As an isolated concept, it is hard to see how it could work as a metaphor at all. What should be the content carried by the metaphor, in a field, where intentionality and interpretation are absent? However, the model makes use of a whole set of interrelated concepts as listed above. If anything, the conceptual set as a whole should be regarded as the metaphor, transporting the structure or the model rather than the semiotic content into the field of molecular biology.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular biology of the cell* (4th ed.). New York: Garland.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 33*(2), 421–441.

Carrier, M. (2000). Multiplicity and heterogeneity: On the relations between functions and their realizations. *Studies in History and Philosophy of Biological and Biomedical Sciences, Part C, 31*(1), 179–191.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science, 68*(1), 53–74.

Cummins, R. (1975). Functional analysis. *The Journal of Philosophy, 72*, 741–765.

Darden, L., & Craver, C. F. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Biological and Biomedical Sciences, Part C, 33*(1), 1–28.

Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science, 59*(2), 195–216.

Floridi, L. (2008). A defence of informational structural realism. *Synthese, 161*(2), 219–253.

Floridi, L. (2009). Against digital ontology. *Synthese, 168*(1), 151–178.

French, S., & Ladyman, J. (2003). Remodelling structural realism: Quantum mechanics and the metaphysics of structure. *Synthese, 136*(1), 31–56.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*(5), 742–752.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*(1), 49–71.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(3), 342–353.

Godfrey-Smith, P. (1999). Genes and codes: Lessons from the philosophy of mind? In V. G. Hardcastle (Ed.), *Where biology meets psychology: Philosophical essays* (pp. 305–331). Cambridge: MIT Press.

Godfrey-Smith, P. (2000). On the theoretical role of 'genetic coding'. *Philosophy of Science, 67*(1), 26–44.

Godfrey-Smith, P. (2007). Information in biology. In D. L. Hull & M. Ruse (Eds.), *The Cambridge companion to the philosophy of biology* (pp. 103–119). Cambridge: Cambridge University Press.

Griffiths, P. E. (2001). Genetic information: A metaphor in search of a theory. *Philosophy of Science, 68*(3), 394–412.

Hempel, C. G. (1980). Comments on Goodman's *ways of worldmaking*. *Synthese, 45*(2), 193–199.

Jablonka, E. (2002). Information: Its interpretation, its inheritance, and its sharing. *Philosophy of Science, 69*(4), 578–605.

Kay, L. E. (2000). *Who wrote the book of life? A history of the genetic code*. Stanford: Stanford University Press.

Krohs, U. (2004). *Eine Theorie biologischer Theorien*. Berlin: Springer.

Krohs, U. (2009a). Functions as based on a concept of general design. *Synthese, 166*(1), 69–89.
Krohs, U. (2009b). Structure and coherence of two-model-descriptions of technical artefacts. *Techné: Research in Philosophy and Technology, 13*(2), 150–161.
Krohs, U. (2011). Functions and fixed types: Biological functions in the post-adaptationist era. *Applied Ontology, 6*(2), 125–139.
Ladyman, J. (1998). What is structural realism. *Studies in History and Philosophy of Science, Part A, 29*(3), 409–424.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Maxwell, G. (1970). Structural realism and the meaning of theoretical terms. In M. Radner & S. Winokur (Eds.), *Analyses of theories and methods of physics and psychology* (Minnesota studies in the philosophy of science, Vol. 4, pp. 181–192). Minneapolis: University of Minnesota Press.
Maynard Smith, J. (2000). The concept of information in biology. *Philosophy of Science, 67*(2), 177–194.
Morgan, M. S., & Morrison, M. (Eds.). (1999). *Models as mediators*. Cambridge: Cambridge University Press.
Moss, L. (2003). *What genes can't do*. Cambridge: MIT Press.
Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science, 61*(2), 297–312.
Sarkar, S. (1998). *Genetics and reductionism*. Cambridge: Cambridge University Press.
Sarkar, S. (2005). *Molecular models of life: Philosophical papers on molecular biology*. Cambridge: MIT Press.
Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge: MIT Press.
Stegmann, U. (2005). Genetic information as instructional content. *Philosophy of Science, 72*(3), 425–443.
Sterelny, K., Smith, K., & Dickison, M. (1996). The extended replicator. *Biology and Philosophy, 11*(3), 377–403.
Tipler, P. A., & Mosca, G. (2007). *Physics for scientists and engineers, extended version* (6th ed.). New York: Freeman.
Worrall, J. (1989). Structural realism: The best of both worlds? *Dialectica, 43*(1–2), 99–124.
Worrall, J. (1994). How to remain (reasonably) optimistic: Scientific realism and the 'Luminiferous Ether'. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1*, 334–342.

Chapter 5
Mechanisms, Patho-Mechanisms, and the Explanation of Disease in Scientifically Based Clinical Medicine

G. Müller-Strahl

Abstract In scientifically based medicine, explanations of normal and deviating organismic properties or events commonly have recourse to the notions of normo- and patho-mechanisms. I will argue – contrary to the shortcut view of most adherents of mechanistic philosophy – that there is a necessarily long but feasible passageway from normo- to patho-mechanisms and will plead for objectivism of the concept of individual diseases on the basis of the concept of a complex mechanistic base supplemented with a general function-analytical account of explanation. This study also considers some of the most prominent ontologies of disease entities, i.e. disease as process or as incapacity. Further, objective criteria are presented which delimit the range of items belonging to a base. These are preparatory steps to carve out the concepts of directionality or connectivity of mechanistic bases which turn out to be the most proximate notions of order in systems combining inciting and inhibitory causal relations. However, a knowledge of the laws of living matter is declined. These findings suggest that mechanistic bases are akin to causal bases and that explanation in medicine is supported by these objectifying concepts. Finally, by introducing a notion of difference among various organismic states, as long as they refer to the same mechanistic base, this contrastive component imbues the underlying mechanistic framework with the distinguishing notions of normo- and patho-mechanisms.

Keywords Mechanistic Explanation • Disease • Negative Causation • Neuron Diagram • Complexity

G. Müller-Strahl (✉)
Institut für Ethik, Geschichte und Theorie der Medizin, Westfälische Wilhelms-Universität Münster, Von-Esmarch-Straße 62, 48149 Münster
e-mail: gerhard.mueller-strahl@uni-muenster.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_5,

5.1 Preliminaries

This study intends to establish an explanatory account of diseases by accounting for both theory and practice of scientifically based clinical medicine. *Explanatory accounts* are distinct from accounts which elaborate classificatory criteria that apply to the entire collective of diseases. In the majority of cases, this *classificatory* task is achieved by creating an overarching disease concept which may rely on objective descriptions (Boorse 1997), or confess to be predominantly normative (Nordenfelt 2007), or incorporate both of these founding principles (Wakefield 1992). That accounts of the classificatory kind outnumber the explanatory ones may be due to two tacitly accepted views in the philosophy of medicine: first, it is believed that a process theory of disease is a reliable point of reference; the problems, however, which arise for an account of explanation when holding such a position, and the alternatives, which may support causal explanations, are rarely addressed. A singular and straightforward examination of these aspects has been given by Whitbeck (1977). Her results will be outlined in Sect. 5.3. Second, the philosophy of medicine did not get rid of the influence experienced by the failure of some early attempts to install deductive-nomological explanatory accounts (Doroszewski 1980; Korab-Laskowska 1980; Ren-Zong 1989; Sadegh-Zadeh 2011), and, therefore again, invested less effort in finding alternative explanatory strategies. Maull (1981) has been among the first to state emphatically that to achieve explanations in the medical context it may be advisable to refrain from deductive theory reduction – not to exclude, however, the possibility to elaborate a well-conceived account of explanatory reduction; since her proposed direction of methodology has not found many followers, it is worthwhile to look out for adequate concepts of explanatory reduction in the domain of medicine. A clear-cut reductive explanation of this type has been proposed by Lange (2007). Since Lange's account both considerably draws from Whitbeck's, and is less known among philosophers of the life sciences, its noteworthy reflections and stimulating theses deserve a thorough reconstruction in order to shape and accentuate more rigorously its major lines of thought. This task will be redeemed in Sect. 5.5.

The slogan "no laws in medicine" not only imparted immunity against deductive explanations but also led to an obvious discrepancy within the philosophy of life sciences: in the context of the philosophy of biology, the problem of epistemic inter-theory reduction according to the special Nagelian version of the deductive-nomological model has received comparatively more attention – e.g., among those adhering or opposing to Schaffner (1993) – than in the philosophy of medicine. This may be due to the fact that the existence of laws is under debate for the realm of biological, hardly, however, medical sciences. Similarly, the recent renaissance of the philosophy of mechanisms has been assimilated successfully to the life sciences (Craver 2007; Bechtel 2011), less, however, for explanatory procedures in medicine (Thagard 1998; Campaner 2011b). Therefore, Sect. 5.6 will demonstrate that it is

worthwhile to make use of a concept of a causal mechanistic base of organismic events; it takes into consideration concepts and developments which have been successively revealed in the medical sciences over the last decades; in fact, such an approach will turn out to be promising in order to understand the fundamentals of medical explanation in the context of disease terms; most importantly, the rationale of diagnosis is at the core of clinical practice and should be elucidated by the concept of a causal mechanistic base; to this intent, first, the roles that explanatory entities may play in the course of diagnostic reasoning will be outlined in Sect. 5.2 right below. Those explanatory entities within a causal mechanistic base that figure in diagnostic inferences will be denominated disease entities; their objective meaning and the differences to other concepts of disease entities are to be clarified in Sect. 5.6.

Thus, three major impulses support this analysis: towards *objectivism* of individual disease concepts by referring to a *mechanistic base* for a general account of *explanation* in scientific clinical medicine. The first impulse is shared with naturalistic positions, e.g., that of Boorse (1975, 1997); this latter position, however, does not support an individuation of diseases; on the organismic level, it solely refers to functions, thus neglecting, on the one hand, the whole variety of symptoms and findings assembled in the clinical picture and natural history of a disease, and on the other hand, the whole spectrum of background conditions that furnish the organism with the functions it is supposed to have. The second impulse of this analysis is shared with a few mechanistic accounts of disease, which, however, skip some important aspects of the explanatory practice in medicine, e.g., the relation of mechanisms to bodily functions or the question what a disease *is* (Thagard 2000; Severinsen 2001; Nervi 2010; Campaner 2011a). Because of the latter omission, they quite frequently just presuppose an *ex ante* distinction of normal and pathological mechanisms and are informed by our acquired and solidified intuitions concerning these notions.

The third impulse which has relatively rarely been employed to analyze the medical disease concept will be shared with the two positions having been announced already and to be presented below in the Sects. 5.3 and 5.5. As has been indicated before, a careful evaluation of their peculiarities will render it worthwhile to elaborate an alternative ontology which invokes the concept of a mechanistic explanatory base; such a causal base turns out to be the source of explanatory power in the context of scientific clinical medicine (Sect. 5.6). The concrete point of reference for this analysis is the prototypic disease phenylketonuria, which is therefore presented in some detail in Sect. 5.4. Since the concept of a mechanistic explanatory base elevates the concept of mechanisms by introducing the notion of certain types of relations between mechanisms, this study finally examines the (objective) normative criteria which codetermine the specific structure of these complexe objects (Sect. 5.6.4). Thus, it can be shown that mechanistic bases are eligible to constitute the disease entities sought for.

5.2 Medical Clinical Phenomenology of the Disease Concept

Diagnosis is a key step in the process of clinical judgment and decision making; its eminent role in the highly complex context of clinical practice is reflected by the fact that diagnosis precedes most instances of treating a patient, and treatment, implying a clear benefit for the patient, is one of the major goals of medicine (Munson and Roth 1994). Among the constitutive elements supporting a diagnosis, the manifestation process of a disease appears to be the most prominent one, but, analytically, it points to a tacitly accepted distinction between an overt phenomenological part and a hidden disease entity which subsists the overt symptoms. Certain clinical expressions have been customized according to this distinction, such as the *full* clinical picture, the latency phase, or the subclinical stage of a disease.

Signs, symptoms, laboratory and other clinical findings met in a patient a at a certain time t can be understood to form a subset S_a of all the characteristic features revealed by the overall manifestation process. To find an apt diagnosis in relation to this simplified clinical situation would mean to make an inference from S_a to a nosological unit N; N is a classificatory term for a disease type and provides the clinician with a set of features S_N according to the general statement $\forall x : (Nx \rightarrow S_N x)$. This inference is supported by a similarity relation $\approx$ between the relevant sets originating from the concrete clinical picture and from the standards relevant for detecting a diseased state (generally referred to as nosology): $S_a \approx S_N$. Taken together, the search for a unique diagnostic entity N according to a given set of symptoms S_a includes the task of solving an inverse problem of the form:[1]

$$\mathrm{P1} : \forall x(Nx \rightarrow S_N x) \wedge \mathrm{P2} : \exists a : S_a a \wedge (S_a \approx S_N) \,//\, \mathrm{C} : Na.$$

The simplified logical scheme of clinical diagnostic reasoning is useful to indicate the obstacles which hinder the formation of a diagnosis as long as it depends on eliminative inductive inference: the generation of hypotheses which allow for the respective abductive inferences is followed by deductive hypothesis testing; after this cycle of abduction, deduction and testing, rather few hypotheses

[1]For literature relating to some aspects of the problem of diagnosis consult Barosi et al. (1993); Bunge (2003); Forber (2011); Rizzi (1994); Sober (1979). Obviously, there are two groups of clinical cases where diagnostic judgments present themselves to be less subtle because there is no need to accentuate the difference of manifest and concealed aspects of an organism whose health is endangered: The one group comprises acute cases of emergencies (which demand direct restitution of vital functions in order to circumvent a life threatening situation) and the other such cases in which the denomination of a disease just appeals to a small set of overt symptoms of limited bodily extent. A short extract of a long list would include: injuries, impairments, variants, malformations, anatomical lesions, inflammations, poisonings, blindness, burns, starvation, drowning, tinnitus, ileus, impingement of an articulation, arthrosis, all kinds of classical triads of symptoms, etc.; multiple sklerosis is another special case (Giovannoni and Ebers 2007).

(ideally, only one) are retained. Since the first obstacle, abductive inference, is already impregnated with uncertainty – the initial hypotheses being obtained by divining – this character cannot be eradicated by the two other steps (Forber 2011). A solution to this loose search for a valuable diagnosis could be to indicate the rational constraints of the aforementioned inferences (Barosi et al. 1993; Sober 1979) or to create an objective foundation of the disease concept. In this study, the last alternative will be pursued, since there are some hints from medical clinical phenomenology which support this route.

5.2.1 *Some Hints for Objectifying Medical Clinical Phenomenology*

If S_a designates the above defined set of manifest features, then it is conceivable that S_a might contain enough information to ground therapy in a number of cases. However, there is certainly more to diagnosis than considering this collection of synchronous symptoms, since the process of diagnosis also aims to include signs which are obtained by elucidating the case history, the *anamnesis* of the individual a. This means that the natural history of a disease conveys diachronic information relevant for the actual state of a and, thereby, helps to identify a disease entity and thus to justify therapy *iff* the prognosis in case of treatment is preferable to one without therapeutic intervention. However, at first sight, this observation does not jeopardize the view that a rigorous unification of synchronous and diachronic data may ground therapeutic efficiency considerably. But then one should also endorse the view that the identification of a diseased state would depend to a great extent on the possibility to ameliorate a state or a condition of an organism by therapeutic interventions. The extension of such a disease concept would, however, not conform with medical clinical practice, since in one sense it is too narrow, in the other it is to broad.[2]

At this stage of the analysis my route departs from this refuge in normativism, and heads for objectivism since clinical phenomenology supports the conjecture that diagnosis is more than to identify a synchronic or diachronic *collection* of signs or symptoms which is denominated by N.[3] Diagnosis does not simply identify sets of

[2]This specific disease concept is too narrow because there are diseases which cannot be treated, and it is too broad because there are ways of effective treatment without necessarily identifying a disease before.

[3]The normativism-naturalism debate for diseases has been appropriately characterized in the following statement:

> Some scholars, *objectivists about disease*, think that there are facts about the human body on which the notion of disease is founded, and that those with a clear grasp of these facts would have no difficulty drawing lines, even in the challenging cases. Their opponents, *constructivists about disease*, maintain that this is an illusion, that the disputed cases reveal how the values of different social groups conflict, rather than exposing any ignorance of

the kind S_a but *above* that a *coordination* among the elements of these sets. This contrast between purely assembled and coordinated signs is a first hint towards the hypothesis that there are entities μ equipped with properties which *explain* some aspects of an organism. It is the intention of the following analysis to show for some prototypical cases which entities μ are good candidates for the explanatory task which is faintly observable in clinical reasoning, how medical explanation of diseases is achieved due to these entities and where their explanatory power originates from. *Prima facie*, there are two approaches which may lend support to the intended demonstration; a more or less indirect one, which tries to give a proof of the existence of such μ by appealing to natural kinds, and a more direct one, which clarifies the fundamental ontology of diseases. If the last approach will turn out to be successful, the question whether, for any individual variable x, the assertion Nx makes reference to a distinct ontological entity μx may be answered positively. Taken together with the preliminaries, disease entities μ may then be used to explain S_N in a manner which is advantageous for diagnostic reasoning according to the schema:

$$\text{P1} : \forall x : (Nx \leftrightarrow \mu x) \wedge \text{P2} : \forall x : (\mu x \leftrightarrow S_N x) \wedge \text{P3} : \exists\, a : S_a a \wedge (S_a \approx S_N) /\!/ \text{C} : Na.$$

5.3 A Process Ontology of the Explanans μ

A process ontology for diseases has been scrutinized by Whitbeck (1977). This position recourses to the following basic statements: generally, it does not appear to be controversial in medical practice that symptoms constituting the clinical picture of a disease and, next, the underlying changes at the microscopic level are treated as separate and coordinated processes. However, there is less conformity concerning the relation between these levels. Again, the dominant view links these layers by a productive account of causation whereby the microscopic pathological events are considered to be causes of the symptoms of the disease. This bipartite view with regard to diseased organisms, however, is not more than an assumption which has to be contradicted; there are no such limits between processes at different levels because mutual interferences by upward and downward causation cannot be eliminated. Therefore, a disease entity is one, and only one, process whose unfortunately very limited information can be obtained by the clinician. Events,

facts, and that agreement is sometimes even produced because of universal acceptance of a system of values [...]. (Kitcher 1997, 208–9)

For an overview of this debate see Murphy (2009).

signs, findings and symptoms are various aspects of this process which are as such of equal value. The entities partaking in a complex disease process are related mutually by effective causation; the entire process may be an effective cause for another later disease process; one stage of a disease may cause a later stage, or the process as a whole might explain effects that appear during the process and remain after the process itself has ended. Further, since the disease process is distinct from a process under healthy conditions, a disease process necessarily has characteristic temporal marks which mark out certain intervals. In consequence, the medical concept of *the* cause of a disease may refer, first, to the transition of the disease-free state of an organism to its diseased state and, second, to the entire process engendered by the initial transition. Such inciting or, respectively, sustaining causes of a disease are the sole candidates to identify a disease in the context of this ontology. The case of sustaining has not been treated by Whitbeck (1977), but the former is explicitly specified by performing a conditional analysis of those factors which elicit a deviating process and by providing some rules which determine the identification of the etiologic agent.

Another consequence of this process account of diseases is a narrowing of the space of explanatory resources; no causative explanation is to be expected by referring μ to a whole of causally connected parts; what remains is the claim that a μ as a process has a classificatory function; it is classification which is a subsidiary vehicle for the explanatory role of μ in clinical practice. Of course, the sense of classification is important in this context; Whitbeck (1977, 620) understands sorting individuals with S_a into one class as grouping them according to criteria which maximize the number of correct inferences that can be obtained from the class the individual belongs to.[4]

An increasing dissatisfaction with the process account of diseases has led more recently to a diametrically opposite conceptualization. Before entering into the evaluation of this incapacity account, a concrete paradigm of a medical diagnostic entity and its explanatory components will be presented in the next section.

5.4 The Individuation of a Disease Category

In this study, phenylketonuria (PKU) will be employed as a prototype of a medical disease entity. Since in the remaining sections of this analysis some statements are exemplified by appealing to features of this disease, it is of interest to give a detailed outline of some physiological mechanisms which are invoked in the context of scientific clinical medicine. The explanations of this section refer to Fig. 5.1.

Phenylalanin (phe) is an essential aminoacid which is sufficiently supplied by the normal dietary protein intake of humans; as aminoacid it can be integrated

[4]Therefore, Lange's reproach that relying on classification for explanation would be a circular construction does not apply (Lange 2007, 274).

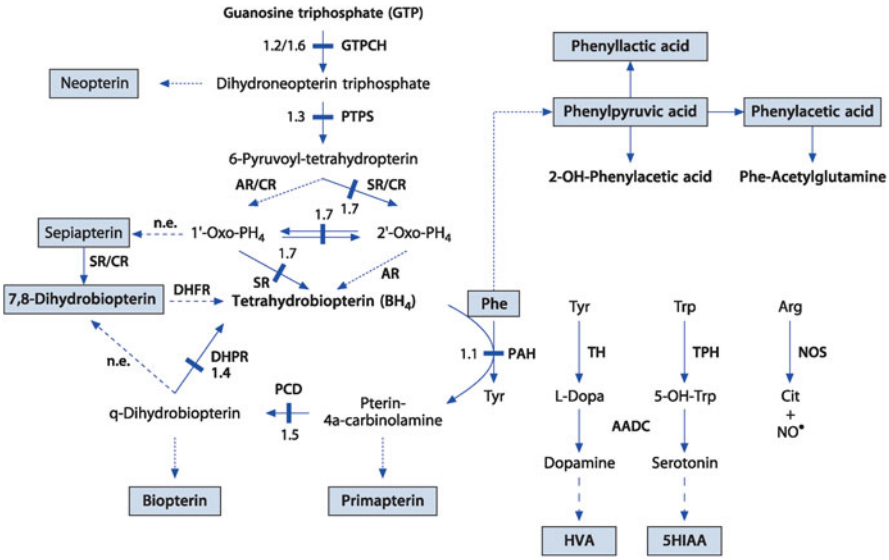


Fig. 5.1 Scheme of some pathways relevant to metabolism of phenylalanin (phe); description in Sect. 5.4; more details are accessible in the paper of Blau et al. (2003) who authorized the reproduction of the original. Enzymes are represented by abbreviations in capital letters close to the shafts of the arrows; n. e. indicates a non-enzymatic (spontaneous) reaction

into certain proteins, or, more peculiarly, it is transformed into tyrosine (tyr) by the iron-dependent enzyme phenylalaninhydroxylase (PAH) which is shown in the center of Fig. 5.1; in order to perform this hydroxylation, PAH requires molecular oxygen and the cofactor tetrahydrobiopterin (BH_4). The enzyme PAH is localized predominantly in the parenchyme of the liver and some other organs, e.g., kidney and pancreas.

In the standard situation, tyr is *not* an essential aminoacid; once produced, it is further distributed over the compartments of the organism by diffusion, membrane bound carriers and convective transport and, e.g., selectively processed within neurons of the *Corpus striatum, Putamen and Nucleus caudatus* of the diencephalon; these cells are equipped with the enzyme tyrosine-3-hydroxylase (TH) which converts tyr into Dopa; Dopa may eventually be transformed into dopamine involving the enzyme AADC; a lack of Dopa establishes symptoms which are categorized as Parkinsonism. Further, dopamine is a precursor of catecholamines in the vegetative nervous system and of melanin, a pigment in the melanocytes of the skin; in the thyroid gland, tyr is a precursor for thyroxine and triiodthyronin.

Renewal of the BH_4-pool is essential for PAH activity and comes from two sources: first, from a recycling pathway shown in the left lower quarter of Fig. 5.1; it involves the enzymes PCD and DHPR; secondly, from a linear pathway starting from the nucleotide GTP which is converted stepwise by the enzymes GTP-cyclohydrolase (GTPCH), PTPS and sepiapterin reductase (SR); SR catalyzes a two-step reaction.

Note that BH_4 is not only a cofactor for PAH – and may thus be responsible for an (extrinsic) deficiency of PAH – but also for TH, TPH, and nitric oxide synthase (NOS); these enzymes are shown in the right lower quarter of Fig. 5.1. Thus, any lack of this cofactor, due to mutations in one of the genes which encode enzymes that contribute to the synthetic or recycling pathway for BH_4, will not only be limited to a decrease of PAH activity accompanied by HPA, but also result in a deficiency of the biogenic amines dopamine, the catecholamines and serotonin. Diseases belonging to this group of BH_4-deficiencies are called *malignant* hyperphenylemias since the evoked type of HPA is difficult to be controlled by treatment. Note, that HPA, due to an intrinsic deficiency of PAH (with the cofactor BH_4 being sufficiently supplied), will similarly result in a depletion of the biogenic amines for the now different reason that phe at high concentrations may act as an inhibitor of TH and TPH (substrate inhibition).

Any deficiency of PAH – either intrinsic or due to inefficient BH_4 metabolism – will therefore have some consequences: Tyrosine will become an essential aminoacid and must be supplied with the diet (what is normally guaranteed). The diminished conversion of phe will lead to an accumulation of phe in the body fluids thus producing the symptom of hyperphenylalaninemia (HPA). Next, the low affinity of some enzymes to phe will be outweighed by its anormally high concentration, so that phe may enter into metabolic pathways which lead to the production of anormal substances; these are shown in the right upper quarter of Fig. 5.1, phenylpyruvic acid being one of them. Further, in the circumstances of HPA, phe will have an inhibitory effect on TH and tryptophan-5-hydroxylase (TPH). Some consequences of a deficiency of TH have already been described. TPH is a key enzyme in the biosynthesis of serotonin from tryptophan (tryp). Serotonin is a neurotransmitter, present in the diencephalon (hypothalamus, nucleus caudatus) and in the pineal gland where it is a precursor of the hormone melatonin. Serotonin is also stored in the enterochromaffine cells of the gut wall.

In summary, there are two distinct groups of HPA: PAH- and BH_4-deficient.[5] The more severe cases of HPA are often identified with PKU, and if they are due to an intrinsic deficiency of PAH, this category of diseases is called *classical* PKU. In a subset of individuals with HPA – mostly moderate or mild HPA – oral supplementation with additional BH_4 can lead to a reduction in blood phe concentration. The mechanism of this BH_4-responsiveness – probably a chaperon-like effect of BH_4 on PAH – is still unclear (Harding and Blau 2010). About 1 to 2 percent of cases of HPA are due to mutations in genes coding for enzymes involved in BH_4 biosynthesis or regeneration. Classic BH_4-deficiencies are characterized

[5] In a strict sense, extrinsic PAH-deficiencies are not only dependent on BH_4, but also on the supply of O_2 and Fe^{2+}; however, an iron deficiency leading to HPA would have to be so severe that any symptoms related to HPA in isolation submerged in the overall clinical picture due to the lack of iron. The same reasoning applies still more obviously to the case of oxygen deficiency – or even to systemic liver diseases which may be very subtle, e.g., minimal hepatic encephalopathy is a disease which symptomatically joins a global insufficiency of the liver (detected by the amount of ammonium in the blood) to a quite specific kind of cognitive impairment.

by HPA and deficiencies of certain neurotransmitters (dopamin, norepinephrine, epinephrine; serotonin; NO). Thus, it can be anticipated that the classification of the *metabolom* HPA is still under debate; one example which is taken from Blau et al. (2010) refers to the biochemical phenotype: the normal range of blood phe concentrations is taken to be 50 to 110 μmol/L. Individuals with blood phe concentrations of 120 to 600 μmol/L before starting treatment are classified as having mild HPA; those with concentrations of 600 to 1200 μmol/L are classified as mild PKU (sometimes a moderate classification is included for concentrations of 900 to 1200 μmol/L); and concentrations above 1200 μmol/L denote classical PKU.

5.5 A Dispositional Ontology of the Explanans μ

Lange (2007) argues for identifying a disease with a kind of incapacity; by introducing a static condition of deficiency in order to explain the dynamic events spreading within an organism this account fills a gap left by the mono- and bilayer views discussed in Sect. 5.3; Lange's concept can even be derived from both views by substituting in the bilayer view an absent capacity for the pathological process on the micro-level, or by adding to the monolayer view a disparate entity that is not a process. The combination of a static and a process-like element in one theory is truly neither a mono- nor a bilayered reconstruction – but it is akin to both: Disease is neither an all-embracing process (which would preclude its explanatory force) nor an underlying process explaining another (which would be a simplification of its complexity) – disease is rather an absent disposition which explains the entire procession of facts and events in the course of an individual's disease.[6] The clinical and pathological phenomenologies of succession can easily be reconciled with the alternative concept that diseases are states, static configurations or incapacities with a dependent sequence of resulting pathological and clinical changes; a disease entity as an incapacity is distinct from the process and it sustains its manifestation.

In order to further clarify his account of a disease, Lange contents himself with PKU in order to approximate the meaning of incapacity. In a first step, a successful identification of this nosological term with six eligible states of incapacities is averted: classical PKU is *not* HPA, since HPA is neither a necessary nor a sufficient state for the usual characterization of PKU; classical PKU is an incapacity which is

[6]All that can be learned from Lange about the concept of an incapacity is condensed into two sentences:

> By an 'incapacity', I mean nothing more than the lack of a certain capacity, and a capacity is simply a disposition (i.e., a power). A fragile vase has the capacity to break and an incapacity to speak, for example.

Further, according to Lange, the distinction of active and inactive enzymes reflects a reference to dispositional terms like incapacity and activity (Lange 2007, 287, FN 18 and 19; 276).

present independent of the level of phe in the blood. Next, PKU is *not* the incapacity to catabolize phe sufficiently, since all other HPAs also involve this incapacity (which may be caused, e.g., by an insensitivity of PAH to BH_4).[7] Further, PKU is *not* the state of *synthesizing* too little PAH, since synthesizing a lot of inactive PAH is in accordance with PKU; classical PKU is *not* the state of *having* too little PAH, since having a lot of inactive PAH is in accordance with PKU; classical PKU is *not* the state of lacking a *particular* sequence of (good) aminoacids in the enzyme PAH, since silent polymorphism allows for many forms of intact enzymes; classical PKU is *not* the state in which PAH has a sequence of aminoacids disjunct to the set of *all (good) sequences* of aminoacids, since this would also be a motley description. Finally, PKU can *not* be identified with the state of a gene if the explanatory role of a disease has to be preserved; because if having a disease amounted to having a gene, this gene could not explain the disease, since otherwise the disease would explain itself. Note that, relative to the incapacity account, the genotype still may explain why a person has that incapacity. In conclusion, classical PKU is defined as the "incapacity to make enough active pheOH [PAH]" or the "incapacity to make pheOH [PAH] with enough activity" (Lange 2007, 275, 276, 287).

According to Lange (2007, 276–79), incapacities explain by invoking a function-analytical account (f.-a.-account hereafter) of explanation. Cummins has identified this important variant by confining function-ascribing statements to disposition statements (Cummins 1975, 758). Then, the disposition K to Φ of a containing system is explained by appealing to component capacities which are organized thus that their programmed manifestation elicits the manifestation of K; and the exercise of an analyzing capacity emerges as the function of the component. Such an explanatory strategy requires a background analytical account which mediates the connection between the functional ascriptions and K:

> The explanatory interest of an analytical account is roughly proportional to (i) the extent to which the analyzing capacities are less sophisticated than the analyzed capacities, (ii) the extent to which the analyzing capacities are different in type from the analyzed capacities, and (iii) the relative sophistication of the program appealed to, i.e., the relative complexity of the organization of component parts/processes that is attributed to the system. (iii) is correlative with (i) and (ii): the greater the gap in sophistication and type between analyzing capacities and analyzed capacities, the more sophisticated the program must be to close the gap. (Cummins 1975, 764)

An explanation of some capacity of a system is rendered interesting by decomposing it into parts such that the capacities of the subcomponents are simpler than the *explanandum*-capacity, different in kind from it and organized in an elaborate manner. In short, two analyzing principles – evoking a gap of sophistication and a difference in kind – are bracketed by a reconciling principle that appeals to

[7] The reader traversing this section in Lange's study will certainly feel the tension between Lange's trial to grasp a *classical* concept of a disease in an article entitled with the *end* of disease. A second tension comes in from the fact that Lange neglects the concept of an intrinsic deficiency of an enzyme which can be defined by including *c.p.*-conditions.

the organizational complexity of the subcapacities which an ambitious research program is normally focused on.

The following analysis makes use of the concept of a correspondence between a capacity and an incapacity: let K be a capacity to Φ, belonging to a containing system which is present in individuals of a subgroup G_1 of an otherwise homogeneous species G, then:

$$(K' \text{ corresponds to } K) \leftrightarrow \forall x \in G_1 : Kx \wedge \forall x \in G \setminus G_1 : \neg\exists x : Kx$$

The non-instantiation of a capacity K for a system of members belonging to a subgroup will be abbreviated hereafter as $K'x$. Next, in order to implement incapacities into an f.-a.-account, two levels and, for each level, a capacity and its corresponding incapacity have to be distinguished; roughly, on the lower level, a capacity k to φ with the corresponding incapacity k', and, on the larger scale, a capacity K to Φ with the corresponding incapacity K' have to be considered separately. It has been proposed by Lange (2007) that the f.-a.-account of a capacity K is the one which supports the f.-a.-account of the corresponding incapacity K'; in other words, there is no f.-a.-account of a higher scale incapacity K', if an f.-a.-account of a higher scale capacity K is not available; the latter explanatory relation may therefore be called the *axis* of the explanatory framework (my terminology). The ultimate point of reference for explaining pathological phenomena, however, is a tacit background of empirical knowledge concerning normal capacities on the macro-level:

> A disease ascription takes place against a (generally tacit) understanding of the sorts of larger capacities [K] that are part of good health. Just as there is a tacit understanding of what a 'normal' diet is, roughly speaking, so there is a tacit understanding that the capacity [K] to eat such a diet (without certain effects) is part of being in good health. This capacity is compromised by the incapacity [k'] to make enough active pheOH [PAH]. (Lange 2007, 276)

How exactly the compromising capacity k' relates to the normal capacity K is revealed by supplementing the complete framework of a medical explanation of diseased states:

> A disease is an incapacity that is explanatory: Insofar as the capacity [k] to X can figure as a component in an interesting function-analytical explanation of the capacity [K] to Y, the incapacity [k'] to X can figure in an interesting explanation of the incapacity [K'] to Y, and so tends to better qualify as a disease. (Lange 2007, 279)

A more concrete statement looks like this:

> [...] the incapacity [K'] to eat ordinary bread without various PKU symptoms is interestingly explained by the incapacity [k'] to synthesize enough active phOH [PAH] because the capacity [K] to eat ordinary bread without various PKU symptoms is interestingly decomposed into the capacity [k] to synthesize enough active phOH [PAH] and other such capacities. (Lange 2007, 279)

Summarizing these statements yields the relations depicted in Fig. 5.2 in order to clarify references in the further analysis.

Consequently, an explanation by incapacities can be reduced to the following logical structure:

$$
\begin{aligned}
N x \leftrightarrow (k' x \wedge (k' \quad \text{f.} - \text{a.} - \text{explains} \quad K')) \\
\text{and}: \quad (k' \quad \text{f.} - \text{a.} - \text{explains} \quad K') \leftrightarrow \exists\, K, k \quad \text{such that}: \\
(k \quad \text{f.} - \text{a.} - \text{explains} \quad K) \wedge (K', k' \quad \text{correspond to} \quad K, k)
\end{aligned}
$$

5.5.1 *Advantages and Drawbacks of the Incapacity Account of Diseases*

The incapacity account of diseases supplemented with an f.-a.-explanatory and contrastive framework is an achievement for the philosophy of medicine; this account effectively interpolates between the bilayer view of organismic processes envisaged in the mind of the clinician who is predominantly engaged in clinical practice, and a strict process theory defended rigorously by some philosophers. Another advantage is reflected by its impetus to deliver a thorough explanatory account of disease phenomena which relies on principles to identify objective individual disease entities. And finally, it holds out the prospect to be adjusted in such a way that an extension of its applicability to states of mental disorders might be possible. In spite of these positive points, the task of this section is to demonstrate the deeply hidden drawbacks of the incapacity account and to motivate another approach which also supports the μ-hypothesis and covers the advantageous points indicated so far. The latter aim will be pursued in Sect. 5.6.

Metaphysical Foundation of Incapacities One major metaphysical problem of the concept of incapacities[8] concerns its consequences for the simple conditional analysis (SCA) of dispositions. In the case of absent dispositions, SCA yields that any stimulus that does not entail the manifestation of D would indicate an

[8]It is taken for granted that the concept of a capacity deserves more precision in the context of medical sciences, e.g., it is not just the capacity to produce enough active PAH which explains the absence of symptoms, but rather the *manifestation* of that capacity. Furthermore, a person afflicted with PKU still has the capacity to eat a normal diet, to chew, to swallow, to digest, to assimilate, etc. Then, a major emphasis has to be placed on a problem concerning the concept of an incapacity: as a general concept for the ontology of diseases it confers a strict and consequently pervasive evaluative notion to nearly all disease terms. It is doubtful whether such a connotation is desirable from a psychological point of view.

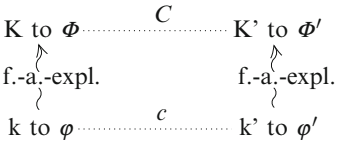


Fig. 5.2 Incapacity account of diseases supplemented with an f.-a.-explanatory (axis k - K; k' - K') and contrastive (c, C) framework

incapacity; in consequence, an inflation of incapacities is imminent. However, this problem can be solved by applying a modified SCA.[9]

Organization of Incapacities Another major problem of the incapacity account consists in the fact that simplistic (in)capacities of subcomponents are appointed to serve the base of the explanatory framework depicted above (see Fig. 5.2) – just recall the incapacity 'not to produce enough active PAH' (or its corresponding capacity); however, these simplistic bases are inconsistent with the requirements of the functional analysis of capacities; a simplistic micro-(in)capacity may be simpler and less sophisticated than the *explanandum*-(in)capacity; it may further be different in kind, but certainly fails if it comes to delineating the organizational complexity established by sub-(in)capacities, since there is only one in each case. Lange seems to have been aware of this inconsistency, since his own use of the incapacity account imports suggestive elements; for instance, the given concrete explanation of PKU refers not only to the already known incapacity but also to 'other such capacities' (compare FN 5.5). Next, lead poisoning is analyzed into several specific biochemical incapacities which reinforces the impression that the simplistic picture conveyed by the major part of the other examples treated in Lange's publication cannot be the last word (Lange 2007, 280).

Interactions of Incapacities Aside from this allusion to a cooperative kind of organization in the case of lead poisoning, there is also the insinuation that one incapacity ignites another incapacity (Lange 2007, 281). Lange admits that one disease (an incapacity according to Lange) can compensate for the compromise received by the first one: it may occur that an incapacity does *not* entail symptoms "not just because something else (perhaps even a second disease) happens to compensate for the incapacity" (Lange 2007, 277; compare 288, FN 21).

Absence versus Appearance of Capacities Furthermore, there is the problem that, in the context of medical explanation, the reference to incapacities neglects the equally justified reference to newly appearing capacities, e.g., somebody who contracts

[9]In accordance with the modified simple conditional analysis, Dx is defined along $Dx \leftrightarrow \exists S(Sx \; \Box\!\!\rightarrow \; Mx)$, wherein $\Box\!\!\rightarrow$ stands for the counterfactual conditional, S for a D-specific stimulus, and M for a D-specific manifestation (Choi 2006). The transformation of this analysis into one for absent dispositions leads to: $\neg(Dx) \leftrightarrow \neg(\exists S(Sx \; \Box\!\!\rightarrow \; Mx))$, and this yields: $\neg(Dx) \leftrightarrow \forall S(Sx \; \Box\!\!\not\rightarrow \; Mx))$. The meaning of the right side of this biconditional is the following: If x were exposed to whichever D-specific stimulus S it would not show any D-specific manifestation.

PKU gains – aside from the specified incapacity – the capacity to synthesize new metabolites.

Differentiation of Incapacities The argument that the incapacity 'not to keep blood levels of phe low' which is characteristic of HPA in general is absolutely distinct from the incapacity 'not to produce enough PAH' is not quite convincing because it is either circular for the case of PKU (active PAH keeps phe blood levels low, *c.p.*) or it implies collateral assumptions which are not specified (Lange 2007, 281).

Localization of Incapacities To conclude this list, there is a problem which is of both metaphysical and epistemic relevance: if PKU is the incapacity 'not to produce enough active PAH', then the question is justified how to specify the carrier of this property or how to localize the inefficient producer. This epistemic aspect is of eminent importance for medical science, since it is in this context decisive to know whether the liver, the genes, or the organism (or a collective of a society) host the incapacity.

Contrastive Contribution The remaining problems relate to the contrastive components C and c of the explanatory relations in Fig. 5.2. The capacity K declared to be normal and the corresponding incapacity K' on the larger scale level have to be tied together in such a way that the claim of correspondence is satisfied. The capacity 'to be able to eat a normal diet (without having PKU symptoms)' corresponds in this manner with the incapacity 'not to be able to eat a normal diet (without having PKU symptoms)'; that is to say that if a normal diet is ingested, PKU symptoms will appear. This example demonstrates that K and K' have to be chosen thus that they include one reference to the underlying capacities k and k' which belong to some pathways of the digestive system and another reference to the presence or absence of symptoms. The use of the first reference allows to shift away from the level of symptoms and to find a capacity that is related to the smaller scale capacities because k and k' are constitutive elements of the pathways of the digestive system, so that the choice of K as the capacity to digest is directly related to both K' (via C) and k and indirectly to k' (via c and C). However, (as this analysis shows) the symptoms play an essential role within this relational framework in order to make the relevant references distinct – otherwise any malfunction of any pathway within the confines of the digestive system would be PKU. To summarize, the problems related to C and c are due to the task of identifying corresponding larger scale (in)capacities rooting in the identification of smaller size (in)capacities; this creates a tendency to search for (in)capacities of whole systems (e.g., the digestive system);[10] in consequence, a satisfaction of the conditions inherent to C and c renders the informative contents of the explanatory axes k-K and k'-K' redundant and reduces symptoms – at first

[10]Indeed, a capacity like 'ingesting a normal diet' is a capacity attributed to one of those systems which a complete and intact organism is composed of – in this case to the digestive system; a look at standard textbooks of physiology and physiological chemistry will reveal that systems of this generic type (e.g., the circulatory or the respiratory system) constitute a clear minority in comparison to the large number of diseases.

sight – to indices attached to these axes. Therefore, the advantage of harmonizing two explanatory accounts by a contrastive supplement leads to a concealment of one of the most important aspects of the medical enterprise – using symptoms as signs for the sake of identifying diseases. There is good reason to circumvent the confrontation of an explanatory account with the symptoms of a disease: it is the unpredictable shift of the (partially functional) symptomatology with a change at the explanatory base (see below).

Finally, the ranks of the *relata* in the contrastive statements C and c of the explanatory scheme in Fig. 5.2 are thus that the explanatory axis k-K has explanatory priority over k'-K'; in the context of discovery it seems to be the other way round: normal (and even already discovered anormal) states are the points of departure – they are taken as given – and serve as reference in relation to which deviations are examined; that what explains the deviations does not explain the normal state in a symmetric manner. This applies to both clinical symptomatology and experimental models of disease.

5.6 A Mechanistic Ontology for Disease Entities

The analysis of two representative ontologies for disease entities has revealed their achievements as well as their limits. The remaining task is to challenge the integration of the concept of mechanisms into the explanatory accounts of scientific clinical medicine. In Sect. 5.6.1 it will be shown that the ontology of mechanical philosophy has not been successfully embedded into this special context and fails to grasp the use of disease terms in a satisfactory manner. Therefore, in contrast to these current conceptualizations and in order to establish a sound foundation of medical science, this study will defend a position which will not only serve philosophy of medicine but will also have some important repercussions on the philosophy of mechanisms. The procedure in the remaining sections of this article will be the following: In Sect. 5.6.2 a neuron diagram for mechanisms relevant to cases of HPA will be presented. In accordance with the characteristic features of this diagram, the concept of a mechanical base of PKU will be introduced, see Sect. 5.6.3.

> Such a (disease-) specific base is both an explanation of what *actually* happens in order to bring about a phenomenon and a base for a function-analytical account which explains properties by demonstrating how *possibly* events may be organized in order to bring about the *explanandum*.

It is a peculiarity of medicine to take into account such *explananda* (e.g., properties, functions, capacities) which are settled on a macroscopic organismic level. The distinction between *what-actually* and *how-possibly* explanations has

been emphasized by Craver (2007, 107–63) and has inspired the distinction between *what-actually* and *what-possibly* explanations (for more details compare the discussion by Piccinini and Craver (2011)). According to this view, medical science is dwelling on the borderline between at least two explanatory concepts and strives, first, to augment knowledge concerning actual mechanisms and, second, to defend functional analyses which consider possibly contributing capacities. By means of two (or several) specific bases the meaning of omissions and preventions will be elucidated in Sect. 5.6.2. Next, the preliminary concept of a specific mechanical base is generalized. A general base imparts a meaning to contrastive statements referring from deviating organismic states back to a standardized one. This step introduces the third explanatory component and establishes the diacritic notion of normal and pathological cases. Finally, in Sect. 5.6.4 the determination of the limits of the extent of mechanistic bases can be shown to be free of arbitrary evaluative notions.

5.6.1 A Survey of Some Mechanistic Ontologies of the Explanans μ

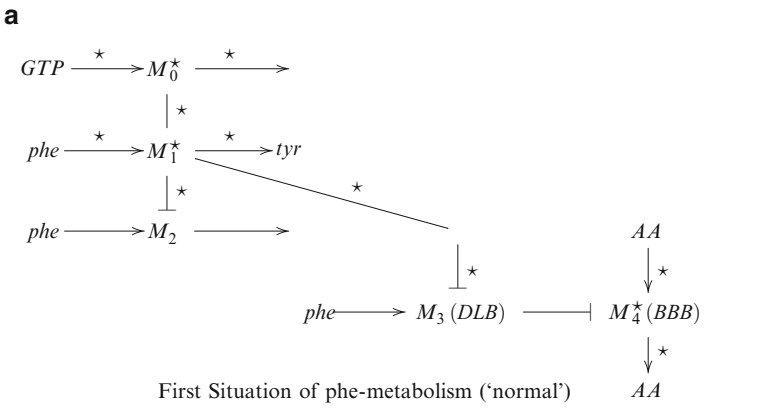
The canonical work of Glennan and Craver in the field of mechanistic philosophy has imparted a major impetus to studies not only in the philosophy of neuroscience, but also of many other disciplines in the special sciences (Craver 2007; Glennan 1996). Among the latter are some which suggest to use the mechanistic account of explanation in the context of medical clinical sciences, a direction which has been sketched with a few strokes by Craver (2007, 110, 132, 147, 155). Most of these studies praise the utility of mechanistic accounts for medicine, but are mainly concerned with the relation of mechanistic correlations to the informational content of correlations obtained by epidemiological or randomized clinical trials (Campaner 2011b; Illari et al. 2011, 25–125). That mechanisms are used for explanatory reasons in medicine is a strict presupposition of these studies. Therefore, the questions of what constitutes a disease in purely mechanistic terms or what distinguishes a physiological organismic mechanism from one which contributes to a diseased state cannot be answered. In consequence, only a few studies remain which come close to the intent of this analysis: Thagard, to begin with, displays flow charts instead of addressing the mechanisms of cancer development or ulcer formation (Thagard 2000); furthermore, the *conditions* for developing an adult-onset diabetes are outlined, whereas the mechanisms responsible for the occurrence of this event are not; to conclude this short overview, mechanisms do not play any circumscribed role at all explaining a disease: “For each disease, epidemiological studies and biological research establish a system of causal factors involved in the production of a disease. [...] We then explain why a given patient has a given disease by instantiating the network, that is by specifying which of the factors operate in the patient” (Thagard 1998, 69, 73 and 74, respectively). The result of combining biological with epidemiological theories is “a narrative explanation of why a person gets sick.” It can easily be inferred that, for Thagard, disease is just a

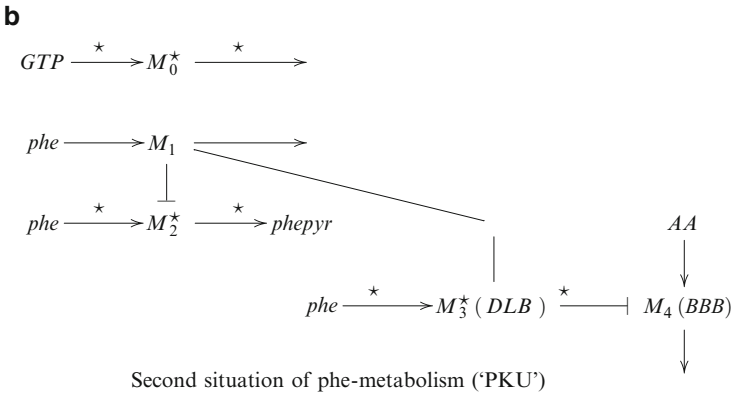
collection of symptoms (e.g., complaining of stomach pains) and that the causal network instantiation (CNI) account of explaining a disease is of diagnostic value mainly, responding to the question 'What disease does this patient have?', but not to the question of what explains a disease if it is conceptualized according to Thagard. Thus, CNI provides "an explanatory schema or pattern" collecting all the "statistically-based causal relations" which are relevant to the case at hand; but it does neither provide a deductive explanatory pattern nor a mechanistic reduction. Apart from this problematic mechanistic account serving the explanation of disease, two further positions remain relevant for this study: The first tackles problems similar to those of Whitbeck's analyses and is therefore settled beyond a mechanistic account of disease entities (Severinsen 2001). The other identifies diseases with malfunctioning mechanisms; a disease is an "impairment of the normal mechanism" which can be attributed to a certain step of a mechanism; "the corresponding disease is then explained with reference to the interruption in the sequence" (Nervi 2010, 217). It is further argued that because of pragmatic relevance patho-mechanisms, even though grafted on physiological mechanisms, are treated in medicine as more or less independent entities. In order to delimit them, three criteria are introduced which are, however, not independent of a reference to a mechanism in its normal state. Apart from this circularity, it can be doubted whether in the section about medical practice a theory about the explanation of diseases is achieved at all, or whether just patho-mechanisms are presented which underly a diseased state of an organism – so that a mechanistic explanatory account of disease does not emerge.

5.6.2 A Neuron Diagram for PKU and Its Generalization

Figure 5.1 illustrates an extract from the metabolic pathways realized in a human organism. There are entities which enter or leave a net of transformational intermediate steps symbolized by arrows which connect entities one- or bidirectionally; transformations are either enabled by enzymes or occur non-enzymatically in a spontaneous manner. Additionally, bars (crossing an arrow) indicate the localizations of several possible metabolic impairments; the related clinical disorders are described in separate tables included in the publication of Blau et al. (2003). Each table lists the symptoms of the respective disease which are arranged according to two dimensions, one for categories comprising sets of symptoms (like characteristic clinical findings, laboratory findings, or symptoms attributed to the nervous, circulatory or respiratory system, etc.) and the other for the period of life of the patient (like neonatal period, infancy, childhood, or adolescence). Obviously, this information is not sufficient to deliver an account of what a disease is and what it explains. This goal can be approached, however, by supplementing the pathways depicted in Fig. 5.1 with a neuron diagram;[11] four mechanisms have been selected, rearranged and the result is communicated in Fig. 5.3: M_0 is the mechanism which

[11] A similarly conceptualized diagram can be found in the paper by Scriver and Waters (1999, 270).





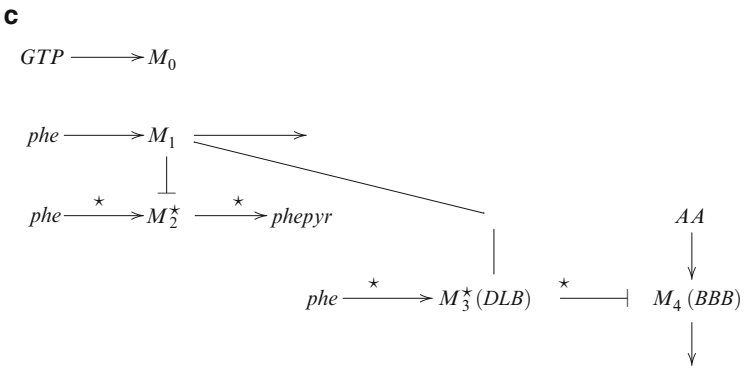


Fig. 5.3 A Neuron Model for three distinct situations of phe-metabolism

supplies the cofactor BH_4 to PAH; M_1 is the mechanism which transforms phe to tyr; M_2 is the mechanism which transforms phe to phepyr; M_3 is the mechanism of diffusion of phe from the liver cells into the blood (DLB) being responsible (in a way) for HPA;[12] a fourth mechanism, M_4, has been added which represents the aminoacid transport across the blood-brain barrier (BBB).

At this point, it is appropriate to consult the philosophy of mechanisms. Regardless of the relative diversity of normative accounts of mechanisms available in the philosophical literature, all of them take some basic concepts for granted (as long as they adhere to the system tradition): a mechanism is a system of parts which are disposed to interact in a certain way – along a characteristic (not necessarily temporally structured) path of changes running through the parts of the system; this pattern of events and interactions is elicited by providing conditions which are in accord with the enduring existence of the parts; such conditions, which lead to the revelation of a pattern of property changes along the path leading through the system, may be referred to as initial conditions. In addition to these singular property changes, there is typically a phenomenon (or behavior) distinct from the properties of the parts, but depending on their organized and orchestrated interaction: a mechanism is not just a mechanism, but a mechanism *for a phenomenon* to occur; mechanisms *explain* a correlated *explanandum*-phenomenon. This succinct overview may suffice to justify at least a regularity account of mechanisms in order to describe mechanistic events.[13]

In Fig. 5.3, a $\star$ attached to a symbol for a mechanism M_n ($n = 1, 2, \ldots$) indicates that the mechanism manifests a path of property changes in virtue of an adequate stimulus condition, which is indicated by an incoming arrow endowed with a $\star$; the outgoing arrow with a $\star$ indicates the presence of the higher-order phenomenon which is explained by the manifestation process occurring within $M_n^\star$. That a mechanism is a mechanism *for a phenomenon* translates into $\rightarrow M_1 \rightarrow$, and if the precipitating conditions obtain, this symbol turns into $\overset{\star}{\rightarrow} M_1^\star \overset{\star}{\rightarrow}$.

The relational networks shown in Fig. 5.3 clarify that the meaning of mechanism is not exhausted by being a mechanism *for a phenomenon*, but has to be supplemented by the relational concept of being a mechanism *for a mechanism*; e.g., in Fig. 5.3a the mechanism of M_0 supplying M_1 with BH_4 is one of the relatively simple cases where a mechanism delivers a positive condition for a succeeding one, M_1, so that it is justified to write $M_1^\star$. Inversely, by virtue of keeping the level of phe in the cytosol of liver cells low, $M_1^\star$ prevents the manifestations to be expected due to M_2 and M_3; therefore, coincidental to $M_1^\star$, neither phepyr is produced (M_2) nor does phe accumulate in the blood (M_3). The claims for such preventive relations are symbolized by blunted arrows; indexing them with $\star$ is analogous to the conventions for the pointed arrows: an incoming blunted arrow with a star indicates

[12] In a paper by Torres (2009) the problem is treated that there are physicochemical events which interrupt the concatenation of property changes which is a demand of the systemic branch of mechanistic theories.

[13] For a more detailed regularity account of mechanisms, see Schrenk (2007, 127–37).

that one positive (negative) factor of the initial conditions for the mechanism under consideration is absent (present) due to the preceding mechanism, which may be realized in different manners, e.g.: M_1 and M_4 are connected by two intermittent preventive relations in series. With $M_1^\star$, M_3 follows, the lack of phe in the blood due to M_3 entails a lack of an inhibitory component directed to M_4. The inhibition on M_4 is therefore [absent],[14] and under the stimulating condition of aminoacids (AA) in the blood, M_4 manifests its transport capacity and qualifies for a $\star$.[15]

Figure 5.3b further conveys that a mechanism M_1 which has [lost] its disposition to transform phe to tyr – in spite of an [*c.p.* adequate] stimulus condition given by phe – will also [lose] its inhibitory (or preventive) effects on M_2 and M_3, so that these two – given the presence of phe – manifest their proper phenomena; $M_2^\star$ generates phepyr, since the level of phe is sufficiently high due to the absence of M_1, and $M_3^\star$ becomes responsible for HPA. This latter phenomenon is a negative condition for $M_4^\star$, the manifestation of which (transport of AA) therefore ceases. The subsequent lack of AA beyond the blood-brain barrier is made responsible for the deterioration of the cognitive development of patients who are equipped with the unresponsive M_1. Thus, an [absence] of a property which M_1 has had [before] leads to an [absence] of a prevention related to M_3 and, therefore, has a permissive (enabling) effect on M_3; in short, the mechanism M_3 is disinhibited by M_1, and therefore M_3 is enabled ($M_3^\star$) to inhibit $M_4^\star$ (M_4).

The exposition of the circumstances in Fig. 5.3 demonstrates that the interrelations of Fig. 5.3b are described by referring to the situation of Fig. 5.3a and *vice versa*. Thereby, contrastive explanations are imported into the overall explanatory account: an absence is the non-instantiation of a property that has been attributed to a mechanism in the contrast situation; and inversely, as there are absences, there are mechanisms which emerge by displaying their phenomena in the contrasted (normal or deviating) situation. Furthermore, there is a basic rule concerning the (not necessarily sufficient conditions for the) attribution of a preventive relation to two phenomena in isolation: if an absence of the property of one mechanism is regularly associated with the appearance of a phenomenon bound to another mechanism and this appearance is due to the former absence (as may be tested by manipulation),

[14] In the text to follow, bracketed terms are considered to prepare the reader for the use of contrastive statements.

[15] The debate about mechanisms and functions has received major impulses recently and therefore the *mechanism-for-mechanism*-concept of mechanistic functions advanced in this study demands a demarcation against prominent different concepts: For obvious reasons, the *mechanism-for-mechanism*-concept does not share the perspectivalist view announced by some authors (e.g., Craver 2013), it does, however, share the independence from an evolutionary selective account and thus differs from the view strengthened by Garson (2011). If admittedly a mechanism produces a plurality of phenomena, this plurality does not preclude a precise determination of some distinct *for*-relations, so that functions of a mechanism can be identified in an objective manner. And since the *for*-relation hands one mechanism over to another one, there is no escape from this layer to a realm of goals: the *for*-relation is not imbued with teleological concepts. Therefore, this concept of a *for*-relation has to be demarcated against the notion of a gene being *for* a phenotypic trait or traits (Kaplan and Pigliucci 2001).

then this correlation is termed an inhibitory (preventive) relation. For instance, the change of $M_4^\star$ to M_4 (an absence) not only coincides with the change of M_3 to $M_3^\star$ (an appearance), but can also be shown to be a change-relating generalization – one intervention is provided by nature itself which presents to us an M_1 which cannot qualify for a $\star$ under the conditions of Fig. 5.3a. Another example: according to the basic rule, the relation between M_0 and M_1 is equivocal when focusing on Figs. 5.3a and 5.3b *or* 5.3a and 5.3c, and therefore in Fig. 5.3b the starred outgoing arrow of $M_0^\star$ points to the right; the M_0-for-M_1 relation is broken, but M_0 may still sustain relations to mechanisms not shown in the diagram. In Fig. 5.3c there is no outgoing arrow from M_0 at all because this mechanism is not in a condition to obtain the phenomenon required for M_1 or for any other mechanism connected to M_0 by a *for*-relation. The basic rule determines type and direction of the *for*-relations between coherent mechanisms.[16] If, in addition to the basic rule, certain background conditions are taken into consideration, then the directionality of the entire cascade of *mechanism-for-mechanism*-relations shown in the diagrams becomes fixed; there is a constant influx of phe towards M_1, M_2 and M_3; M_2 demands higher levels of phe, so that its initial conditions are obtained (its affinity to phe is lower than in the case M_1); M_3 is a complex, but passive mechanism (phe diffuses downhill and is then transported as a solute by convection); M_4 is supposed to be a (passive) transporter with saturation kinetics, and therefore an elevation in the plasma concentration of one amino acid (e.g., phe) will reduce the uptake of other amino acids into the brain, so that it is not the transporting mechanism that is inhibited but the flux of AA. It can be generally asserted that the signalized inhibition of a mechanism does not mean a rupture of the mechanism as such, but indicates a contribution to the initial conditions of the associated mechanism so that they remain incomplete; neither does inhibition mean that the inhibited mechanism is the cul-de-sac which would not allow the transit of property changes to an adjoining third mechanism. In fact, *directionality* is the common and overarching feature of the ensemble of pointed *and* blunted arrows of the neuron diagram.

In conclusion, the relations shown in Fig. 5.3 differ from the more skeletal presentation of pathways shown in Fig. 5.1 in so far as the former include additional information which concerns incitement as well as inhibition; physiological inhibition alone does not affect directionality; thus, an additional layer of explanatory import is installed above the net of relations which arise from the simple dual connection of mechanisms focused on in isolation. It may be hypothesized that directionality is a concept that unifies types of relations which are kept apart by manipulative procedures alone; therefore, directionality may come close to a concept of causality.

[16] Applying only the first part of the basic rule would leave the choice of direction to you.

Finally, it is a simple exercise to see that defining the situation in Fig. 5.3b as 'normal' and the one in Fig. 5.3a as 'deviating' will result in a change of the connoted appraisals. The intuition demanded for this change – for example, of the values attributed to $M_3^\star$ and M_4 – can be challenged by supposing, for example, that Fig. 5.3a represents an enhanced state which is not available under normal conditions; then removal of phe from the blood might be one measure to advance enhancement. Or, suppose that AA is an (endogenous or exogenous) toxin; then the situation in Fig. 5.3b will be preferable and the infusion of phe (or inhibiting M_1) might protect the organism in the deviating situation of Fig. 5.3a against the toxin.

5.6.3 Identifying Diseases Within Complex Mechanistic Bases

The analysis so far has emphasized that

> four explanatory parts have to be kept apart: the relation between a mechanism and its *explanandum*-phenomenon, the *mechanism-for-mechanism*-relation, the f.-a.-account of symptomatology and the contrastive explanation. Directionality appears to point to a causal organismic structure uniting several mechanism.

It is also important to note that to assert a *mechanism-for-mechanism*-relation is independent of the contrastive statements which are implemented if a conventional normal situation is kept fixed and confronted with one or more deviating situations; such comparisons need not be necessarily explicit nor can they be read off from the mechanisms in isolation (e.g., if examined in an experimental set-up) – even if they are directly related to other mechanisms by means of the *mechanism-for-mechanism*-type; rather, a contrastive statement depends on the confrontation of two situations, one of which is chosen to be the reference for certain exogenous reasons; e.g., in Fig. 5.3b the intuition of a deviation from the situation in Fig. 5.3a comes in from two sides: First, from the (contrastive) statement that under standardized initial conditions (and you are rather free in choosing the standard without getting disturbed by your choice) M_1 does not respond to phe in the usual manner known from the standard situation in Fig. 5.3a, which may correspond to a quite objective evaluation of the facts. The other reasoning emerges from the judgment that the result (probably) produced by M_4 is not simply different from that produced by $M_4^\star$ – but undesirable; this undesirable aspect of the effect, first, enjoins a definite asymmetry on the situations under comparison (you will not voluntarily inverse your judgment on the preferences), and, second, may be attributed to the contrastive statement relating to M_1; without a preceding objective statement, the undesirable effect (probably) due to M_4 could be projected onto the whole chain of entities being in continuity with M_4 by relations of the *mechanism-for-mechanisms*-type;

or, it could equally well be projected onto one single entity out of these or onto any set of them.

Before introducing the concept of a complex mechanistic base, it is useful to recall, first, the clinical spectrum of symptoms which collectively indicate PKU: in medical literature the major (undesirable) effect of PKU is described as a severe mental retardation quantified by means of the IQ score; apart from this cognitive phenotype, untreated phenylketonuria is clinically further characterized by epilepsy (a disease within a disease), hyperactivity, peculiarities of gait, stance, and sitting posture, delayed psychomotor development, motor disturbances, increased muscle tone and brisk tendon reflexes, EEG alterations, fairer hair and lighter cutaneous pigmentation than other family members, eczema-like rash, vomiting, typical mousy odor of the urine (a pathognomic sign due to phenylpyruvate extruded from M_3); there may be accompanying behavioral psychiatric manifestations (e.g., autism) and some other less frequent satellite symptoms.[17]

Secondly, the task remains to shed a light on another aspect not disclosed in Fig. 5.3: the depicted neuron diagram is not the whole story – but it represents a center of growth for further relations of the *mechanism-for-mechanism*-type. Some hints may suffice to clarify this idea: HPA inhibits TH (see Fig. 5.1); in consequence, tyr is not effectively converted to dopa; since dopa is the precursor of dopamin, catecholamines, melanin, and thyroid hormones, some effects of HPA on diverse extrahepatic systems of the human organism may be anticipated: the lack of Dopa in the nigrostriatial system of the diencephalon may induce a Parkinsonism with clinical signs similar to those found in M. Parkinson; the lack of melanin in melanocytes yields a light-coloured skin; the lack of catecholamines may be responsible for symptoms related to the sympathetic part of the vegetative nervous system. Next, the aminoacid trp is not converted (effectively) to serotonin anymore, since TPH is inhibited by HPA as well (see Fig. 5.1); since serotonin serves as neurotransmitter in the CNS, disorders of neural circuits are to be expected; serotonin is also stored as a tissue hormone in the enterochromaffine cells of the digestive tract, and lacking this hormone may be responsible for peristaltic disorders; in the pineal gland, serotonin is converted to melatonin, so that the shortage of this hormone will have a variety of repercussions on the mechanisms depending on the presence of this hormone. What these sketches intend to reveal is a fifth explanatory aspect in conjunction with systems of mechanisms for mechanisms: The concatenations displayed in Fig. 5.3 do not form an independent island in the organism, but may be *extended* according to regulations given by the set of possible links of the *mechanism-for-mechanism*-type. There are two ways to perform such extensions – by completing an already established base and by looking for adjoining bases. Thus, there is a distinction

[17] See the clinical synopsis in OMIM database under no. 261600; some clinical findings are available in the papers of Pietz et al. (1998); Pérez-Dueñas et al. (2005); interestingly, it is difficult to encounter clinically useful diagnostic criteria in recent literature which may be due to the fact that diagnosis for PKU is made by biochemical methods in the early period of life: the biochemical phenotype has replaced the clinical.

between a mechanistic base with its proper (direct) extension and its (indirect) connections with other mechanistic bases, which again have their proper extensions. The dominant base of PKU, e.g., assembles the relevant parts of a containing system S with a capacity K' and their respective component capacities $k'_1, \ldots, k'_n$ so that the parts establish an explanatory f.-a.-account of K' on a higher (and finally clinical) level. Within a base, one can make the distinction of a fixed point (e.g., the step catalyzed by PAH) and some adjoining elements, e.g., the transport across the BBB, or – in more distant organismic parts – the lesions to myelin sheets of cerebral cells, the liberation of peroxides in the brain matter, etc., which together constitute the extension of the dominant base and contribute to the f.-a.-explanation of the clinical phenomenon K' (mental retardation). This special base of PKU is the one mainly under examination in scientific medical literature. It has, for instance, been a highlight to figure out that there are persons with HPA who take a normal developmental course – due to preemption: in these particular individuals there are gene loci coding for proteins which intervene with the transport across the BBB such that the brain tissue is protected from the deleterious effects arising from a high phe-level in the blood. Apart from this base of major importance (consisting of a fixed center, an extension of this center and a fluctuating periphery) there are also bases for other (less threatening) symptoms of PKU which are in mechanistic contiguity with the major base; e.g., the catalytic steps performed by TH and TPH are the centers of such associated bases which assemble the parts for an f.-a.-account of higher (and finally clinical) level phenomena.

In sum, the explanatory account of PKU also consists of several interrelated bases among which one, say B_0, is of major importance since it reveals to be a common cause for the other adjoining bases $B_1, \ldots, B_n$. The elements of each (eventually extended) base serve as constituents both in a mechanistic *what-actually* explanation and in an f.-a.-account of explanation of the *how-possibly* kind referring to certain clinical signs or findings in the case of a specific disorder, e.g., of the PKU type. A formalization of these aspects clarifies the explanatory account achieved so far for the dominant base B_0:

A complete set of mechanisms $M_0 = \{M_0^1, M_0^2, \ldots, M_0^n\}$ and the relations $R_0 = \{R_0^{kl}, k, l \in (1, \ldots, n), k \neq l\}$ between its elements forms a *general* (complex and coherent mechanistic) base $B_0 = <M_0; R_0>$ which has to be specified further by selecting a set of mechanisms M_0^φ from M_0 and its corresponding set of relations R_0^φ from R_0 in order to chose a coherent explanatory *specific* base $B_\varphi = <M_0^\varphi; R_0^\varphi>$ with capacity k to φ within a (containing) system S equipped with (normal) capacity K to Φ. If B_φ is exposed to certain precipitating conditions, then B_φ manifests φ and B_φ (or rather its extension) f.-a.-explains Φ. In quite a similar manner, a base $B_{\varphi'} = <M_0^{\varphi'}; R_0^{\varphi'}>$ may be construed for a system S equipped with (deviating) capacity K' to Φ', and so on.

According to this definition, Fig. 5.3 displays three special bases drawn from a general base B_0 for explaining the capacity to perform certain intellectual tasks (measured by the IQ score); each base corresponds to one type of symptomatic state and each of them can be transformed into the other by rearranging the mechanisms and *for*-relations within B_0, as long as the constraints of the whole structure are respected. For the cases in Figs. 5.3b and 5.3c, the corresponding symptom is a common effect. Furthermore, the direct extension of these latter two bases is quite similar, since, supposedly, mainly HPA gives rise to those effects at a larger distance which enter into the f.-a.-account for the capacity K' to perform intellectual tasks (a hypothesis still under debate). However, the indirect extension of these bases is quite different.

The previous discussion concerning the concept of a complex mechanistic base can be summarized thus: First, a pivotal base comprises various sets of relations among a fixed set of mechanisms; the discrete patterns of interactions of a special base serve as elements within both a *what-actually* explanation and an f.-a.-account of organismic properties.[18] Each such pattern further explains its extension to other adjoining bases. Since each base comprises several patterns of interactive mechanisms, a base also provides a possibility to take into consideration manipulations and the consequences of changing from one pattern to another; in other words, it provides a lower level lever for manipulating higher level features of an organism; and since these features are explained by a pattern of a base, it can be imagined that with a base at hand one may make a good bet to get support from the *c.p.*-biconditional: $\mu_1 x \longleftrightarrow S_1 x$; with μ_1 to be identified with a pattern, say P_1, within a general base B_0, and S_1 being a symptom from the set S_N.

Having fixed the concept of a base so far will help to bring forth two related questions that will be treated in the remaining section: First, what are the norms for delimiting the constituents of a base? Secondly, if a base is given, what are the sources for a differentiation within the various groups of patterns belonging to one base? The quest for the concept of patho-mechanisms requires an answer to the latter question.

5.6.4 *Normative Criteria of a Complex Mechanistic Base*

A catalog of objective criteria is presented with the intention to demonstrate its general validity, since so far it has proved to be quite reliable for some standard cases of scientific clinical medicine.[19] Some peculiarities of a base have been mentioned

[18] Amundson and Lauder (1994) apply the function-analytical account of Cummins to explanatory accounts of functional anatomy.

[19] E.g., Tretter (2010, 44) describes a base for Parkinson's disease, schizophrenia, obsession and addiction in detail without invoking the concept of a base; similarly, in the fourth section of a publication by Craver (2013), a tacit reference is made to the concept of a mechanistic base.

already during the elaboration of this concept: a collection of mechanisms has to be endowed with a relationally coherent structure of the *mechanism-for-mechanism*-type. Since there are innumerable collections which fulfill the condition of coherence, it is further necessary to identify a pivotal mechanism (and its pivotal relations) within the global relational network. Quite frequently, pivotal mechanisms are distinguished by being correlated to a gene.[20] In the case of PKU, about 500 phenotype-modifying mutations within the locus of the phenylalanine hydroxylase gene (PAH) on chromosome 12q23.2 exist. Pivotal relations have been depicted in Fig. 5.3; M_1 relates to M_0, M_2, M_3, phe and tyr. The further extension of this central mechanism is determined by finding delimitations such that the resultant collection of mechanisms – designated as base – can be characterized by simple input and background conditions; the relations within a base display directionality; the output of a base is generally simple, but some outgoing relations connect the superior base with secondary adjoining bases. The extension of the limits around the pivotal mechanism are drawn such that the resulting structure is not only as local, but also as dense as possible. The density of a base is a measure which accounts for the number of distinctive patterns within the chosen base relative to the number of mechanisms in isolation, but unified in a base. E.g., Fig. 5.3 represents three patterns within one base composed of five mechanisms. Thus, within a base and among interconnected bases there is a hierarchical organization imparted by objective criteria.[21] Further, a base is called (theoretically) symmetric if it is conceivable to transform one pattern of its mechanistic states into another by applying certain feasible operations which change some selected properties of one or another mechanism of the base. A base is said to be practically symmetric if (therapeutic) methods are available which realize the foreseen (divined) transforming operations. An asymmetric base is defined correspondingly. Asymmetry might be a parameter to be employed in order to justify an objective foundation of disease entities. This line of thought, however, is not pursued in this study and, consequently, in order to focus the remaining discussion, bases will be assumed to be symmetric. One final decisive note with regard to the normative criteria of a base concerns the constraints it imposes on the patterns within the adjoining bases; as a matter of fact, the choice of one pattern of mechanistic states for a dominant base has repercussions onto the patterns of the adjoining bases; or, conversely, one pattern within the dominant base excludes at least some of the possible patterns of states in the adjoining bases.

The objective criteria of a (symmetric) base are not sufficient to assign distinctive roles to the one or the other of its potentially available patterns giving rise to a trace within the mechanistic framework of that base. One final hypothesis of this

[20]The idea that it is true for genes to be causes for the presence of organic components but not for that of diseases is taken from Scriver (2007). I use this relativity of the gene concept here in order to attribute to it a role of an objective confirmation of structures and orders present in complex mechanistic bases.

[21]It is this concept of hierarchy which is the clearest line of demarcation against the set-theoretic account of medical explanation advanced by Schaffner (1993) and still defended by Kendler and Parnas (2008).

study is that this separation comes from the *explanandum* of the f.-a.-account made available by exactly this base. Two aspects of this import from the higher to the lower level have to be distinguished: first, the assertion that a *difference* between two constellations of clinical signs, say Δ and Γ, can be explained by traces in the respective bases, say B_δ and B_γ. Such a Δ and Γ may coexist peacefully; they are just explained by their patterns within the base plus the relevant annexes. Second, it is inviting to escape this neutrality of explanatory accounts if a rather mild way of differentiation is introduced by choosing one as a standard case and, consequentially, regarding the others as deviating cases. This also adds – though only seemingly – a new explanatory strategy applicable to the so-called deviating cases:

> *Not* B_δ *but* B_γ is responsible *for* Γ and *not* Δ; standardization implies contrastive explanation (or even talk about absences as causes). In consequence, there is a surplus of explanatory content on the side of deviating cases – since the standard case lacks this contrastive attribution. This asymmetry due solely to the contrast on a macroscopic level may reverberate onto the bases; the one that corresponds to the deviating macroscopic state is generally called a *patho-mechanism* in medical literature. In philosophy of science, patho-mechanisms form a subgroup among the various posssible patterns of an explanatory base. It is still another matter to assign an evaluative notion to states like Δ or Γ, e.g., a normal (or good or – more explicitly – preferable) state to Γ and a deviating (or bad or – more explicitly – rejectable) state to Δ.

In a final step, the clarification of the explanatory structure of medical terms can be used to grasp one necessary condition for determining the sought meaning of the disease concept:

> A disease has to be identified with one coherent trace within a mechanistic base. The selected mechanism would then be baptized patho-mechanism. That is the point at which the majority of mechanistic ontologies referring to medical explanation sets in (compare Sect. 5.6.1).

5.7 Summary and Conclusion

A close examination of the role of diagnostic entities in clinical practice has motivated a search for an ontological characterization of the notion of a disease. Two views, i.e., a process ontology and an incapacity account, have been presented,

and both their strengths and inadequacies have been pinpointed. Because of the remaining insufficiencies, this study has turned to an alternative ontological conceptualization – that of a complex explanatory mechanistic base. Contrary to the shortcut view of most adherents of mechanistic philosophy, there is a long way from mechanisms to patho-mechanisms if the hybrid explanatory framework of scientific clinical medicine is taken into account: it is not patho-mechanisms which *bona fide* explain a diseased state of an organism; the concept of a mechanistic base (joining mechanisms by *mechanism-for-mechanism*-relations) is prior which again is grafted with an f.-a.-explanation of the symptomatic level. Thus, explanation in clinical medicine is based on the pivotal concept of a complex mechanistic base – at most a causal concept – the meaning of which is to unify a collection of separate mechanisms by stable (and not perspective) *mechanism-for-mechanism*-interrelations. A base, once fixed, on the one hand explains *what actually* happens, and on the other hand, by referring to clinical properties, *how* symptoms *possibly* may emerge; thus, a base unifies a mechanistic with a function-analytical explanatory account. The notion of a patho-mechanism comes in when alternative explanations for organismic states are already available for choosing one as the standard and the other ones as deviating cases. Thus, alternative explanations refer to *one and the same* complex explanatory base and are therefore entitled to be differentiated (e.g., when performing differential diagnosis) and compared – and the clinical *explananda* of the explanations similarly refer to this base. The choice of a standard organismic state determines a standard explanatory account among the range of available explanations. Then, the alternative explanations obtain an additional contrastive component relative to the standard case. This explanatory surplus does not yet imply that the respective explanation is endowed with an evaluative meaning. Standardization is a more general concept than that of normalization or even that of statistical norms, which for their part are surpassed again by evaluative judgments. Thus, there are several separating tendencies beyond (neutral) differentiation. Standardization is an early step which adds an explanatory surcharge to all deviating cases, or, inversely, infers a relative lack of explanatory power onto the standard case; this contrastive asymmetry imparts the contrastive component to the mechanistic patterns underlying the respective explanatory accounts; the causal bases of those cases which are differentiated against the standard are, in a second step, referred to as patho-mechanisms.

The concept of a mechanistic base determines local entities; it captures neither global processes (as in the process account) nor infinitesimally small pinpoints underlying a process (as in the incapacity account). A base conceptualizes a rather finite set of spatio-temporal entities (mechanisms) and their coherent relations (patterns). Note, however, that such local mechanistic bases decisively determine the associativity with distant adjoining mechanistic bases; and also within a base, a hierarchical structure could be verified, which consists of a central mechanism, pivotal relations and further more distant mechanisms; they all are enclosed within certain limits to be found at a distance from the center. Against this background, a base with its mutually excluding patterns provides insight into natural and therapeutic manipulatory influences on organismic states, and is a starting point

for pathogenetic and etiological concepts of the development of diseases. Objective normative criteria have been invoked in order to determine the structure and the limits of a (symmetric) base and its extension; different criteria have been advanced for its central and peripheral items; in this context, for instance, genes assume the normative role to confirm and fix the centers of the various hierarchies within and among mechanistic bases – not to *cause* a disease; genes establish only points of reference.

The ontology of a disease entity has been confined to a trace or pattern within a general base. This pattern is the μ that in the ideal case would fulfill the biconditional $\forall x : (\mu x \leftrightarrow S_N x)$ (see Sect. 5.2.1). It has been shown that it is possible to reduce this concept to that of grades of order among entities if the concepts of directionality and connectivity of relations between these entities are included; but no deeper insight into an objective landscape of organismic states can be acquired since a knowledge of the fundamental laws of life is not available at present.[22]

The findings of this study have ethical implications: if in some debates 'goods' are understood as the set of concrete and abstract environmental objects in correlation to the genes of an organism, then the related debates which concern the distribution of genes and goods among organisms will not only be relative to the applied concept of disease but will have to accept the present veil of ignorance which envelops the objective laws of life – and they will hardly be disclosed in the near future.

Acknowledgements This study has been supported by a grant of the German Research Counsel (DFG) and has derived benefit from its integration into the DFG Research Group 'Causality, Laws, Dispositions and Explanation at the Intersection of Sciences and Metaphysics'.

References

Amundson R, Lauder GV (1994) Function without purpose. Biology & Philosophy 9:443–469

Barosi G, Magnani L, Stefanelli M (1993) Medical diagnostic reasoning: Epistemological modeling as a strategy for design of computer-based consultation programs. Theoretical Medicine and Bioethics 14(1):43–55

Bechtel W (2011) Mechanism and biological explanation. Philosophy of Science 78(4):533–557

Blau N, Bonafé L, Blaskovics ME (2003) Physician's Guide to the Laboratory Diagnosis of Metabolic Diseases, 2nd edn, Heidelberg, chap Disorders of Phenylalanine and Tetrahydrobiopterin Metabolism, pp 89–106

Blau N, van Spronsen FJ, Levy HL (2010) Phenylketonuria. The Lancet 376(9750):1417–1427

Boorse C (1975) On the distinction between disease and illness. Philosophy & Public Affairs 5(1):49–68

Boorse C (1997) What Is Disease?, Humana Press, Totowa, NJ, chap A Rebuttal on Health, pp 3–134

[22]This consequence is emphasized as an objection to the presentiment expressed by Lange (2007, 282–285).

Bunge M (2003) Emergence and Convergence: Qualitative Novelty and the Unity of Knowledge. Toronto Studies in Philosophy, University of Toronto Press, Toronto, Buffalo, London
Campaner R (2011a) Explanation, Prediction, and Confirmation, Springer, Dordrecht, Heidelberg, London, New York, chap Causality and Explanation: Issues from Epidemiology, pp 125–135
Campaner R (2011b) Understanding mechanisms in the health sciences. Theoretical Medicine and Bioethics 32:5–17
Choi S (2006) The simple vs. reformed conditional analysis of dispositions. Synthese 148:369–379
Craver CF (2007) Explaining the Brain, paperback edn. Oxford University Press, Oxford
Craver CF (2013) Functions: Selection and Mechanisms, Synthese Library, Boston, chap Functions and Mechanisms: A Perspectivalist View, pp 199–220
Cummins R (1975) Functional analysis. Journal of Philosophy 72:741–765
Doroszewski J (1980) Hypothetico-nomological aspects of medical diagnosis part i: General structure of the diagnostic process and its hypothesis-directed stage. Theoretical Medicine and Bioethics 1:177–194
Forber P (2011) Reconceiving eliminative inference. Philosophy of Science 78(2):185–208
Garson J (2011) Selected effects and causal role functions in the brain: The case for an etiological approach to neuroscience. Biology & Philosophy 26(4):547–565
Giovannoni G, Ebers G (2007) Multiple sclerosis: the environment and causation. Current Opinion in Neurology 20(3):261–268
Glennan S (1996) Mechanisms and the nature of causation. Erkenntnis 44:49–71
Harding C, Blau N (2010) Advances and challenges in phenylketonuria. Journal of Inherited Metabolic Disease 33:645–648
Illari PMK, Russo F, Williamson J (eds) (2011) Causality in the Sciences. Oxford University Press, Oxford
Kaplan JM, Pigliucci M (2001) Genes 'for' phenotypes: A modern history view. Biology & Philosophy 16(2):189–213
Kendler KS, Parnas J (2008) Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology. Philosophical Issues in Psychiatry, Johns Hopkins University Press
Kitcher P (1997) The Lives to Come: The Genetic Revolution and Human Possibilities, revised edn. Simon & Schuster, New York
Korab-Laskowska M (1980) Hypothetico-nomological aspects of medical diagnosis part ii: Formal model of the explanation and testing procedures. Theoretical Medicine and Bioethics 1:195–205
Lange M (2007) The end of disease. Philosophical Topics 35:265–292
Maull N (1981) The practical science of medicine. The Journal of Medicine and Philosophy 6:165–182
Munson R, Roth P (1994) Testing normative naturalism: The problem of scientific medicine. The British Journal for the Philosophy of Science 45(2):571–584
Murphy D (2009) Concepts of disease and health. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, summer 2009 edn
Nervi M (2010) Mechanisms, malfunctions and explanation in medicine. Biology & Philosophy 25:215–228
Nordenfelt L (2007) The concepts of health and illness revisited. Medicine, Health Care and Philosophy 10:5–10
Piccinini G, Craver C (2011) Integrating psychology and neuroscience: Functional analyses as mechanism sketches. Synthese 183:283–311
Pietz J, Dunckelmann R, Rupp A, Rating D, Meinck HM, Schmidt H, Bremer HJ (1998) Neurological outcome in adult patients with early-treated phenylketonuria. European Journal of Pediatrics 157:824–830
Pérez-Dueñas B, Valls-Solé J, Fernández-Alvarez E, Conill J, Vilaseca M, Artuch R, Campistol J (2005) Characterization of tremor in phenylketonuric patients. Journal of Neurology 252:1328–1334
Ren-Zong Q (1989) Models of explanation and explanation in medicine. International Studies in the Philosophy of Science 3:199–212

Rizzi DA (1994) Causal reasoning and the diagnostic process. Theoretical Medicine and Bioethics 15:315–333

Sadegh-Zadeh K (2011) The logic of diagnosis. In: Gabbay DM, Gifford F, Thagard P, Woods J (eds) Philosophy of Medicine, Handbook of the Philosophy of Science, vol 16, North-Holland, Amsterdam, chap The Logic of Diagnosis, pp 357–424

Schaffner KF (1993) Discovery and Explanation in Biology and Medicine. Science and its conceptual foundations, University of Chicago Press

Schrenk MA (2007) The Metaphysics of Ceteris Paribus Laws. Ontos, Heusenstamm

Scriver CR (2007) The pah gene, phenylketonuria, and a paradigm shift. Human Mutation 28(9):831–845

Scriver CR, Waters PJ (1999) Monogenic traits are not simple: Lessons from phenylketonuria. Trends in Genetics 15(7):267–272

Severinsen M (2001) Principles behind definitions of diseases : a criticism of the principle of disease mechanism and the development of a pragmatic alternative. Theoretical Medicine and Bioethics 22:319–336

Sober E (1979) Clinical Judgment: A Critical Appraisal, Reidel Publishing Company, Dordrecht, Boston, London, chap The Art and Science of Clinical Judgment: An Informational Approach, pp 29–44

Thagard P (1998) Explaining disease: Correlations, causes, and mechanisms. Minds and Machines 8:61–78

Thagard P (2000) How Scientists Explain Disease, 2nd edn. Princeton Univ. Pr., Princeton, NJ

Torres P (2009) A modified conception of mechanisms. Erkenntnis 71:233–251

Tretter F (2010) Systems Biology in Psychiatric Research. From High-Throughput Data to Mathematical Modeling. Wiley-VCH Verlag, Weinheim

Wakefield J (1992) The concept of mental disorder: On the boundary between biological facts and social values. American Psychologist 47:373–388

Whitbeck C (1977) Causation in medicine: The disease entity model. Philosophy of Science 44(4):619–637

Chapter 6
The Generalizations of Biology: Historical and Contingent?

Alexander Reutlinger

Abstract Several influential philosophers of biology have raised the claim that the generalizations of biology are *historical and contingent* (Beatty J (1995) The evolutionary contingency thesis. In: E. Sober (Ed.) (2006) *Conceptual issues in evolutionary biology* (pp. 217–247). Cambridge: MIT Press; Schaffner, K. (1993). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press; Rosenberg (*British Journal for Philosophy of Science, 52*(4): 735–760, 2001); Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon; Mitchell, S. D. (2009). *Unsimple truths: Science, complexity and policy*. Chicago: The University of Chicago Press). This claim divides into the following subclaims, each of which I will contest: *firstly*, biological generalizations are restricted to a particular space-time region. I argue that biological generalizations are universal with respect to space and time. *Secondly*, biological generalizations are restricted to specific kinds of entities, i.e., these generalizations do not quantify over an unrestricted domain. I will challenge this second claim by providing an interpretation of biological generalizations that do quantify over an unrestricted domain of objects. *Thirdly*, biological generalizations are contingent in the sense that their truth depends on special (physically contingent) initial and background conditions. I will argue that the contingent character of biological generalizations does neither diminish their explanatory power nor is it the case that this sort of contingency is exclusively characteristic of biological generalizations.

Keywords Evolutionary contingency thesis • Laws of nature • Biological generalizations • Universality • Ceteris paribus laws

A. Reutlinger (✉)
Philosophisches Seminar, DFG Research Group "Causation and Explanation",
Universität zu Köln, Richard-Strauss-Str. 2, 50931 Köln, Germany
e-mail: Alexander.Reutlinger@uni-koeln.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_6,

6.1 Introduction: The Universality of Laws

Many philosophers of biology are convinced that there are important differences between (fundamental) physics and the biological sciences. One salient way in which biology is, according to these philosophers, unlike physics concerns the features of generalizations that play an epistemic role in the scientific practice of these disciplines. It is a majority view in philosophy of biology that (fundamental) physics states universal and exceptionless laws, while the biological sciences rely on nonuniversal and physically contingent generalizations (cf. Beatty 1995; Schaffner 1993; Rosenberg 2001; Mitchell 2002, 2009; Craver 2007).[1] This majority view in the philosophy of biology converges with the results of the debate on ceteris paribus laws since the mid-1990s: generalizations in the special sciences (such as neuroscience, psychology, sociology, economics, medical science, and the biological sciences) have different features than the laws of (fundamental) physics (cf. Reutlinger et al. 2011 for a survey).[2]

In this chapter, I will agree with these philosophers that the dynamical laws in fundamental physics and the laws in the special sciences differ in the way they describe.[3] However, despite the differences between laws in fundamental physics and generalizations in the special sciences (including biology), most philosophers believe that, in physics as well as in the special sciences, laws are important because they are statements used to explain and to predict phenomena, they provide knowledge how to successfully manipulate the systems they describe, and they support counterfactuals. Statements that are apt to play these roles in the sciences I call *lawish.* Similarly, Mitchell (1997, 2000) characterizes generalizations in the biological sciences (and in the special sciences in general) as "pragmatic laws" in virtue of performing at least one of these roles.

So, one might wonder what exactly the target of philosophers of biology is, who stress differences between the features of generalizations in fundamental physics and in the biological sciences. Philosophers of biology are worried that logical-empiricist views have created certain philosophical prejudices about how we think about laws of nature (e.g., Beatty 1995; Mitchell 2009). In the early debate on

[1] A terminological clarification: my focus is on *law statements* rather than on laws themselves. My aim is not to argue for any particular metaphysical claim (such as a regularity view and a dispositionalist account).

[2] Cf., for instance, Earman and Roberts (1999), Earman et al. (2002), Lange (2000), Loewer (2008), Roberts (2004), Woodward (2003, 2007), Maudlin (2007), Strevens (2009), and Reutlinger (2011).

[3] Many of the problems I will discuss in this chapter would be even trickier if one disagreed with the majority view in philosophy of biology and in the debate on ceteris paribus laws at this point. Some philosophers (e.g., Cartwright 1983, 1989; Mumford 2004) believe that even fundamental physics deals (at least in part) with nonuniversal laws. However, this would rather encourage the debate in philosophy of biology: if this were the case, the issue of nonuniversal laws might turn out to be even more pressing.

laws of nature, empiricist philosophers of science believed that *lawlikeness* is the crucial concept in order to find out which statements are law statements and which are not. Most importantly for our purposes, lawlikeness is commonly associated with universality (cf. Braithwaite 1959, p. 301). Philosophers of biology argue that the logical-empiricist view is a *philosophical prejudice* that ought to be overcome because it has been developed by focusing exclusively on physics while ignoring the biological sciences and other special sciences. It is simply false to believe that the generalizations of the latter scientific disciplines are universal.

By contrast to lawlikeness, I use "lawish" in the following way: a general statement is lawish if it is of explanatory and predictive use, successfully guides manipulation, and supports counterfactuals. Contrary to the traditional understanding of laws, being lawish does neither require universality nor other characteristic features of fundamental physical laws (such as the feature of satisfying symmetry principles). It is a matter of convention whether one would still want to use the term "law" for nonuniversal (i.e., not lawlike) general statements.[4] In other words, whether you want to refer to lawish statements by the honorific term "law" is merely a *verbal issue and not an interesting philosophical problem*. One can either use a new term for lawish, nonuniversal explanatory, or general statements. For instance, Woodward and Hitchcock (2003) introduce the concept of an explanatory generalization. Or, as I maintain in this chapter, one can insist that if a statement plays a lawish role, then it shares sufficiently many properties with universal laws in order to be called a law. Christopher Hitchcock and James Woodward admit that their account may be read as a *reconceptualization* of lawhood (cf. Woodward and Hitchcock 2003, p. 3). In order to avoid a fruitless quarrel about verbal issues, my strategy in this chapter will be to address two questions:

1. Are the laws of biology nonuniversal – and, if so, in which sense?
2. If the generalizations of biology are indeed in some sense nonuniversal, does this fact question their ability to play a lawish role?

Before I go on to answer these questions, let me provide a few examples of candidates for lawish generalizations in biology. The following five generalizations are classic examples in the debate on whether there are any laws of biology:

> *Mendel's law of segregation*: "In a parent, the alleles for each character separate in the production of gametes, so that only one is transmitted to each individual in the next generation." (Rosenberg and McShea 2008, p. 36)

[4]This is not to deny that the unique features of laws in physics are a topic of its own philosophical interest. Let me mention two questions of the greatest philosophical interest that are both related to the symmetry principles that constraint the law statements of physics: (a) how can we explain the existence of time-directed processes in a physical world that is governed by time-reversal invariant fundamental dynamical laws (cf. Albert 2000; Loewer 2008)? (b) Are symmetry principles laws? Are they empirical or a priori statements? Do they govern first-order laws (cf. Loewer 2009)?

Hardy-Weinberg law: "In an infinite, randomly mating population, and in the absence of mutation, immigration, emigration, and natural selection, gene frequencies and the distribution of genotypes remain constant from generation to generation." (Rosenberg and McShea 2008, p. 36; cf. Beatty 1995, p. 221)

The Krebs-cycle generalization: "In aerobic organisms, carbohydrate metabolism proceeds via a series of chemical reactions, including the eight steps of the Krebs cycle." (Beatty 1995, p. 219)

Bergmann's rule: "... given a species of warm-blooded vertebrates, those races of the species that live in cooler climates tend to be larger than those races of the species living in a warmer climates." (Beatty 1995, p. 224)

Allen's rule: "... given a species of warm-blooded vertebrates, those races of the species that live in cooler climates have shorter protruding body parts like bills, tails, and ears than those races of the species that live in warmer climates." (Beatty 1995, p. 224)

Recently, the debate has been enriched by a large number of interesting examples of lawish generalizations (cf. especially Lange 2000; Elgin 2006; Hamilton 2007; Raerinne 2011a, b). It is important to present a few of these example in order to prove the point that the above-listed classic examples of lawish generalizations are not an exceptional (and sometimes even outdated, no longer accepted) part of scientific practice in biology. Quite to the contrary, biology seems be full of lawish generalizations (which, admittedly, do not live up to the standard of lawlikeness):

The area law: "... the equilibrium number S of a species of a given taxonomic group on an island (as far as creatures are concerned) increases [polynomially][5] with the islands area [A]: $S = cA^z$. The (positive-valued) constants c and z are specific to the taxonomic group and island group." (Lange 2000, 235f)

The classic Lotka-Volterra model: "The classical Lotka-Volterra prey–predator model's equations are the following. Prey's growth equation is

$$dN_1/dt = rN_1 - bN_1N_2$$

Predator's growth equation is

$$dN_2/dt = ebN_1N_2 - cN_2$$

In the equations, r is the intrinsic growth rate of prey (in the absence of predation), c is the intrinsic death rate of predator (in the absence of their prey), b is the predation rate coefficient, e is predation efficiency, N_1 is the population size of prey at time t, and N_2 is the population size of predators at time t. These equations describe the dynamics in which populations of both prey and predators exhibit periodic oscillations." (Raerinne 2011a, p. 222)

The Volterra rule: "... any biotic or abiotic factor that both *increase* the death rate of predators and *decrease* the growth rate of their prey has the effect of *decreasing* the predator population size, whereas the population size of its prey *increases*." (Raerinne 2011a, p. 228)

Kleiber's rule: "... basal metabolism, an estimate of the energy required by an individual for the basic processes of living, varies as $aW^{0.75}$, where W is its body size [and a is a constant – A.R.]." (Raerinne 2011a, p. 219)

[5]Lange mistakenly writes "exponentially."

The exponential population growth model: "population growth is density independent, and it can be described by the equation

$$N_t = N_0 e^{rt},$$

where N_t is the population size at time t, N_0 is the initial size of the population, and r is the growth rate of the population, called the intrinsic rate of increase." (Raerinne 2011a, p. 212)

Mechanistic models: in the recent literature, the focus is on a large class of generalizations describing the steps in a mechanism such as the mechanism of photosynthesis, the LTP mechanism. (cf. Craver 2007)

Generalizations like these are believed to be *lawish*, although they are not universal generalizations.

So, why is it important to understand lawishness? One weighty reason stems from the conceptual connection of laws to causation and explanation. According to the empiricist interpretation, the most important feature of lawlikeness is *universality*. The idea to understand lawhood mainly in terms of universality has lead many theories of causation and explanation to rely on universal laws. This assumption turns out to be problematic: the central challenge for any theory of nonuniversal laws in the biological sciences is to account for their apparent lawish function (in the sense introduced above). If we are not able to provide an explication of nonuniversal laws, then (at least) the philosophy of biology faces a severe problem concerning causation and explanation in its domains. Many theories of causation and explanation in their *standard* form presuppose universal laws of nature (cf. Reutlinger 2011, p. 99 for a detailed discussion). The problem stemming from many theories of causation and explanation consists in a logical tension between three assumptions:

1. The biological sciences (a) refer to causes in their domains (i.e., some causal statements in biology are true) and (b) provide explanations in their domains.
2. It is a plain fact that the biological sciences – in contrast to physics – cannot rely on universal laws.[6]
3. Most philosophical theories of causation and explanation – in their standard form – essentially presuppose universal laws.

This tension can be formulated as the *nomothetic dilemma of causality and explanation* (cf. Pietroski and Rey 1995, p. 85; Woodward and Hitchcock 2003, p. 2):

[6]Cf. Earman et al. 2002, 297f), Woodward 2002, p. 303), and Roberts 2004). As noted above, Cartwright 1983, 1989) and Mumford 2004) dispute the claim that paradigmatic laws of physics conform to the received philosophical picture e.g., being universal). However, they do not deny that laws in the special sciences are nonuniversal, have exceptions, etc.

First horn: If it is a plain fact that the biological sciences cannot rely on universal laws (assumption 2) and if most philosophical theories of causation and explanation essentially involve universal laws and we do not reject these theories (assumption 3), then there is no causation and also no explanation in the biological sciences (negation of assumption 1).

Second horn: If there is causation and also explanation in the biological sciences (assumption 1) and if it is a plain fact that the biological sciences cannot rely on universal laws (assumption 2), then there is causation and explanation that does not involve universal laws (negation of assumption 3), i.e., we have to reject the above-listed theories of causation and explanation in their standard form.

If we do not want to give up the immensely plausible opinion that the biological sciences refer to causes and provide explanations (assumption 1) *for purely philosophical reasons*, then we are in need of a theory of nonuniversal lawish generalizations.

In this chapter, I will proceed as follows: in Sect. 6.2, I will provide several alternative meanings of the ambiguous concept of universality. I suggest that the claims made by philosophers of biology about the nonuniversality of lawish statements ought to be distinguished into three claims: *first*, the lawish statements are restricted to a space-time region. *Second*, the lawish statements are restricted to specific kinds of entities. *Third*, the lawish statements are true only if special physically contingent initial and background conditions obtain. In Sect. 6.3, I argue against the claims that lawish generalizations are historical in the sense that they are restricted to a specific spatiotemporal region and specific kinds of entities. In Sect. 6.4, I question the view that the feature of contingency undermines the lawish character of a statement. I argue for this claim by showing that the feature of contingency is compatible with four standard accounts of laws in the special sciences (i.e., completer, normality and statistical, invariance, and dispositionalist theories). In Sect. 6.5, I summarize the results of the preceding sections. I conclude with an outlook on future research concerning the features of laws describing biological complex systems.

6.2 What Is Universality?

As stated in the introduction, many philosophers of biology believe that the lawish generalizations of biology are – unlike the laws of fundamental physics – not universal. But what does it mean to be *universal* and, respectively, to be *nonuniversal*? It is an astonishing fact that this question is seldom answered in a

systematic way.[7] The lack of a systematic approach is a serious problem, because universality is an ambiguous concept. In accord with Andreas Hüttemann (2007, pp. 139–141), we may distinguish four *dimensions* of universality with respect to a law statement:

1. *First dimension – universality of space and time*: Laws are universal_1 iff they hold for all space-time regions.
2. *Second dimension – universality of domain of application*: Laws are universal_2 iff they hold for all (kinds of) objects.
3. *Third dimension – universality for external circumstances*: Laws are universal_3 iff they hold under all external circumstances, i.e., circumstances that are not referred to by the law statement itself. One useful way to interpret Hüttemann's reference to external conditions is to say that laws are true for *all initial* and *background conditions* of the system whose behavior is described by the law.
4. *Fourth dimension – universality with respect to the values of variables*: Laws are universal_4 iff they hold for all possible values of the variables[8] in the law statement. Universality in this sense acknowledges that laws are usually quantitative statements (and, thus, the predicates contained in these statements are to be conceived as variables ranging over a set of possible values).

Paradigm examples of fundamental physical laws (such as Newton's laws (supposing that they are true), Einstein's field equations, and the Schrödinger equation) are usually taken to be universal in all four dimensions (cf. Schurz 2002, Sect. 6.1; Hüttemann 2007, pp. 139–141). One might add the fifth dimension of universality to Hüttemann's list (to which I will return in Sect. 6.5):

[7] Mitchell (2000), Schurz (2002), Hüttemann (2007), and Reutlinger (2011) are notable exceptions.

[8] A variable X (in the terminology of statistics and causal modeling) is a function X:D ran(X), with a domain D of possible outcomes, and the range ran(X) of possible values of X. For quantitative variables X, ran(X) is usually taken to be the set of real numbers (cf. Pearl 2000; Eagle 2010, Chap. 0.9). For example, temperature is represented by a variable T that has several possible values such as $T = 30.65°$. However, in the debate on causation, philosophers often use qualitative, binary variables with $\text{ran}(X) = \{0; 1\}$ – whether a binary variable takes one of its values is taken to represent whether or not a certain type of event occurs (cf. Hitchcock 2001). On notation: capital letters, such as X, Y, …, denote variables; lowercase letters, such as x, y, …, denote values of variables; the proposition that X has a certain value x is expressed by a statement of the form $X = x$, i.e., $X = x$ is a statement about an event-type (cf. Woodward 2003).

5. *Fifth dimension – universality with respect to material constitution*: the same macro-behavior can be realized by microscopically different systems (cf. Batterman 2002; Hüttemann 2009). That is, a law describing the macro-behavior of a system is universal$_5$ if the generalization describing the macro-behavior is true for all microscopic realizers of the system in question. This concept of universality is a technical term in physics.

The crucial question in this chapter is which dimension of universality is at stake when philosophers of biology claim that the lawish generalizations of their discipline are nonuniversal. Philosophers of biology seem to refer to several dimensions of (non)universality. Hence, we need to disambiguate their claims. I think it is a fair reconstruction to say that three claims, with respect to three dimensions of universality, prevail in the debate:

1. *Historicity claim I:* The lawish generalizations of biology are historical because they are spatiotemporally restricted (cf. Rosenberg 2001, pp. 755–758). That is, the generalizations are nonuniversal$_1$.
2. *Historicity claim II:* The lawish generalizations of biology are historical because they are restricted to certain kinds of objects that exist in a limited space-time region (cf. Rosenberg 2001, pp. 755–758). In other words, the generalizations do not have the feature of being universal$_2$.
3. *Contingency claim:* The lawish generalizations of biology are true only if (a) certain physically contingent initial and background conditions C obtain and (b) these conditions C lead to the evolution of those biological entities that the biological generalizations in question describe (cf. Beatty 1995, 218f). I interpret Beatty's influential evolutionary contingency claim as a special case of nonuniversality$_3$: lawish generalizations in biology are true only if specific initial and boundary conditions obtain.

In Sect. 6.3, I will argue that we can easily reject *historicity claim I* and *historicity claim II*. Hence, the lawish generalizations of biology can indeed be regarded as universal$_1$ and universal$_2$. In Sect. 6.4, I will agree with most philosophers of biology that the lawish generalizations are true only if certain physically contingent initial and background conditions obtain. However, I will argue that this kind of contingency does not prevent generalizations to play a lawish role.

Before I take up the tasks of Sects. 6.3 and 6.4, I will briefly add three disclaimers.

First, some philosophers of biology have observed that certain alleged examples of laws in biology are no longer part of the presently established and accepted biological knowledge. For instance, Rosenberg argues that Mendel's laws have been

replaced by more accurate models (Rosenberg 2001, p. 747). I certainly do not have a problem with rejecting examples of law statements for the reason that they are no longer part of the widely accepted biological knowledge. However, merely pointing out that Mendel's laws have been replaced does not provide any threat to someone who claims that the lawish generalizations are nonuniversal$_3$.

Second, a standard objection to lawish generalizations is that they are either *false* (because disturbing factors occur) or *trivially true* (if the generalization is qualified by a ceteris paribus clause stating "if disturbing factors are absent . . . "). This dilemma is known as Lange's dilemma[9] in the debate on ceteris paribus laws (cf. Reutlinger et al. 2011, Sect. 4). Lange's dilemma is usually accepted as a challenge because philosophers expect and many scientists seem to believe lawish statements in biology (and laws in physics) to be *true and empirical* – and not trivially true – statements.[10] In the past two decades, various strategies have been developed in order to avoid Lange's dilemma (e.g., by Pietroski and Rey 1995; Lange 2002; Schurz 2002; Woodward 2003; Woodward and Hitchcock 2003; Maudlin 2007; Reutlinger 2011; Strevens 2012; Hüttemann manuscript). In this chapter, I will take it as a *premise* that one of these strategies is able to deal with Lange's dilemma. In any case, I believe that the question "*Can the generalizations of biology be contingent in Beatty's sense and still play a lawish role*?" ought to be distinguished from the challenge posed by Lange's dilemma.[11]

Third, a common complaint related to Lange's dilemma is this one: in the case of biological systems, background factors change over time – they are not constant as the literal reading of "ceteris paribus" suggests (cf. Beatty 1995; Rosenberg 2001). However, in the debate on ceteris paribus laws, several solutions have been proposed how to account for situations in which other conditions are not equal (cf. Cartwright 1983, 1989; Pietroski and Rey 1995; Maudlin 2004; Reutlinger 2011; Hüttemann manuscript). I will bracket this issue here by assuming that the problem of changing background conditions can be separated from the historicity claims I and II and the contingency claim.

[9]Named after Marc Lange (cf. Lange 1993, p. 235).

[10]Not everyone agrees: one option to avoid Lange's dilemma is to reject the assumption giving rise to Lange's dilemma. That is, to reject the claim that lawish statements are true and empirical statements. Following this line of reasoning, some philosophers have argued that biological generalizations lack empirical content and should be interpreted as a priori *truths* (cf. Sober 1997; Elgin 2003). This might be a fall-back option, but it cannot be the first choice, in my view, because most of the examples above clearly appear to be empirical statements.

[11]As Raerinne (2011b) points out, even if special initial and background conditions are necessary for a generalization to hold, the fact that these conditions obtain is not sufficient for the generalization to be true. If it is the case that *not all* initial and background conditions are considered (and this seems to be Beatty's claim), then disturbing factors might still occur.

6.3 Against the Alleged Historical Character of Biological Generalizations

Are the lawish generalizations of biology universal$_1$ and universal$_2$? I think the answer is yes. Being universal$_1$ and universal$_2$ are features that the lawish generalizations of biology and the laws of physics have in common. My answer is in conflict with Rosenberg's historicity claims I and II (see Sect. 6.2). Contrary to Rosenberg, I will argue for two claims: first, lawish generalizations in the biological sciences hold for *all* space-time regions (i.e., they are universal$_1$). This kind of universality allows that these generalizations simply lack an application in some space-time regions. Secondly, lawish statements can be formalized such that they quantify over an *unrestricted* domain of objects (if so, they are universal$_2$).

Arguing for these claims might not seem plausible at first glance, because generalizations in the biological sciences are usually interpreted as system laws.[12] Gerhard Schurz (2002, Sect. 6.1) introduces the notion of system laws as follows: while fundamental physical laws "are not restricted to any special kinds of systems (be it by an explicit antecedent condition or an implicit application constraint)" (Schurz 2002, p. 367), system laws refer to particular systems of a certain (biological, psychological, social, etc.) kind K in a specific space-time region. Hence, so the usual characterization continues, lawish statements in the special sciences typically have an *in-built historical dimension* which the fundamental physical laws lack, because they are restricted to a limited space-time region where the objects of a certain kind K exist (for instance, cf. Beatty 1995; Rosenberg 2001). I will argue that Schurz is absolutely correct in characterizing lawish statements in the biological sciences as being "restricted to [. . .] special kinds of systems (be it by an explicit antecedent condition or an implicit application constraint)" (Schurz 2002, p. 367). However, if one adopts Schurz's characterization of generalizations in biology as system laws, then one is still entitled to believe that these statements are universal$_1$ and universal$_2$. Let me explain why I think Schurz's interpretation of biological generalizations as system laws differs from Rosenberg's spatiotemporally restricted laws. I will argue for this claim in two steps: first, I will argue for the universality$_1$ of lawish statements and then for their universality$_2$.

6.3.1 Argument for Universality$_1$

Does Schurz's characterization of system laws imply that the generalizations of biology are nonuniversal$_1$? No. Simply because a generalization G does not have an application in some space-time region s, that does not mean that the law does not

[12] Cf. Cartwright (1983, Essay 6) for a similar notion of a phenomenological law.

hold at *s*. In order to be truly nonuniversal$_1$, G would have to conform to a thought experiment of "Smith's Garden" by Tooley:

> All the fruit in Smith's garden at any time are apples. When one attempts to take an orange into the garden, it turns into an elephant. Bananas so treated become apples as they cross the boundary, while pears are resisted by a force that cannot be overcome. Cherry trees planted in the garden bear apples, or they bear nothing at all. If all these things were true, there would be a very strong case for its being a law that all the fruit in Smith's garden are apples. And this case would be in no way undermined if it were found that no other gardens, *however similar to Smith's garden in all other respects*, exhibited behaviour of the sort just described. (Tooley 1977, p. 686, my emphasis)

According to Tooley, a law L can be spatiotemporally restricted to a space-time region *s* (as the laws in Smith's garden) in the sense that L fails to be true in a situation that is *perfectly similar* to the situation in *s*, except for the fact that this perfectly similar situation is located in a different space-time region *s** (cf. Earman 1978).

I think the generalizations of biology that are truly nonuniversal$_1$ would be similar to the laws that hold for various fruit in Smith's garden. But it seems to be a too strong claim that laws in the biological generalizations are local in the same way as the laws in Smith's garden are. It seems to be a more promising option to say that (a) biological generalizations are universal$_1$ and (b) these generalizations simply lack application in some space-time regions. For instance, Bergmann's rule, the classic Lotka-Volterra model and Mendel's law of segregation do not hold on Mars because there are neither warm-blooded vertebrates nor anything standing in a predator–prey relation, nor cells with alleles. However, this situation does not indicate that Bermann's rule, the classic Lotka-Volterra model, and Mendel's law of segregation are *local* laws – as the laws of Smith's garden are. A better understanding seems to be that these statements happen to have no application on Mars (e.g., if there are no warm-blooded vertebrates on Mars, then the conditions of application for Bergmann's rule are not satisfied; cf. Strevens 2012, Sect. 3). To illustrate my claim in another way, consider the following scenario: suppose we were to find a space-time region *s* that is in biological aspects perfectly isomorphic to Earth (including certain physically contingent initial and background conditions) – that is, the only difference between life on Earth and life in this region *s* is the spatiotemporal location. Suppose further we were to discover that none of the generalizations of current terrestrial biology is true in region *s*. Would we not demand an explanation for this local inapplicability? It is precisely this demand for an explanation that reveals the intuition that Bergmann's rule is quite dissimilar to the laws of Smith's garden.

6.3.2 Argument for Universality$_2$

Does the characterization of lawish statements in the biological sciences as system laws imply that these statements are nonuniversal$_2$? No, it does not. At first glance, biological generalizations, if viewed as system laws, appear to be nonuniversal$_2$:

special science laws quantify over a restricted domain of objects of a certain kind – *not* over a domain of objects of *all* kinds. For instance, consider *Bergmann's rule* once more: "given a species of warm-blooded vertebrates, those races of the species that live in cooler climates tend to be larger than those races of the species living in a warmer climates" (Beatty 1995, p. 224). Bergmann's rule seems to be restricted to warm-blooded vertebrates – it does not make any claim about electrons, atoms, neurons, rational agents, markets, etc. One might get the idea that generalizations of biology refer to a *restricted* domain D that is a proper subset of the domain X of all things. Bergmann's rule can be formalized as quantifying over a restricted domain D of warm-blooded vertebrates (with d as an individual variable of domain D):

$\forall(d)$((lives in cooler climates)$d \rightarrow$ (tends to be larger than those races of the species living in a warmer climates)d).

But is this really a convincing reconstruction of lawish statements in the special sciences? I can provide an alternative formalization that quantifies over the domain of *all* objects. This formalization interprets the kind of object (here: warm-blooded vertebrates) as a *predicate* and not as a restriction of the domain. In the alternative formalization, x is an individual variable for the unrestricted domain X:

$\forall(x)$((is a member of a species of warm-blooded vertebrates)$x \wedge$ (lives in cooler climates)$x \rightarrow$ (tends to be larger than those races of the species living in a warmer climates)x).

The alternative, unrestricted formalization of Bergmann's rule is a way to save universality$_2$. By formalizing lawish generalizations in this way, I provide a reason to reconstruct them as generalizations quantifying over all kinds of objects.[13]

This is not a trivial result at all, because philosophers of biology, such as Beatty (1995) and Rosenberg (2001), insist that generalizations in the biological sciences should be regarded as (a) being historical in the sense of applying only to a specific space-time region (this is in contradiction with universality$_1$) and (b) as referring to a restricted domain of objects (this contradicts universality$_2$). Contrary to these philosophers, I want to emphasize that one *can* maintain that lawish generalizations

[13]One might want to dispute the claim that even the fundamental laws do not apply to *everything* (contra Schurz 2002; Hüttemann 2007). One objects that the fundamental laws, for instance, do not apply to angels and numbers. However, I think that, even if this were the case, we could preserve the universality$_2$ for the fundamental laws by exactly the same strategy which I just used for preserving universality$_2$ for lawish statements in the special sciences. Further, my arguments do not have to rely on the characterization of fundamental physical laws which Schurz and Hüttemann provide.

in the biological sciences are universal$_1$ and universal$_2$. In other words, the lawish generalizations do not differ from the fundamental laws of physics with respect to the first and the second dimension of universality.

6.4 The Case for Nonuniversal$_3$ Generalizations

In Sect. 6.2, I interpreted Beatty's evolutionary contingency thesis as a special case of nonuniversality$_3$: lawish generalizations such as Allen's rule, the Volterra rule, and the exponential growth model hold *only if* very specific initial and background conditions obtain. Allen's rule, the Volterra rule, and the exponential growth model do not hold under all (physically) possible initial and background conditions. This is why I interpret these lawish generalizations as being nonuniversal$_3$. There is good evidence for the view that the biological sciences are not exceptional in postulating contingent laws. Physically contingent lawish generalizations are of importance in the physical sciences as well. Let me provide a famous example from the physical sciences: the second law of thermodynamics (for short, the second law). The second law is a nonfundamental physical law. The second law is usually taken to play a role in physical explanation, prediction, and manipulation – i.e., it performs a lawish role. The standard formulation of the second law is:

> The total entropy of the world (or of any isolated subsystem of the world), in the course of any possible transformation, either keeps at the same value or goes up. (Albert 2000, p. 32)

Craig Callender provides an example as an illustration of the second law:

> Place an iron bar over a flame for half an hour. Place another one in a freezer for the same duration. Remove them and place them against one another. Within a short time the hot one will 'lose its heat' to the cold one. The new combined two-bar system will settle to a new equilibrium, one intermediate between the cold and hot bar's original temperatures. Eventually the bars will together settle to roughly room temperature. (Callender 2006)

It is a majority opinion that an explanation of why the second law obtains has to require more than just the fundamental laws of physics. According to a tradition originating in the work of Ludwig Boltzmann, one has to rely on physically contingent initial conditions – among other things – in order to explain why macroscopic physical systems conform to the second law. An influential proposal for such an initial condition is the so-called past hypothesis, i.e., the claim that the initial macro state of the universe (or an isolated subsystem thereof) was a state of low entropy (cf. Albert 2000, p. 96; Loewer 2007, pp. 298–304, 2009, pp. 156–158). The upshot of the Boltzmannian explanation of the second law is as follows: the second law is a lawish statement which is true only if special initial conditions (expressed by the past hypothesis) obtain – and these special initial conditions are a physically contingent fact with respect to the fundamental dynamical laws of physics.[14]

[14]Cf. Roberts (2008) and Strevens (2008) for further examples of physically contingent lawish statements.

The question I would like to answer in this section is the following: *If the generalizations of biology are indeed nonuniversal$_3$, does this fact undermine their ability to play a lawish role*? I will provide arguments for the following answer: no, a generalization might be nonuniversal$_3$ and lawish at once. I will argue for this claim by showing that several standard theories of lawish statements (or ceteris paribus laws) are consistent with the fact that the truths of some lawish statements depend on whether special initial and background conditions obtain (cf. Reutlinger et al. 2011 for a survey of these and other accounts of ceteris paribus laws).

6.4.1 Completer Accounts

The basic idea of completer approaches consists in regarding lawish generalizations in the biological sciences – such as Bergmann's rule, the area law, and the Volterra rule – to be *incomplete as they stand*. The generalizations are *completed* by adding missing conditions to the antecedent of the law statement. The guiding thought is that the completed antecedent implies the consequent of the lawish statement. Jerry Fodor motivates the completer account of laws in the special sciences (including the biology) as follows:

> Exceptions to the generalizations of a special science are typically inexplicable from the point of view of (that is, in the vocabulary of) that science. That's one of the things that makes it a *special* science. But, of course, it may nevertheless be perfectly possible to explain the exceptions *in the vocabulary of some other science*. [...]. On the one hand the [special sciences'] ceteris paribus clauses are ineliminable from the point of view of its propriety conceptual resources. But, on the other hand, we have – so far at least – no reason to doubt that they can be discharged in the vocabulary of some lower-level science (neurology, say, of biochemistry; at worst physics). (Fodor 1974, p. 6)

Fodor's idea is that the additional, completing factors whose existence is required by the ceteris paribus clause cannot be entirely specified within the conceptual resources of, for instance, biology. However, the completion can (at least in principle) be achieved within the vocabulary of some fundamental science, such as neurophysiology or physics. A physical microdescription of the antecedent condition A is called a realizer of A (the same A may have several different realizers). Fodor (1991, p. 23) defines the completer more precisely:

> A factor C is a completer relative to a realizer R of A and a consequent predicate B iff:
>
> (i) R and C is strictly sufficient for B.
> (ii) R on its own is not strictly sufficient for B.
> (iii) C on its own is not strictly sufficient for B.

Based on this notion of a completer, Fodor defines the truth conditions of a cp law as follows:

"cp(A→B)" is true iff for every realizer R of A there is a completer C such that (A∧C)→B.[15]

The crucial question for my purposes is whether the completer approach is compatible with lawish generalizations that have the feature of being nonuniversal$_3$. The answer is yes, I believe. The natural place for listing the specific physically contingent initial and background conditions – that Beatty (1995) emphasizes – is the completer condition C. For instance, in the case of Allen's rule, the completer consists of certain physically contingent initial and background conditions without which a species of warm-blooded vertebrates that live in cool climates would not have evolved. It is a controversial matter whether adding the evolutionary history to the antecedent of the lawish generalization is strictly *sufficient* for the truth of the consequent of the law statement (cf. Sober 1997 and Elgin 2006 versus Raerinne 2011b).[16] However, what matters most for the problem that this chapter is concerned with is that there is nothing in the completer account itself which prevents lawish generalizations from being dependent on specific initial and background conditions.

6.4.2 Normality and Statistical Accounts

The main idea of normality theories consists in advocating the following truth conditions for laws in the biological sciences: Allen's rule is a true lawish generalization iff it is *normally* the case that, given a species of warm-blooded vertebrates, those races of the species of warm-blooded vertebrates that live in cooler climates have shorter protruding body parts like bills, tails, and ears than those races of the species that live in warmer climates (cf. Reutlinger et al. 2011, Sect. 8). Schurz (2001, 2002, §5) analyzes lawish statements in biological sciences as *normic* laws of the form "As are normally Bs." Schurz explicates normality in terms of a high probability of the consequent predicate, given the antecedent predicate, where the underlying conditional probabilities are objective statistical probabilities. According to the *statistical consequence thesis*, normic laws imply numerically unspecified statistical generalizations of the form "Most As are in fact Bs," by which they can be empirically tested.

Is it thus compatible with the normality account that the truth of lawish statements of biology depends on specific physically contingent initial and background conditions? Here the answer is also positive: normality statements can have a complex antecedent which lists further conditions. In analogy with the completer approach, these conditions might include those physically contingent conditions without which – in the case of Allen's rule – warm-blooded vertebrates would not have evolved in a cool climate.

---

[15]Cf. Pietroski and Rey (1995), Maudlin (2007), and Reutlinger (2011) for variants of the completer account.

[16]This controversy is concerned with Lange's dilemma which I will not address in this chapter.

An analogous strategy can be applied to the statistical approach to lawish generalizations proposed by Earman and Roberts (1999). Their view is closely related to Schurz's normic account. According to Earman and Roberts, a typical special science generalization "asserts a certain precisely defined statistical relation among well-defined variables" (Earman and Roberts 1999, p. 467). That is, special science laws are *statistical generalizations* of the following form: "in population H, a variable P is positively statistically correlated with variable S across all sub-populations that are homogeneous with respect to the variables $V_1, \ldots, V_n$" (Earman and Roberts 1999, p. 467). The obvious place to mention the physically contingent conditions without which, for instance, warm-blooded vertebrates would not have evolved in a cool climate are the variables $V_1, \ldots, V_n$. It is worth pointing out a genuine feature of normic and statistical accounts: unlike in the case of completer accounts, it is not the case that a proponent of the statistical and the normic account claims that the antecedent of the lawish statement is sufficient for the consequent.

Moreover and most likely in agreement with Beatty and Rosenberg, Schurz (2001) defends the statistical consequence thesis by appealing to an *evolution-theoretic argument*.[17] Schurz argues that evolutionary systems are self-regulatory systems whose self-regulatory properties have been gradually selected according to their contribution to reproductive success. He claims that the temporal persistence of self-regulatory systems is governed by a certain range of prototypical norm *states*, in which these systems constantly have to stay in order to keep alive. According to Schurz, regulatory mechanisms compensate for disturbing influences of the environment. Although the self-regulatory capacities of evolutionary systems are the product of a long adaptation history, they are not perfect. Some organisms may be dysfunctional, and their normic behavior may have various *exceptions*. However, Schurz claims that it has to be the case that these systems are in their prototypical norm states in the *statistical majority* of cases and times. Otherwise, these systems would not have *survived* in evolution.

The upshot of this discussion is that the normality account is not merely compatible with nonuniversality$_3$. In fact, one of its main proponents, Gerhard Schurz, even provides an evolution-theoretic argument in favor of the account. If Schurz's argument is sound, then it implies that normic laws are a direct result of biological evolution.

[17]One of the problems of Schurz's approach arises as soon as one starts to apply his theory of normic laws to nonbiological (e.g., economic) examples. His argument is based on a *generalized theory of evolution* which does not only apply to biological evolution but also to cultural evolution. The common domain of the life sciences (which, according to Schurz, include biology, psychology, as well as the social sciences and the humanities) are evolutionary systems or their products. One might worry, though, whether such a *generalized theory of evolution* is sufficiently confirmed.

6.4.3 *Invariance Accounts*

In accord with invariance theories, the distinctive feature of lawish generalization is their invariance. Invariance is the feature that separates lawish and accidentally true generalizations. A generalization is invariant if it holds for some, possibly a limited, range of the possible values of variables figuring in the generalization. According to Woodward and Hitchcock (2003, p. 17) and Woodward (2003, p. 250), a statement *G* is minimally invariant iff the testing intervention condition holds for *G*. The testing intervention condition for a generalization G of the form $Y = f(X)$ states:

1. There are at least two different possible values of an endogenous variable X, x_1 and x_2, for each of which Y realizes a different value (y_1, y_2) in the way that the function *f* in G describes.
2. The fact that X takes x_1 or, alternatively, x_2 is the result of an intervention.

Take the *Volterra rule* as an example of an invariant generalization. According to Woodward and Hitchcock's account, the Volterra rule is minimally invariant if there is an intervention ("any biotic or abiotic factor") such that if the death rate of predators (counterfactually) *increases* and the growth rate of their prey (counterfactually) *decreases*, then the predator population size *decreases* and the population size of its prey *increases*.

Again, is the invariance account of lawish generalizations compatible with the contingency claim? Yes, it is. Invariance is defined relative to a set of variables (such as the death rate of predators and the population size) and a set of functions relating the variables (such as an increase-decrease function). An invariantist is free to embrace the view that biological entities (e.g., an ecosystem rabbits and foxes) to which these variables apply have evolved. And she is free to say that it is a physically contingent fact that biological entities of this kind have evolved. The crucial point for the advocate of an invariance account is this: given that certain entities of a kind K have evolved, the lawish generalizations about members of K are the invariant generalizations.

6.4.4 *Dispositionalist Accounts*

According to the dispositionalist account, a law statement is true if the type of system in question (i.e., those entities to which the law applies) has the disposition that the law statement attributes to the system (cf. Cartwright 1989; Hüttemann 1998; Bird 2005). For instance, the Krebs-cycle generalization states that aerobic organisms are the kind of system *disposed to* have a carbohydrate metabolism proceeding via a series of chemical reactions, including the eight steps of the Krebs

cycle. The manifestation of this disposition might be disturbed, but aerobic might still have the disposition for Krebs-cycle behavior. That is, dispositionalists reconstruct law statements as statements about dispositions, tendencies, and capacities, etc., rather than about overt behavior.[18] The claim is that certain kinds of systems have certain kinds of tendencies or dispositions.

Is it the case that the dispositionalist account is compatible with the claim that the lawish generalizations of biology are nonuniversal$_3$? We can provide a positive answer. The dispositionalist can happily accept that the dispositions of biological systems have evolved, and at the same time, he/she can maintain that lawish generalizations ought to be interpreted as claims about the dispositions of biological entities, such as aerobic organisms (cf. Hüttemann and Kaiser 2014 for a number of examples of biological dispositions).

What has been established in this section? I have started out by interpreting Beatty's evolutionary contingency claim as a special case of nonuniversality$_3$. Then I have pointed out that biology is not the only science that relies on generalizations that depend on physically contingent initial conditions (the second law is an example from physics). The main result of this section is that four standard theories of lawish statements in the special sciences (i.e., the completer account, the normality account, the invariance and the dispositionalist account) are compatible with the feature of nonuniversality$_3$. Thus, it might be the case that the generalizations of biology differ from the fundamental physical laws because the former are not true for all initial and background conditions (as Beatty and Rosenberg argue). However, this result need not impress us since the generalizations of biology might still play a lawish role. This result requires a qualification: these generalizations play a lawish role to the extent that discussed theories of special science laws can be integrated into theories of explanation, prediction, and manipulation. One can be optimistic about the prospects of a successful integration of lawish statements into theories of explanation because several recent theories of explanation do not require universal laws and rely on nonuniversal generalizations instead (cf. Woodward 2003; Craver 2007; Mitchell 2009; Strevens 2009).

Before concluding this section, I will add two disclaimers:

First, even if each of the standard accounts of lawish statements is compatible with nonuniversality$_3$, a proponent of these accounts still has to deal with other problems, in order to be convincing. The most pressing problems are (a) to avoid Lange's dilemma and (b) to distinguish lawish generalizations from accidentally true generalizations. However, I think that several convincing cases have been made in favor of each of the above-presented accounts (cf. Reutlinger et al. 2011; Reutlinger 2011; Strevens 2012).

Second, it is not the case that I have to accept that *every* generalization that is true of evolved biological entities can play a lawish role. In order to support this claim,

[18]The main motivations to adopt a dispositionalist theory consist in (a) having a strategy to avoid Lange's dilemma and (b) explaining why idealized laws can be applied in nonideal situations (cf. Reutlinger et al. 2011, Sect. 7).

I can rely on a distinction proposed by Waters (1998). Waters distinguishes two classes of generalizations about evolved entities: the first class of generalizations concerns the *architecture* of a biological entity, i.e., the way it is built (such as "all major arteries have thick layers of elastic tissues around them," "all birds have wings," and "all zebras have stripes"). The second class of generalizations describes how a biological entity changes over time. The lawish role seems to be primarily ascribed to members of the second class – the dynamical generalizations (or "causal" generalizations, as Waters refers to them). Let me put it more cautiously: it is at least not clear why I would have to accept that *all* architecture-generalizations do, in fact, play a lawish role in scientific biological practice. The epistemic role of architecture-generalizations might be limited to classifying systems of a certain kind (which is the product of evolution and which might also be described by a dynamical generalization).[19]

6.5 Conclusion and Outlook

What has been achieved in the preceding sections? In Sect. 6.1, I reconstructed a view held by many philosophers of biology: the generalizations occurring in the biological sciences differ from the fundamental laws of physics, as the former are *not universal*. But what exactly does universality amount to? In Sect. 6.2, I attempted to disambiguate "universal" by suggesting several alternative meanings of the concept of universality (cf. Hüttemann 2007; Batterman 2002). Based on these alternative meanings, I proposed to understand the claims made by philosophers of biology about the nonuniversality of lawish statements in the following ways: *first*, the lawish statements are restricted to a space-time region, i.e., the statements are nonuniversal$_1$. *Second*, the lawish statements are restricted to specific kinds of entities, i.e., the generalizations are nonuniversal$_2$. *Third*, the lawish statements are true only if very special physically contingent initial and background conditions obtain. I took this kind of contingency to be a special case of nonuniversality$_3$. In Sect. 6.3, I argued against the claims that lawish generalizations are historical in the sense that they are restricted to a specific spatiotemporal region and to specific kinds of entities. I opposed to nonuniversality$_1$ and to nonuniversality$_2$. The upshot is that lawish generalizations and the laws of physics resemble one another because they share the features of universality$_1$ and universality$_2$. In Sect. 6.4, I raised objections to the view that the feature of contingency somehow undermines the lawish character of a statement. I argued for this claim by showing that the

[19]This distinction might also be regarded as a defense of Schurz's normic approach to lawish generalizations, because Schurz's main example of a normic law is "normally, birds can fly." This is an unfortunate choice, I think, since the immediate response to this example is to deny that this statement plays a lawish role. Rather Schurz's example ought to be classified as an architecture-generalization. Schurz's account is strong when applied to dynamical generalizations.

feature of contingency is compatible with four standard accounts of laws in the special sciences. This compatibility suggests that a contingent generalization G of biology is lawish to the extent to which the presented standard accounts of laws in special sciences permit that G is used for explanatory and predictive purposes, that G guides manipulations, that G supports counterfactuals, etc. One significant result of this discussion was that it does not matter at all whether one is willing to call, for instance, Bergmann's rule or the exponential growth model a *law*.

Let me conclude with an outlook on future research. The observation that biological generalizations are only true provided the presence of certain initial and background conditions raises an important issue that is often neglected in the debate. Although the laws of biology *depend* on the presence and absence of *some* conditions, it is a highly nontrivial feature of these laws that they are also *independent* of many initial and background conditions. These *independence conditions* can be illustrated by generalizations describing the macro-behavior of biological complex systems. For my present purpose, a complex systems can be preliminarily understood as consisting of many parts (i.e., the micro-level of the system) whose (random) interactions alone – and not the influence of an external cause – result in ordered macro-behavior (cf. Strevens 2003; Ladyman et al. 2013). It is important to draw a distinction between two concepts of complexity: compositional and dynamic complexity (cf. Mitchell 2009; Kuhlmann 2011). The compositional reading refers to the spatial arrangement and to the number, as well as to the kinds of parts. The dynamic reading of complexity characterizes the temporal evolution of the systems on a macroscopic scale. An example of a biological dynamical complex system is an ecosystem whose parts consist of the members of various species inhabiting a particular territory. The classic Lotka-Volterra model, the Volterra rule, and the exponential growth model are examples of generalizations keeping track of how an ecosystem develops over time on a macroscopic scale. For instance, the Prey's growth equation in the classic Lotka-Volterra model describes how the prey population changes over time.

So, in what respect is the dynamical macro-behavior of a complex system *independent* of certain conditions? The macro-behavior of dynamically complex systems is typically *robust*. The macro-behavior is robust if it is invariant with respect to a range of changes in the initial micro-conditions and the background conditions (cf. Strevens 2003; Woodward 2007; Ladyman et al. 2013). Are the classic Lotka-Volterra model and the Volterra rule descriptions of robust macro-behavior? Certainly, these generalizations do not remain true for *all* possible initial and background conditions. However, this is not required by robustness. The macro-behavior is robust if it holds for *some* range changes in the initial micro-conditions and for the background conditions. One can be easily convinced that this range of changes exists: suppose that Volterra rule applies to a group of foxes and rabbits in an ecosystem. Do we think that the Volterra rule generalization would have been false if the rabbits number 12 and 42 had been placed 10 m west of their actual position at some initial time t_0? No, certainly not; we suppose that the Volterra rule is invariant under these changes of initial micro-conditions (cf. Strevens 2003, Chap. 1; Ladyman et al. 2013). The Volterra rule is robust under these – and possibly

more interesting – changes in the initial micro-conditions. However, the robustness of generalizations, such as the Volterra rule, is a nontrivial feature of a lawish generalization in biology.

It is a challenging task for future research to investigate two topics: *first*, Beatty (1995) and others have focused on the (physically contingent) conditions that have to obtain for biological generalizations to be true. *Second*, it has usually been neglected that there is another side of the coin: the truth of lawish generalizations in biology is also independent of certain initial and background conditions. I discussed the case of robustness, while Elgin (2006), Hamilton (2007), and Raerinne (2011b) focus on the invariance of scaling laws in biology (such as Kleiber's rule) with respect to many material micro-details about different species. That is, Elgin, Hamilton, and Raerinne focus on the *universality with respect to material constitution*, i.e., the same macro-behavior can be realized by microscopically different systems (see Sect. 6.1; cf. Batterman 2002). Two interesting question for future research could be: (i) If a generalization G is robust and invariant with respect to many material micro-details, do we then have a good (although not sufficient) reason to believe that G is not merely accidentally true but a statement that is able to play a lawish role? (ii) What is a good explanation for the surprising fact that many biological generalizations have the features of robustness and universality with respect to material constitution?

References

Albert, D. (2000). *Time and chance*. Cambridge: Harvard University Press.
Batterman, R. (2002). *The devil in the details*. Oxford: Oxford University Press.
Beatty, J. (1995). The evolutionary contingency thesis. In: E. Sober (Ed.), (2006) *Conceptual issues in evolutionary biology* (pp. 217–247). Cambridge: MIT Press.
Bird, A. (2005). The dispositionalist conception of laws. *Foundations of Science, 10*(4), 353–370.
Braithwaite, R. B. (1959). *Scientific explanation*. Cambridge: Cambridge University Press.
Callender, C. (2006). Thermodynamic asymmetry in time. In: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2011 Edition).http://plato.stanford.edu/archives/spr2011/entries/time-thermo/
Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.
Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford: Oxford University Press.
Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon.
Earman, J. (1978). The universality of laws. *Philosophy of Science, 45*(2), 173–181.
Earman, J., & Roberts, J. (1999). Ceteris paribus, there is no problem of provisos. *Synthese, 118*(3), 439–478.
Earman, J., Roberts, J., & Smith, S. (2002). Ceteris paribus lost. *Erkenntnis, 57*(3), 281–301.
Elgin, M. (2003). Biology and a priori laws. *Philosophy of Science, 70*(5), 1380–1389.
Elgin, M. (2006). There may be strict and empirical laws in biology, after all. *Biology and Philosophy, 21*(1), 119–134.
Fodor, J. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese, 28*(2), 97–115.

Fodor, J. (1991). You can fool some people all of the time, everything else being equal: Hedged laws and psychological explanations. *Mind, 100*(397), 19–34.
Hamilton, A. (2007). Laws of biology, laws of nature: Problems and (dis)solutions. *Philosophy Compass, 2*(3), 592–610.
Hüttemann, A. (1998). Laws and dispositions. *Philosophy of Science, 65*(1), 121–135.
Hüttemann, A. (2007). Naturgesetze. In A. Bartels & M. Stöckler (Eds.), *Wissenschaftstheorie* (pp. 135–153). Paderborn: Mentis.
Hüttemann, A. (2009). Physikalische Realisierung in der Physik. In M. Backmann & J. G. Michel (Eds.), *Physikalismus, Willensfreiheit, Künstliche Intelligenz* (pp. 67–73). Paderborn: Mentis.
Hüttemann, A. (manuscript). Ceteris paribus laws in physics. Manuscript submitted to Synthese
Hüttemann, A., & Kaiser, M. (2014). Dispositions in biology. In: K. Engelhardt & M. Quante (Eds.), *Oxford handbook of potentiality*. Dordrecht: Springer.
Kuhlmann, M. (2011). Mechanism in dynamically complex systems. In P. McKay Illary, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 880–906). New York: Oxford University Press.
Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science, 3*, 33–67.
Lange, M. (2000). *Natural laws in scientific practice*. New York: Oxford University Press.
Lange, M. (2002). Who's afraid of Ceteris Paribus Laws? Or: how i learned to stop worrying and love them. *Erkenntnis, 52*, 407–423.
Lange, M. (2009a). *Laws and lawmakers*. New York: Oxford University Press.
Lange, M. (2009b). Why is there anything except physics? *Synthese, 170*(2), 217–233.
Loewer, B. (2007). Counterfactuals and the second law. In H. Price & R. Corry (Eds.), *Causation, physics and the constitution of reality: Russell's republic revisited* (pp. 293–326). Oxford: Clarendon.
Loewer, B. (2008). Why there is anything except physics. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation and causation* (pp. 149–163). Oxford: Oxford University Press.
Loewer, B. (2009): Why is there anything except physics? *Synthese, 170*(2), 217–33.
Maudlin, T. (2007). *The metaphysics within physics*. Oxford: Oxford University Press.
Mitchell, S. D. (1997). Pragmatic laws. *Philosophy of Science, 64*(4), 242–265.
Mitchell, S. D. (2000). Dimensions of scientific law. *Philosophy of Science, 67*(2), 468–479.
Mitchell, S. D. (2002). Ceteris paribus – An inadequate representation of biological contingency. *Erkenntnis, 52*(3), 329–350.
Mitchell, S. D. (2009). *Unsimple truths: Science, complexity and policy*. Chicago: The University of Chicago Press.
Mumford, S. (2004). *Laws in nature*. London: Routledge.
Pietroski, P., & Rey, R. (1995). When other things aren't equal: saving Ceteris Paribus Laws from vacuity. *British Journal for the Philosophy of Science, 46*, 81–110.
Raerinne, J. (2011a). *Generalizations and models in ecology: Lawlikeness, invariance, stability and robustness*. Academic dissertation, University of Helsinki, https://helda.helsinki.fi/handle/10138/24583
Raerinne, J. (2011b). Allometries and scaling laws interpreted as laws: A reply to Elgin. *Biology and Philosophy, 26*(1), 99–111.
Reutlinger, A. (2011). A theory of non-universal laws. *International Studies in the Philosophy of Science, 25*, 97–117.
Reutlinger, A., Hüttemann, A., & Schurz, G. (2011). Ceteris Paribus laws. In: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2011 Edition). http://plato.stanford.edu/archives/spr2011/entries/ceteris-paribus/
Roberts, J. (2004). There are no laws in the social sciences. In C. Hitchcock (Ed.), *Contemporary debates in the philosophy of science* (pp. 168–185). Oxford: Blackwell.
Roberts, J. (2008). *The law-governed universe*. New York: Oxford University Press.
Rosenberg, A. (2001). How is biological explanation possible? *British Journal for Philosophy of Science, 52*(4), 735–760.

Rosenberg, A., & McShea, D. (2008). *Philosophy of biology: A contemporary introduction*. London: Routledge.
Schaffner, K. (1993). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press.
Schurz, G. (2001). Pietroski and Rey on *Ceteris Paribus* laws. *British Journal for Philosophy of Science, 52*(2), 359–370.
Schurz, G. (2002). Ceteris Paribus Laws: classification and deconstruction. *Erkenntnis, 52*, 351–372.
Sober, E. (1997). Two outbreaks of lawlessness in recent philosophy of biology. *Philosophy of Science, 64*(4), 458–S467.
Strevens, M. (2003). *Bigger than chaos: Understanding complexity through probability*. Cambridge: Harvard University Press.
Strevens, M. (2008). Physically contingent laws and counterfactual support. *Philosophers' Imprint, 8*(8), 1–20.
Strevens, M. (2009). *Depth: An account of scientific explanation*. Cambridge: Harvard University Press.
Strevens, M. (2012). Ceteris paribus hedges: causal voodoo that works. *Journal of Philosophy, 109*, 652–675.
Tooley, M. (1977). The nature of laws. *Canadian Journal of Philosophy, 7*(4), 667–698.
Woodward, J. (2003). *Making things happen*. Oxford: Oxford University Press.
Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, part I: a counterfactual account. *Noûs, 37*(1), 1–24.
Woodward, J. (2007). Causation with a human face. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's republic revisited* (pp. 66–105). Oxford: Oxford University Press.
Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, part I: A counterfactual account. *Nous, 37*(1), 1–24.

Chapter 7
Evolutionary Explanations and the Role of Mechanisms

Gerhard Schurz

Abstract In the first section I outline the three basic theoretical assumptions of a generalized theory of evolution: the Darwinian modules of reproduction, variation, and selection. The analysis of functional traits of evolutionary systems which I give in Sect. 7.2 is grounded on these assumptions. The evolutionary explanation of the emergence of functions leads me to an abstract schema of evolutionary explanations that is based on iterations of variation and selection processes. In the final Sect. 7.3, it is shown, at hand of the problem of explaining evolutionary *macrotransformation*, that abstract evolutionary explanations are considered as inadequate by evolutionary scientists as long as not at least some *plausible* mechanism can be given, both for the production of complex variations and for their selection.

Keywords Evolutionary explanation • Macrotransformation • Evolutionary function • Evolutionary mechanism

7.1 General Evolution Theory: The Three Darwinian Modules

According to Darwinian evolution theory in its contemporary stage, evolutionary processes consist of the following "Darwinian" postulates or modules (Schurz 2011):

G. Schurz (✉)
Heinrich-Heine-Universität Düsseldorf, Institut für Philosophie,
Universitätsstraße 1 Geb.23.21, 40225 Düsseldorf, Germany
e-mail: schurz@phil.uni-duesseldorf.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_7,

Module 1 – Reproduction: There are entities – organisms or more generally evolutionary systems – which again and again reproduce themselves in regard to certain significant traits; these traits are called reproduced or inherited traits, and every such reproductive process produces a new generation.

Module 2 – Variation: Reproduction brings along variants that are reproduced or inherited at the same time.

Module 3 – Selection: There is selection, because certain variants are fitter under given environmental conditions, that is, they reproduce more quickly than others, thus replacing the other variants in the long run. The selecting parameters of the environment are also called *selection parameters*.

Sober (1993, p. 9) summarizes the three modules as "inheritable variation of fitness." Thereby "fitness" is understood as the *effective* rate of reproduction, that is, the average number of reproducing offspring. Varying rates of reproduction alone lead only to a *weak* selection in the sense of a continuous decrease of the relative frequency of the less fit variants. This does not yet mean that these variants have to become extinct. Yet in all realistic examples there exist upper bounds to population size due to the environment's limitation of resources. As a result, *strong* selection occurs, that is, the frequency of the less fit variant does not only decrease, but at one point in time, these variants eventually become extinct.

We should add that in contemporary evolution theory, an "overall adaptationism" is no longer tenable. Evolutionary processes are not solely the result of selection and adaptation; a further important kind of evolutionary processes are different kinds of random drifts that are caused by selectively neutral variation mechanisms. More importantly, not every phenotypical treat which is the result of selection processes is the cause of a selective advantage and hence has a direct adaptive explanation; many traits have been selected as mere causal side effects of other traits which have an adaptive explanations (see (EE)(2.) in Sect. 7.3).

Dennett (1995, 64f.) has emphasized that the three Darwinian modules make up an *algorithmic* process, whose fundamental properties are as follows:

1. The evolutionary process can be outlined in an abstract and object-neutral manner, which is why at least in principle evolution theory can be generalized to many object areas, also outside of biology, for example, to the evolution of culture (see below).
2. The algorithmic process consists of certain basic steps: (a) the reproduction of the genes or, in general words, of the system's "reprons" (see below);

(continued)

(continued)

(b) their variation; (c) the causal creation of organisms with their phenetic traits; as well as (d) the selection of these organisms and their genes, based on varying rates of reproduction of their genes.

3. Algorithmic processes are *recursive* (or iterative), that is, *the same* sequence of simple steps is repeatedly applied to the result produced in the meantime.[1] In this way, from very many *local* steps strung together bit by bit, a *global* result of development emerges, which is in no way already discernible from the "internal nature" of the local steps and often enough cannot even be calculated mathematically in advance but only be understood and explained evolutionarily. This recursiveness is indeed the secret of all evolutionary processes. It leads to highly complex structures resulting from the iteration of astonishingly *simple* basic components, which then look as if a "superior designer" had conceived them. Recursive procedures are also the most important foundation of formal logics and computer programs.

Typical evolutionary processes are *quasi-teleological*: From their selective directedness a goal *seems* to result, which is pursued. However, a lineage owes its directedness only to the *stability* of selective forces over many generations. If the selective forces or selection criteria change strongly, the direction of evolution subsequently changes, too. On the basis of such changes of direction, evolution can be divided into stages, for example, anaerobic versus aerobic unicellular organisms. In contrast, it is no longer possible to speak of directed evolution, if the selection parameters change in a quick and irregular manner, with alteration rates of a similar magnitude than the generation rates. This can lead to strong fluctuations or even to chaotic developments. It is questionable whether in this case one can still speak of evolution at all – at least one cannot speak here of *directed* evolution anymore. As a prerequisite of directed evolution, one has to assume a *fourth* condition (in addition to the three modules mentioned above):

Condition for directed evolution – stability of selective forces: The alteration rate of the selective forces is either low compared to the generation rate, or else the changes are regular or predictable.

[1]This condition is curiously missing in Dennett (1995); though it is the most important one (see also Boyd and Richerson 1985, 20f.).

The or-else phrasing is necessary, because organisms can adapt to changing environments very well, so long as the alterations are regular. Accordingly, species are differentiated into *specialists* and *generalists*; the latter adapt to altering conditions (Sober 1993, p. 21). A simple example is the adaptation to the times of day and year. There are also more complex examples, namely, the amphibian arrowhead, whose leaves assume a sea grass-like shape under water, one similar to the leaf of a water lily on the water, and an arrow-shaped one on land (Wilson 1998, 185f.). The most generalist living being is, without doubt, man.

Because of its abstractness and object-neutrality, evolution theory can, at least in principle, be generalized to other domains in which the three Darwinian modules are realized in some form. One example is *cultural* evolution. Let us begin with the negative demarcation of the theory of cultural evolution from sociobiology. Sociobiology (similar to evolutionary psychology) considers the cultural development of humans as being ultimately determined by their genes. In contrast, in the theory of cultural evolution, cultural development is precisely *not* reduced to the genetic-biological level and tried to explain from there. Rather, a *distinct level* of cultural (social, technical) evolution is assumed: the evolution of *memes*. The meme concept has been introduced by Dawkins (1976, Chap. 11) as the "cultural counterpart" to the genes.[2] With memes are meant human ideas and skills, which are reproduced by the mechanism of *cultural tradition*. In that regard it should be emphasized that "culture" is always understood *in a broad sense* here, as everything human made, which cannot be reduced to the human genes – cultural evolution therefore comprises not only cultural history in a narrow sense of moral and religion, art, and literature but also social, political, and legal history and in particular the evolution of science and technology.

For the evolution of memes, it is not important which position one assumes in the mind-body controversy – whether one sees memes rather as neuronal brain structures or as mental thought structures. Essential is only the presence of the three Darwinian modules. In order to be able to describe these modules sufficiently generally, we introduce a few additional object-neutral concepts of the generalized theory of evolution (GE), summarized in Table 7.1. Every kind of evolution consists first and foremost of its specific *evolutionary systems* – these are those systems that are in a direct interaction with the environment. In biological evolution (BE) these are the organisms – in cultural evolution (CE) the cultural systems created by humans. Evolutionary systems always possess certain subsystems or parts, which are more or less *directly* replicated or reproduced from each other: we call these subsystems in a generalized manner the *reprons* or repron complexes. The reprons of BE are the genes, gene complexes, and genotypes; the reprons of CE are the memes or meme complexes, that is, stored information in the human brain or mind, respectively. BE is characterized by the additional condition of sexual reproduction and genetic diploidy; this peculiarity does not occur in CE.

[2]On meme theory cf. Blackmore (1999), Aunger (2000, 2002), Mesoudi et al. (2006), Schurz (2011).

Table 7.1 Concepts of GE, applied at the levels of BE and CE

Generalized evolution	Biological evolution	Cultural evolution
Evolutionary systems	Organisms	Human societies
Reprons	Genes in the nucleus	Memes or acquired information
Phenetic traits	Organs, abilities	Skills, procedures, language, ideas and thought patterns
Reproduction	Replication, DNA copy	Passing on to next generation by imitation and learning
Variation	Mutation and recombination	Interpretation and variation of passed on memes
Selection	——— Higher rates of reproduction due to higher ———	
	rate of propagation	cultural attractiveness
Inheritance	Sexual (diploid)	Asexual (blending inheritance)

We call those traits and skills of an evolutionary system, which are produced by the reprons in the course of its individual development, the *phenetic* traits of the evolutionary system. In BE these are the organismic traits, in CE the cultural or technical products or the institutions, which have emerged from human memes. *Selection*, finally, comes about on all levels by certain kinds of evolutionary systems and underlying reprons reproducing more quickly under the given environmental conditions than others. Beside these fundamental similarities between BE and CE, there are of course also a number of important differences, for example, intentionally directed variations in CE, which however constitute no fundamental obstacle to the application of the three Darwinian modules. Table 7.1 compiles the fundamental concepts of GE and their counterparts at the levels of BE and CE.

Opponents of the generalized theory of evolution (GE) have often reproached it with the claim that the transfer of the theory of biological evolution (BE) to cultural evolution (CE) is a mere metaphor. Yet as we have developed GE here, it involves entities on the cultural level which are by no means merely metaphorically; they rather *literally* reproduce and in doing so are subject to processes of variation and selection. For instance, according to cultural evolution theory, the carriers of technical evolution are precisely not the technical appliances or resp. artifacts – in this traditional view evolution would indeed only be a metaphor, as technical appliances do not reproduce. Rather, the carriers of technical evolution are the culturally reproduced skills as well as manners of the production and utilization of technical appliances, and they obviously do reproduce.

7.2 Functional and Evolutionary Explanations

Evolutionary explanations in biology have predominantly been discussed in the context of the controversy about evolutionary function concepts and functional explanations (cf. Allen et al. 1998). The function concept is a philosophical problem of a long tradition, whose discussion is directly linked to the dispute between the theory of evolution and creationism or teleology. Let us consider the basic form of a functional explanation:

> (FE) Basic form of a functional explanation: Systems (or species-members) of type S possess a certain trait T in order to perform a certain function F, which has a high value for S.

For example, the vertebrates possess a heart that circulates the blood in the body, in order to provide the body with oxygen.

Functions are certain causal effects of the underlying organs or resp. traits of evolutionary systems. Cummins (1975) suggested analyzing functions as common effects of complex systems. This analysis, however, is unable to clarify what distinguishes effects performing biological purposes (like the heartbeat) from nonfunctional effects (like the falling of a stone, when I let it go). So, the central task lies in working out the difference between nonfunctional and functional causal effects. In principle there seem to be only three strategies that correspond to time-honored philosophical doctrines for this purpose:

First, one can perceive "in order to" in the sense of the intentional function concept as a creator's intention, who has purposively constructed system S with trait T in this way, so that it has the effect F. Applied to the macroperspective of evolution, this intentional function concept leads straightaway to creationism.

Second, one can perceive the "in order to" as an ontologically distinct "force" by which the future attracts the past, which can in no way be reduced to scientific causation. In this way one arrives at the Aristotelian conception of teleology.

Both creationism and teleology are hardly tenable from a contemporary scientific viewpoint. This leaves, *third*, the evolutionary analysis of functions, which makes the concept of the function compatible with a causal-naturalist analysis, neither requiring divine creators or teleological forces nor relinquishing the quasi-directedness of function. On that note, Millikan (1989, p. 13), Neander (1991, 174), Sober (1993, p. 84), Schurz (2001, §4), and others have proposed different variations of the following evolutionary analysis of the function concept:

(EFC) The evolutionary function concept: A causal effect E of a subsystem (organ) of an evolutionary system (organism) S is an evolutionary function iff (1) E is a reproduced ("heritable") trait and (2) the reprons (genes) on which E is based were selected because they have predominantly contributed to the evolutionary fitness of species S in its evolutionary history through the effect E.[3]

Addition (see below): If only condition (1) but not (2) is fulfilled, E is called a mere evolutionary side effect of (a subsystem of) S.

The concept of the evolutionary function is a special case of the so-called etiological function approach (Wright 1976), according to which a system S possesses a trait T with function F if and only if S causes F by means of T and F is in some manner valuable to S. Thereby, the prima facie normative condition of valuableness can be characterized differently (Bedau 1998); in the concept of the evolutionary function, it means as much as selective advantage and can therefore be reduced to a purely descriptive condition (Wachbroit 1994, p. 580).

Distinctive of the concept of the evolutionary function is its just-mentioned historical nature: functions are constituted by the relevant selection history of the relevant trait T of species S. Bigelow and Pargetter (1987), by contrast, have proposed a function concept that depends only on the present time, identifying an organ's function with the organ's present disposition of contributing to the fitness. Millikan (1984, p. 29) objects to this, quite rightly, that this presence-related explication can no longer distinguish between evolutionarily normal functions and dysfunctions. Accordingly, it is still the evolutionarily normal function of a damaged pancreas to produce insulin, and only for this reason we can say that the pancreas of a diabetic human, which no longer produces enough insulin, is no longer able to perform its evolutionary function, that is, it is biologically defective. Bigelow and Pargetter would have to say that in the contrary, the pancreas of a diabetic does no longer have the function to produce insulin. In short, an organ can also have an evolutionary function, without in fact performing it or even being able to perform it (likewise Laurier 1996, 27f.).

Not every evolutionary selected trait needs to have a direct evolutionarily adaptive function – many such traits are mere side effects of such functions. In Schurz (2001) the selected traits of evolutionary systems (whether they are functional or side effects) are called "prototypical" traits and are defined as in the addition to the abovementioned explication (EFC). A prototypical trait or effect E performs an evolutionary function if and only if the selection of the underlying

[3]The explication corresponds to Neander's short version of Millikan's concept of the *proper function*, enriched by two additions suggested in Schurz (2001), that T has to be a heritable trait and that the reprotype has to have predominantly contributed to the evolutionary fitness. The additions are meant to solve the problems explained below.

repron RE happened because of E itself; otherwise, E is merely a side effect of other functions caused by RE. For instance, it is the evolutionary function of the heart to pump blood, whereas the sound of the heartbeat only is a side effect of this function that by itself does not have any biological function (Cummins 1975; Bigelow and Pargetter 1987). The distinction between functional-adaptive traits and mere side effects forestalls an excessively adaptationist perspective.

Let me explain these notions by means of a few additional examples. In the framework of cultural evolution (CE), it is an evolutionary function of matches to catch fire if struck against an ignition surface, as they have been selected in CE for this purpose. That in striking the match one occasionally burns one's fingers is a typical side effect of this, while the color of matches is not a prototypical trait at all (neither as a function nor as a side effect). Analogously, in the framework of biological evolution (BE), it is an evolutionary function of noses to be able to smell and to protrude from the face, and it is a typical side effect of this that in winter noses cool down comparably fast, while – in contrast to the claim by Voltaire's Dr. Pangloss (Gould and Lewontin 1979, p. 583) – it is not a biological-evolutionary trait of noses to be able to wear glasses. Yet, conversely, it is certainly a cultural-evolutionary trait of glasses to be able to sit on noses.

Fodor and Piatelli-Palmarini (2010) have argued that the distinction between functions, that is, selectively advantageous traits, and mere side effects would be "intensional" and thus would constitute a fundamental obstacle to the theory of selection, as selection processes are "extensional." An excellent refutation of this view is given by Block and Kitcher (2010): they point out that the distinction between functional traits and mere side effects is an obvious and harmless causal distinction, which presumes nothing more than basic assumptions of causality (directed cause-effect relations), which are almost universally accepted in the sciences. Selection process is defined in terms of causal relations, and therefore, the distinction between a trait T that is the cause of a reproductively advantageous effect E and another trait T' which is a causal side effect of T is not at all obscure or "intensional", at least not in any sense of this word in which cause-effect relations are not intensional.

In common sense, those traits of the environment of the evolutionary system, towards which the system has adapted, are often regarded as "functional traits" of the environment. For example, it is said that the function of rain and sunshine was to feed plants water and energy. But neither the Sun nor the Earth's water cycle is an evolutionary system. Conversely, it is rather the plant which is the evolutionary system, whose functionality systematically utilizes these environmental traits. Nevertheless, we can accept the common sense mode of speaking as a derived-functional mode of speaking and accommodate it through the following additional convention: An environmental trait U is evolutionary-functional in a derived sense, iff there are evolutionary systems S with evolutionary functions F, which have adapted towards U. This means that the selective advantage caused by F would not have come about, if the trait U had been predominantly absent in the history of S's

environment. With this extended concept of derived-functional environmental traits, we evolutionarily grasp the entire scope of facts, on which the creationist argument from design is based, without leaving the naturalist perspective.

The condition in (EFC) that the effect E must predominantly contribute to the evolutionary fitness of species S in its evolutionary history ensures a connection between evolutionary normality and statistical normality. A number of philosophers of biology (e.g., Millikan 1984; Neander 1991; Wachbroit 1994; Laurier 1996) have argued against this by claiming that evolutionary normality would be independent from statistical normality. In contrast, Schurz (2001) attempts to show by means of a logically elaborate argument that, while evolutionary normality is not identical with statistical normality, the former implies the latter – at least, if one understands evolutionary normality in the sense of the above-defined evolutionarily prototypical traits. For, roughly speaking, there are only three possible reasons for why a selectively advantageous trait does not also become statistically dominant:

First, the trait is not predominantly a heritable or reproduced trait but its appearance strongly depends on the environmental conditions. For this reason we (unlike Millikan 1984, pp. 20, 29) restricted our explication (EFC) to reproduced traits. This restriction solves Millikan's objection (Millikan 1989, 62ff.) that many evolutionary functions of organs are only performed rarely. For instance, sexual reproduction is certainly the most important function, yet in many species it is only performed by a few individuals – namely, only by those of the numerous offspring that survive until reproductive maturity. But the genetically determined phenotypic condition that is selected here is not the actual performance of the function but the disposition to the performance of the function under suitable circumstances, and this disposition is present in almost all members of the species. In general, normic-evolutionary traits as a rule are not actual traits but dispositional traits.

Second, a trait may not become dominant if in the selection history of the trait, alongside stages of positive selection, there have also been long stages of negative selection. In these cases the formation of a normic and statistically dominant trait does not occur, but rather, a polymorphism of traits is generated. In order to rule out selectively ambivalent cases, we have required in the explication (EFC) that the reprons on which the trait is based have in the history of S predominantly contributed positively to S's evolutionary fitness.

Third, it may be that, while a trait E had been predominantly positively selected in the history of S, the evolutionary history of S was interrupted by an externally triggered catastrophe, so that the time was too short for E to become statistically dominant. Catastrophes admittedly occur repeatedly in evolution, but for evolution to be able to take place at all, they have to be sufficiently rare. For this reason we restrict the claimed connection between evolutionary and statistical normality to the "major part" of evolution and permit exceptions caused by catastrophes.

By means of this analysis, all objections against the connection between evolutionary and statistical normality that are known to me are no longer applicable (for further elaborations cf. Schurz 2011, 2012)

7.3 Evolutionary Explanation and Mechanism

The preceding evolutionary explication of functions can be turned into the following explication of evolutionary explanations:

(EE) Evolutionary Explanations:

Explanandum: An evolutionary species (organism) S possesses a certain subsystem (organ) with a certain effect E that S's ancestor species (S′) did not possess.

Explanans: (1.) Certain combinations of variations (mutations) in the ancestor species led to the appearance of a new complex of reprons (genes) R_E that produced the effect E in the normal environment of the ancestor species, leading to a new variant S*.

(2.) In the subsequent history of the ancestor species, the causal effect E was selected because it had predominantly contributed to the evolutionary fitness of the new variant S*, which by successive reproductive isolation evolved into the new species S* (while the ancestor species S′ either died out or transformed into a distinct species S**).

A remarkable feature of this explication of evolutionary explanations is that it does not inform us about *causal mechanisms*. This lack of mechanisms arises at two places:

Mechanism of variation: We are not informed about the mechanisms by which a series of variations lead to the new genes or reprons R_E which produce the new phenotype E. This is not a problem in the case of so-called micro-transformations, in which the new phenotype differs only a little from the old one, because mechanisms for microvariations (such as mutation in biology) are well known. Examples are leg length of hoofed animals, or beak sizes of birds, etc. However, in the case of so-called macrotransformations, where an entire new type of organism appears, such as the transition from water-living to land-living animals or from nonflying into flying animals, the lack of causal explanations indeed constitutes a problem. Many critics of Darwinian evolution have objected that the combination of *independent* improbable mutations which are necessary

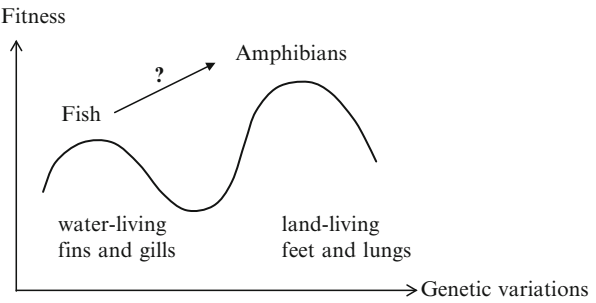


Fig. 7.1 Passage through a fitness valley

to produce the required new macrotrait seems to be far too improbable to be possible without creationist assumptions.

Mechanisms of Selection: Also, abstract evolutionary explanations do not inform us about the mechanisms of selection, that is, *how* the new phenotypic effect E leads to the selective advantage. This may be no problem for fully developed macrotraits whose function is clear; but it is much more unclear for intermediate forerunners of these traits. For example, the wings of present birds enable them to fly, and this is a clear selective advantage, but what was the selective advantage of vestigial wings in bird-forerunners who were too small to enable their possessors to fly?

The remainder of this section contains an analysis of the evolutionary explanations of *macrotransformations*. The analysis will show that, in fact, evolutionary explanations are not considered as adequate by evolutionary scientists as long as not at least some *plausible* mechanism can be given, both for the production of complex variations and for their selection. The "plausible" mechanism need not be empirically confirmed to a high degree, but it must not be *too* improbable in the given background knowledge. In this respect, the mechanisms cited in evolutionary explanations often may generate only a *how-possible* explanation rather than a full causal explanation (see Schurz 1999, 110f.).

The fundamental problem of the origin of new macrostructures by successive mutations is the apparent necessity of the passage through a *fitness valley*. One example is the transition of water-living fish to land-living amphibians and reptiles. Of course, fishes who occasionally rob with their fins in shallow water and subsequently on land suffer drastically in their fitness. So how could those fishes who supposedly evolved into amphibians have survived their first steps without God's help? The problem is illustrated in Fig. 7.1.

If the origin of a new macrotrait every time requires the passage through a fitness valley that is life threatening for the evolving species, why then have so many new macrotraits originated in the evolution, without all species having become extinct in the process? For this there is an explanatory solution, which at least in most cases, has subsequently delivered the missing explanation and thereby increased the probability of the process. It consists in the existence of specific transition forms during a macrotransformation, possessing a rudimentary antecedent of the new macrotrait, which in the given environment performs some function *other* than

the later one, so to speak a *proto-function*, due to which the antecedent was already able to prove fitness-increasing.

We at first illustrate the process in the example of the evolutionary transition of fishes to amphibians. How could the transition from the fishes to the amphibians run its course, if, on the one hand, the amphibians' new macrotraits, that is, feet and lungs, would only be disadvantageous in fishes and, on the other hand, fishes without feet or lungs would die very rapidly on land?

Absolute all terrestrial vertebrates (the tetrapods or quadrupeds) descend from the ancestors of the tetrapodic bony fishes, that is, the coelacanths (of that time). In contrast to (almost) all other fishes – summarized as ray-finned fishes – these do not have vertically attached fins but similar to the feet of a vertebrate have laterally-horizontally attached front and back fins. It is assumed that the coelacanths were the ancestors of the first amphibians and "waddled" with their fins in shallow waters on the bottom of the water and close to the shore. They were able to procure food there, which other fishes were unable to reach. In doing so, they occasionally also waddled out of the water and deposited their eggs in the wet sludge outside of the water. This constituted an enormous selective advantage because there were not any predators there. During a transition period of millions of years, fins more and more similar to feet were selected as well as lung-like respiratory organs (next to the gills) for respiration outside of the water. In principle, the skin and specifically the mucous membrane are able to absorb oxygen from the atmosphere, and it is assumed that antecedents of lungs developed from an enlargement of the oral mucous membrane turned to the inside. As soon as the modified proto-amphibians were able to stay on land for a longer time, an explosion-like multiplication and diversification of the new beings on land was the consequence, as this region had so far been unoccupied and accommodated huge quantities of novel ecological niches, giving space to new life forms.

From the perspective of the early form, or the antecedent, this process is called *exadaptation*: a trait that has once been selected for other purposes assumes a different function. From the perspective of the later function, one also speaks of *preadaptation*: the trait performing the proto-function was "created in order to" fulfill a different and completely novel function later – whereby "created in order to" must not be misunderstood in the teleological or creationist sense, as if this had been planned at a higher level (cf. also Ridley 1993, 329f.). In this sense, the tetrapodically arranged fins of the coelacanths were a preadaptation for the amphibians' feet, or the latter were an exadaptation of the former. The process of preadaptation or exadaptation is ostensively displayed in Fig. 7.2. The fitness landscape is transformed by it from a roller coaster to a smooth climb.

In an analogous manner a number of additional macrotransformations could be explained:

1. *The transition from the saurians to the birds*: How did the birds develop their feathered wings? Birds originate from certain saurians. In some saurians a plumage developed with the proto-function of thermoregulation (Millikan 1989, p. 44). Meanwhile, fossils of feathered saurians of the size of contemporary

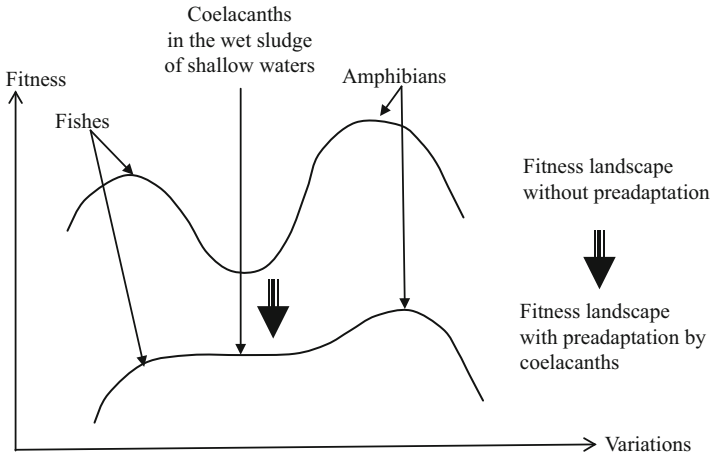


Fig. 7.2 Transformation of the fitness landscape by preadaptation (or exaptation). Transition from the fishes to the amphibians via the coelacanths

flightless birds have been found. To be sure, birds, like mammals, are warm-blooded animals and so are able to keep their body temperature high even in a cold environment, which reptiles cannot – in the case of cold, they fall into a state of motionlessness, which has possibly also been responsible for the extinction of the saurians in a cold period after the comet impact. The light skeletal structure likewise had already developed in saurians, as these due to their size are not able to move without extremely light bones. Wings could have developed from feathered flying membranes. Spreads of skin between finger or toe bones and also between body and extremities have indeed developed several times in evolution: in aquatic mammals and birds to fins and in tree-living reptiles and mammals to means of gliding from tree to tree.

2. *The origin of warm-blooded mammals*: The first mammals during the reign of the saurians have predominantly evolved to be nocturnal. With respect to the requirement of the upkeep of the necessary body temperature during the cold night, they were therefore subject to strong selection. By means of their warm-bloodedness, they were able to adapt much better to the global ice ages than the saurians.
3. *The transition from land mammals to aquatic mammals*: Whales (and later other aquatic mammals) have evolved about 50 million years ago from hippopotamus-like hoofed animals from Pakistan. During a transition period of several million years, their feet have again transformed to fin-like extremities; fossil transition forms, like the Pakicetus, are known.[4] Gills did not form again; instead, aquatic

[4]See en.wikipedia.org/wiki/Evolution_of_cetaceans.

mammals can hold their breath (breathed via nostrils) for a long time, but they have to surface regularly in order to breathe.

4. *The development of an adaptive brain in Homo sapiens*: The steadily growing brain of the hominids requires an increasingly longer and more risky gestation period of the embryo. A solution could have been to allow the brain to continue growing after gestation. The plasticity of the child's brain originating from this proto-function could in the following have been the basis of the development of the systematic learning ability of the brain of *Homo sapiens*.
5. *The transition from the prokaryotes to the eukaryotes*: The first one-cellular organism in evolution, the prokaryotes (bacteria, algae), consisted (basically) only of a cell membrane with RNA in it. The cell membrane of the prokaryotes was not as permeable as that of today's eukaryotes (which are cells with organelles such as nucleus, plastids, and mitochondria). The prokaryotic cell membrane rather contains *murein* as a solid supporting layer that is much more rigid than the eukaryotic cell membrane and which, in contrast to the latter, can only let through smaller molecules, but no macromolecules or even small prokaryotes. For prokaryotic cells to become eukaryotic cells being able to perform phagocytosis, that is, to swallow entire prokaryotes, the prokaryotes first had to abandon the rigid cell wall. But the rigid cell wall *protected* the prokaryotes from diverse harmful influences, while the eukaryotes' more complex protection mechanisms, like primitive perception and locomotion, were not available to the prokaryotes. The prokaryotes consequently had to first pass through a *fitness valley*, that is, become more vulnerable, in order to travel from their previous fitness peak to a still higher fitness peak, that of the eukaryotes (Maynard-Smith and Szathmáry 1995, pp. 122–126). This necessity of the passing of fitness valleys explains the long period of stagnation of almost 200 billions of years, before the level of eukaryotes could be reached and a new stage of evolutionary explosion could begin. How can the transition through this fitness valley be explained? Even today, this question is not answered. A clue is provided by the fact that there is a special group of bacteria, the archaebacteria, which were previously taken for an especially old bacterial stem species, but which according to more recent findings have more in common with the eukaryotes than with the remaining bacteria (the eubacteria), for which reason they are today considered as a sister species of the eukaryotes.[5] Archaebacteria, as opposed to eubacteria, do not possess a rigid supporting cell wall made of murein, so that their predecessors might have been the point of origin of the eukaryote evolution according to the endosymbiotic theory.

The preceding examples of evolutionary explanations of macrotransformations can be summarized as follows: evolutionary explanations are not considered as adequate unless they do not contain *plausible* (though not necessarily empirically

[5]Cf. Maynard-Smith and Szathmáry (1995, 125f.), Cavalier-Smith (2002), Szathmáry and Wolpert (2003, p. 272), as well as en.wikipedia.org/wiki/Evolution

confirmed) mechanism, both for the variations that are necessary to produce of new phenotypic traits or functions and the selective advantage that they conferred to the new variant. Typically this is done as follows: the macrotransformation from species S to S* are decomposed into a sequence of plausible microvariations V_1, $V_2, \ldots$, such that for every of these microvariations V_i, a plausible mechanism M_i (being based on mutation and recombination in BE) can be provided, by which the corresponding intermediate species variant acquires a selective advantage to its immediate predecessor. In conclusion, our explication (EE) of evolutionary explanation in the beginning of this section is incomplete and has to be complemented by the following third condition:

(EE)(3.) The required variations in (1.) are produced by a sequence of microvariations V_1, $V_2, \ldots$ each of which possesses a selective advantage as required in (EE)(2.) by way of mechanism $M_1, M_2, \ldots$

Let us finally ask why the problem of macrotransformations has been so intensively discussed as a problem for the theory of biological evolution, but not in the domain of cultural evolution. There is a simple answer to that. The mechanisms of variations in cultural evolution (CE) are way different from the mechanisms of variation at the biological level. Cultural variants do not appear "blindly" like biological mutations but are usually goal intended and rationally planned (cf. Boyd and Richerson 1985, p. 9). This difference does not constitute any real objection to the applicability of the Darwinian modules to cultural evolution. While technical inventions, for example, are not blind mutations, they are in a multifarious way *flawed* and *imperfect*. Thus, they are capable of a systematically optimizing selection, and this is all that Darwinian evolution requires. However, the directedness of cultural variations makes an important difference concerning the possibility of macrotransformations. In CE it often happens that a human individual *simultaneously* varies several connected but different ideas or skills in a directed manner. If this is the case, then something like a cultural "macromutation" results. For instance, with the invention of cooking on the fire, cooking stoves, cooking containers, etc., were invented at the same time. The inventors of the wagon wheel at the same time invented axles, the chassis, and roads. When Einstein postulated the speed of light as the maximum velocity of propagation, for the sake of consistency, he simultaneously replaced the Galilean transformations of velocity with the Lorentz transformations. In this way, spontaneously successful memetic macromutations, paradigm shifts, or mental subversions can indeed oftentimes occur in CE, which in BE they are very improbable. The probabilistic reason why coordinated macrovariations in CE are no longer improbable is simple: *given* the *intentions* of the intentional subjects who produce these variations, they are no longer probabilistically independent from each other (as mutations) but are positively probabilistically dependent.

References

Allen, C., Bekoff, M., & Lauder, G. (Eds.). (1998). *Nature's purposes*. Cambridge, MA: MIT Press.

Aunger, R. (Ed.). (2000). *Darwinizing culture: The status of memetics as a science*. Oxford: Oxford University Press.

Aunger, R. (Ed.). (2002). *The electric meme: A new theory of how we think*. New York: Simon & Schuster, Free Press.

Bedau, M. (1998). Where's the good in teleology? In C. Allen, M. Bekoff, & G. V. Lauder (Eds.), *Nature's purposes* (pp. 261–291). Cambridge: MIT Press.

Bigelow, J., & Pargetter, R. (1987). Function. *Journal of Philosophy, 84*(4), 181–196.

Blackmore, S. (1999). *The meme machine*. Oxford: Oxford Paperbacks.

Block, N., & Kitcher, P. (2010). Misunderstanding Darwin. *Boston Review*, Mar/Apr 2010. bostonreview.net/BR35.2/block_kitcher.php

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Cavalier-Smith, T. (2002). The neomuran origin of archaebacteria, the negibacterial root of the universal tree, and megaclassification. *International Journal of Systematic and Evolutionary Microbiology, 52*, 7–76.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72*(20), 741–765.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

Dennett, D. C. (1995). *Darwin's dangerous idea*. New York: Simon & Schuster.

Fodor, J., & Piatelli-Palmarini, M. (2010). *What Darwin got wrong*. London: Profile Books.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian paradigm. *Proceedings of the Royal Society of London B, 205*(1161), 581–598.

Laurier, D. (1996). Function, normality, and temporality. In M. Marion & R. S. Cohen (Eds.), *Québec studies in the philosophy of science* (pp. 25–52). Dordrecht: Kluwer.

Maynard Smith, J., & Szathmáry, E. (1995). *Evolution*. New York: Oxford University Press.

Mesoudi, A., Whiten, A., & Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behavioral and Brain Science, 29*, 329–347.

Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.

Millikan, R. G. (1989). In defence of proper functions. *Philosophy of Science, 56*, 288–302.

Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science, 58*(2), 168–184.

Ridley, M. (1993). *Evolution*. Oxford: Blackwell Scientific.

Schurz, G. (1999). Explanation as unification. *Synthese, 120*(1), 95–114.

Schurz, G. (2001). What is "normal"? An evolution-theoretic foundation of normic laws and their relation to statistical normality. *Philosophy of Science, 68*(4), 476–497.

Schurz, G. (2011). *Evolution in Natur und Kultur. Eine Einführung in die verallgemeinerte Evolutionstheorie*. Heidelberg: Spektrum Akademischer Verlag.

Schurz, G. (2012). Prototypes and their composition from an evolutionary point of view. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford handbook of compositionality* (pp. 530–553). New York: Oxford University Press.

Sober, E. (1993). *Philosophy of biology*. Boulder: Westview Press.

Szathmáry, E., & Wolpert, L. (2003). The transition from single cells to multicellularity. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 271–290). Cambridge: MIT Press.

Wachbroit, R. (1994). Normality as a biological concept. *Philosophy of Science, 61*, 579–591.

Wilson, E. O. (1998). *Consilience. The unity of knowledge*. New York: Knopf.

Wright, L. (1976). *Teleological explanations*. Berkeley: University of California Press.

Part III
Explanation in the Historical Sciences

Chapter 8
Explaining Roman History: A Case Study

Stephan Berry

Abstract The Roman Empire occupies a pivotal position in modern perceptions of history, and it is certainly one of the most intensely investigated cultures of the past. Nevertheless, we are far from knowing "everything," and the concept of explanation becomes crucial in particular for those phenomena that are adequately represented neither in the written records studied by historiography nor in the material remains studied by archaeology. One example is the question whether the Romans had a Grand Strategy and how the geographic boundaries of their empire can be explained: such issues refer to plans, intentions, concepts of geography, and the like, which have to be reconstructed in a tedious way from the scarce surviving evidence, in order to obtain explanations for the strategic decisions made by the Romans.

Keywords Ancient historiography • Greece • Imperialism • Roman Empire • War

8.1 Introduction

The notion of explanation is fundamental in both scientific theories and philosophical accounts of how science works. A number of chapters elsewhere in this volume will address the issue of explanation in historiography and evolutionary

This chapter is based on the talk "Late Roman Decadence and Beyond: Explaining Roman History" that was presented at the workshop "Types of Explanation in the Special Sciences – The Case of Biology and History," organized by the Research Group "Causality, Laws, Dispositions, and Explanation in the Intersection of Science and Metaphysics (DFG 1063)," September 30–October 3, 2010, Cologne (Germany).

S. Berry
Freelance Science Author, Berlin, Germany
e-mail: stephan.berry@web.de; http://www.stephan-berry.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_8,

science from a philosophical perspective, but this chapter will look at the topic in a complementary way: Debates from current historiography of ancient Rome will be presented as a case study. This is intended to elucidate some of the problems that historiographers actually encounter in their work, i.e., problems that can be observed particularly in cases where unequivocally accepted explanations are absent.

In principle, problems of methodology are similar across all disciplines that deal with history in a broad sense of the word: history, archaeology, linguistic science, and so on in the humanities but also paleontology, evolutionary biology, or cosmology in the natural sciences. This fundamental similarity is a basic tenet of several works elsewhere in this issue, and the author of this chapter has also argued in favor of this methodological similarity (Berry 1999, 2008). However, *similar* problems and approaches do not imply *identity* in all respects. Human actors have attributes, such as intentions, plans, and beliefs, which are usually absent from the objects of study in historical fields of natural science: the emergence of our solar system can be reconstructed without any reference to the intentions of the planets, while mental states are essential for understanding events in human history.

This gives rise to a special situation in historiography: *apparently* plausible explanations for historical phenomena are usually rather easy to present because the modern observer feels related to the actors of the past through concepts such as feelings, needs, intentions, and the like. The fundamental similarity between observer and observed enables the construction of seemingly plausible ad hoc explanations in many cases: actor X did Y because he wanted to achieve Z (see, for instance, the habit of archaeologists to ascribe a "religious" or "ritual" purpose to any object for which other functions are not obvious).

Completely unexpected and unexplainable phenomena probably will be rare in human history, but occur rather often in the fields of natural history. When in 1995 the first planets outside our solar system were detected, they exhibited a number of features that were unexpected and unaccounted for in astrophysical theory. The existence of exoplanets *as such* had actually been anticipated. But their large atmosphere in combination with the small distance to their sun was not merely unexpected; it was something that one would have considered as outright impossible prior to this discovery: intense heat and radiation from the parent star should have blown all remnants of an atmosphere away, according to generalizations from observations of our own solar system. By contrast, the major problem that historians and archaeologists frequently face is to select between several *equally conceivable* explanations; the complete absence of explanations, however, is less frequent.

To start the tour through Roman history, the next section will take a look at the available sources. In addition to problems of interpretation, with which anybody working with documents from the past is confronted, Roman sources have peculiar features that affect the problem of finding explanations for some aspects of Roman history.

We will then turn to two major questions: First, the rise of Republican Rome from a city to a world empire. How can we explain that a single city became a superpower that could dominate the whole Mediterranean basin and many adjacent territories?

Second, the inner workings of the Empire once it had conquered the world: how can we explain the decision-making process in the center of power? The logical third step would be the decline and collapse of the Empire. This is the mystery of mysteries in history, and countless explanations have been offered. Therefore, this can of worms will not be opened here, given the limitations of space (see Ando 2008; Wolfram 1990, pp. 422–441; Demandt 1989, pp. 470–492 for an overview).

8.2 The Problems of the Sources

Rome is one of the most intensely investigated cultures and states of the past. The amount of things that we know about ancient Rome is enormous. Nevertheless, some areas are particularly prone to produce long-standing controversies and, correspondingly, a lack of undisputed explanations. So what are the gaps in the available evidence, leaving questions of Roman history unanswered, despite centuries of scholarly work?

There are two main types of evidence: material remains, as studied by archaeologists or art historians, and written accounts, studied by historians and philologists. In between, there are categories such as *papyri*, inscriptions on buildings, and coins, which fall in both realms because they are material objects on the one hand but contain textual information on the other.

Material evidence can tell us a great number of things about ancient conditions of life, of trade routes, of production processes, and so on. But with respect to causal explanations, in particular when it comes to explaining political or social processes, material evidence has limits. Let us take as an example a Roman glass vessel that is found in *Germania Magna*, the unoccupied part of Germany on the right-hand side of the Rhine. Chemical analyses will reveal the composition of the glass, and by some fancy methods, it may be possible to trace the provenience of the raw materials that were used, and perhaps one can even locate the workshop where the glass was made.

But how did it come into the soil in *Germania*, perhaps hundreds of miles away from the Roman *limes*? Conceivable explanations could be:

- It reached the *barbaricum* by means of normal trade.
- It was loot that Germanic raiders of the Roman Empire had brought home.
- It was given by the Romans to some Germanic king or chieftain, as part of diplomatic exchange of gifts.
- It was a piece that a Germanic mercenary in Roman service had acquired and brought home, when he returned after his term of service.

This example is intended to show that, using material evidence, we can answer many questions regarding *how*, *when,* and *where* the people of the past did what they did, but the central questions "*why?*" and "*in which historical context?*" are generally more difficult or impossible to answer, based on material evidence alone.

Answers to such questions concerning the intentions and motivations of the players in history are more aptly sought in the historical accounts, but these have their limits, too.

First, not everything that has happened was captured in written form, and, second, not everything that was written down has survived to the present day. These are trivial problems that relate to any written account of the past. Likewise, writing history was a pastime for members of the upper classes, giving rise to a considerable social bias in their writings, but again this is a problem that we are frequently confronted with in any historiography of the premodern age. Some more specific problems that relate especially to Roman history are depicted in the following section.

8.2.1 The Classical Model

Ancient historiography of later eras, i.e., Hellenistic Greece as well as Republican and Imperial Rome, took the earlier works of Classical Greece to be an authoritative model that was to be emulated as far as possible. This gave rise to an approach that squeezed the material into a canonical form, irrespective of whether this did justice to the matter at hand or not.

And historiography had a number of different purposes: to educate, entertain or surprise the reader, or to make a political point, rather than to capture the course of history in an objective or scientific way. The problem that the ancient sources deliberately blur the picture and that an elegant reading and the adherence to the canonical pattern is more important than accurate detail can affect any context that is "technical" in the broadest sense: technology proper, military matters, economic, administrative, or legal affairs. And then there were other topics which were not deliberately blurred but which were simply too trivial and too self-explanatory for the ancient reader to be expounded explicitly.

8.2.2 Rome as the Center of the Universe

For Roman historians, the city of Rome was the center of the universe, and "Roman history" was the history of that city. The empire-wide effects of Roman rule and decisions were not relevant, at least not in themselves. This means that issues which are important for modern historians, such as social and economic history, have to be reconstructed in a tiresome way from pieces of scattered evidence, often archaeological or epigraphic in nature, because the writings of ancient historians offer only meager material on these questions. In general, the written accounts are either Roman or Greek in perspective, and other people living under Roman rule remain silent for us.

The only and notable exception are the Jews because a number of sources allow to see the Empire with their eyes, in particular the works of Flavius Josephus. He is essentially the only author who follows the patterns of Greek and Roman historiography but writes from the perspective of a people on the periphery of the Greco-Roman world. Due to a lack of comparable sources, it is impossible for us to complement this with a national history of, say, the life of Numidian or Illyrian tribes under the empire.

8.2.3 The Not So Impartial Observer

Many Roman historians of the imperial time were senators which means that they, in addition to the mentioned general upper-class bias, also had a marked anti-emperor bias because the relations between emperor and the senate were frequently strained. This caused senatorial historians to use their writings for revenge, usually after the emperor in question was dead.

8.2.4 Politics in Secret

While politics in the republic had been at least partially a public affair, being discussed in the senate, the forum and the people's assembly, it had become something essentially secret under the emperors, with decisions being made in the inner circle of power. Therefore, in the imperial era, historians were able to properly expound the backgrounds and causes of political decisions only to a limited extent – and this limitation has, of course, been inherited by their modern successors. And then there are examples of ancient historians who actually belonged to these inner circles at some point of their career. Yet this does not guarantee that their accounts are particularly reliable, because of their involvement in court intrigue and the urge to use their knowledge for retaliation, as described above.

The previously mentioned problems relate to the interpretation of available sources, but with regard to the earliest phases of Rome, we face an additional difficulty: there were no contemporary writers, so that all material on the times of the kings, the origin of the republic, and the beginnings of its rise to dominant power in Italy originated at a later date. Now it is time to look at the rise of Rome as a superpower.

8.3 The Rise of Rome

Already in ancient times, observers were bothered by the question how a single city came to dominate the whole Mediterranean world. The Greek Polybius, who wrote in the second century BC, was the first historiographer to tackle this issue in a

systematic manner. His approach, i.e., asking for the cause of the rise of Rome, stands within a tradition of causal analysis that goes back to some of the most important Greek historians: Herodotus, the so-called founder of historiography, was interested in the cause of the wars between Persians and Greeks (the intricate story in Herodotus 5.23–97 includes earlier conflicts *within* the Greek world that preceded the Ionian Revolt against the Persians). Likewise, Thucydides wrote his work to elucidate the cause of the Peloponnesian War between Athens and Sparta (immediate causes of the war, Thucydides 1.23–87; the underlying long-term conflict between Athens and Sparta, 1.88–118).

One causal factor that Polybius regards as crucial is the constitution of Rome, which, according to him, is a perfect balance of different types of constitution known from the Greek cities (Polybius 6.18; comparison of Sparta's and Rome's ability to create a stable hegemony, 6.48–50). In addition, he considers the Roman army to be a crucial factor (6.19–42); this type of reasoning is popular even today: the Romans dominated the world because they had the best army in the world.

One will not encounter this explanation in academic circles today, but in the mass media, in accounts for a general audience, it is still alive. But what does "best" army in the world mean? – This is an undefined term. Moreover, here it seems appropriate to make a direct comparison to the present age; having the best army in the world is not enough: the US forces may be called the best army of the present, but winning battles is not the same as winning wars or creating a stable peace order, as the situation in Iraq or Afghanistan shows so clearly.

For Polybius, it was in particular the tactical superiority of the more flexible Roman legion over the rigid Macedonian *phalanx*, which had become a standard formation for many powers around the Mediterranean (Polybius 18.31 f.). This explanation of Roman success is also found in the Roman historian Livy, writing in the time of Augustus, i.e., about 150 years after Polybius (Livy 44.40–42). But looking at the encounters of *legio* and *phalanx* in detail, one sees that several times the Romans avoided defeat only by a hair's breadth, so the notion of a general superiority cannot be maintained (analyses of such battles in Cowan 2009, pp. 103–147;Pietrykowski 2009, pp. 195–236).

In general, Roman history was full of severe defeats, what calls the whole approach to this explanation into question. In spite of the undeniable qualities of the Roman army, one has to conclude: Rome was not invincible, and the crucial feature that requires an explanation is the resilience of the Roman state, i.e., the ability to create a military and political system that remained intact even in the face of a total disaster, such as the catastrophic defeats against Hannibal's army in the Second Punic War.

At this point, we must turn to the debate among modern historians. Their discussions of Roman expansion are centered on the notion of imperialism, which reveals that it originates in modern political science. In 1979, William Harris published "War and Imperialism in Republican Rome," which became one of the most influential books on Roman expansion. At the beginning of Chap. 1, he states:

> Since the Romans acquired their empire largely by fighting, we should investigate their attitudes towards war. (Harris 1979, p. 9)

In fact this is, in a nutshell, Harris' program: he investigates the Roman mentality and concludes that it is the extraordinarily warlike character of the Romans that compelled them to uninterrupted warfare, year after year, for centuries, until they had finally conquered anything that was worth to be conquered.

In his review of the debate, Rich (2004) identifies two major reasons why Harris' views became dominant: First, his theory replaced the older theory of "defensive imperialism," which had been promoted by Mommsen and had become widely accepted. According to that view, the Romans only went to war because they had to; they were fighting essentially defensive wars during which, as a side effect, their empire constantly grew. This paradoxical view of unintentional world conquest had become untenable, and Harris' approach seemed to offer a much more plausible explanation of the rise of Rome.

Second, the time at which Harris conceived his book was thoroughly influenced by modern anti-imperialism. The European powers had already lost their colonial empires, and exposing the evils of Roman imperialism clearly hit a nerve with many readers. The personal setting of Harris, who was writing as an Englishman in the United States during the Vietnam War, may explain in part the polemical character of his book, as Rich believes (see also Fitzpatrick 2010 on the comparison of ancient and modern "imperialism").

Harris' theory of a specific Roman urge to go to war had two facets. One is the immediate material benefits of conquest, i.e., the increase of territory, the influx of loot and money from plundering cities and selling their population as slaves, and so on. Concerning this aspect of the theory, Erich Gruen has demonstrated that the crucial decisions of the senate, when and where to go to war, were not generally dominated by economic motives:

> A growing body of scholarly literature finds war and greed tantamount to imperialism. The equation may be too simple. Distinctions need to be made and emphasized. The prospect of loot could entice generals and stimulate recruiting – which is not the same as determining a senatorial decision to make war. The carrying off of spoils and the exaction of indemnity might enrich the state, but would not necessarily impel it toward an enduring system of regulation and exploitation. Enslavement or sale of defeated enemies helped stock the plantations of rural Italy; yet nothing shows that this either inspired Roman expansion or dictated imperial control. The leaps of logic too easily distort and mislead. (Gruen 2004, p. 30)

But mere greed, the drive for material rewards, would at least have had some rational core. According to Harris, there is another, even darker, and wholly irrational side of the Roman attitude. According to him, the Romans overrated warrior ethos and military glory to such a degree that their attitude became outright pathological. And it is because of the focus on the notion of a pathological Roman lust for war that Harris' view has become popular.

Tim Cornell summarizes this standard view of Roman militarism in his comprehensive study of Rome's early history:

> For most of its history the Roman Republic was constantly at war, and a very high proportion of its citizen manpower was committed to military service. Its institutions were military in character and function, and its culture was pervaded by a warlike ethos. (Cornell 1995, p. 365)

However, he then introduces a new turn to the story, because he goes on:

> These facts are important, but they do not explain Roman imperialism; rather, they are themselves symptoms of the phenomenon that needs to be explained. Why were the Romans so belligerent? How did they manage to conquer Italy so quickly, and why was their control of the conquered peoples so thorough and long-lasting? In the last analysis, the answer to all these questions is the same, and is to be found in the nature of Rome's relations with her neighbors from the earliest times.
>
> The foundations of Roman military power were firmly laid in the settlement that followed the Latin revolt in 338 BC. [. . .] The settlement of 338 established a hierarchy of relationships in which the subject peoples were categorized as full citizens, citizens *sine suffragio*, Latins and allies. These various groups had one thing in common: the obligation to provide troops for the Roman army in time of war. The result was that the Roman commonwealth possessed enormous reserves of military manpower, and in 338 was already the strongest military power in Italy.
>
> As it proceeded on its triumphant course, the Roman state expanded by adding an ever widening circle of dependent communities to the commonwealth. Defeated peoples were annexed with either full or partial citizenship, Latin colonies were founded, and an increasing number of states became allies. (Cornell 1995, p. 365)

Now here we have something completely different: the explanation put forward by Harris and his followers is an essentialist one – it was the Roman's nature to be so belligerent. Cornell offers a causal mechanism instead: By turning defeated enemies into allies and, in the long run, allies into citizens of their own state, the Romans created a system that was able to expand continually, because each successful integration of a former enemy into this system increased its military resources. One might describe this as a positive feedback loop and compare it to biological modes of growth.

Independent of Cornell, Arthur Eckstein (2006) has identified the same cause for the sustained expansion of Rome, but he also contributed another perspective to the debate which seems crucial. It is important because there had been the paradox of comparative science without comparisons; by claiming that the Romans were *exceptionally* bellicose, one makes a statement that inevitably requires a basis for comparison, but this issue had been neglected.

Of course, it is well known how many wars had been fought in the Greek world, and the many pieces of evidence for the Greek's appreciation of military glory are well known, too. And it is no secret that Athenian democracy had its origin in a total mobilization and militarization of the society. Nevertheless, this fact, which contributed to the aggressive stance of Athenian politics against other cities, is frequently overlooked, and it appears as if Athenian democracy arose by abstract reasoning in the lofty heights of political philosophy.

Such lines of evidence had not yet been discussed in context, from the broad perspective of a cross-cultural comparison on Rome, her Italian neighbors, and the Greek states as well as other states in the ancient world. By providing this broad comparison for the first time, by assembling a large amount of material on the role of war in the ancient world in general, rather than focusing on the Romans in isolation, Eckstein reaches conclusions which allow to see Roman militarism in a new light:

> One theme, however, has come to dominate modern scholarship on this problem: that Rome was exceptionally successful within its world because Roman society and culture, and

> Rome's stance toward other states, were exceptionally warlike, exceptionally aggressive, and exceptionally violent and not merely in modern terms but in ancient terms as well. [. . .]
>
> The present study takes a different approach. It applies to other ancient states the insights and method of analysis pioneered by Harris concerning Rome. It finds militarism, bellicosity, and diplomatic aggressiveness rife throughout the polities of the ancient Mediterranean both east and west. [. . .] Moreover, the present study finds the origins of the harsh characteristics of state and culture now shown to be not just Roman but common to all the ancient Mediterranean great powers, all the second-rank powers, and even many minor states as well, not so much within the specific pathological development of each state (what the political scientists call "unit-attribute" theory), but rather proposes that these characteristics were caused primarily (though not solely) by the severe pressures on all states deriving from the harsh nature of the interstate world in which they were forced to exist. (Eckstein 2006, p. 3)

This gives rise to the central question:

> The fundamental question is not why Roman society was militaristic and often at war, but why the Roman city-state was able to create a very large and durable territorial polity when so many other city-states failed at that task. Athens, Sparta, and Thebes all ultimately failed at it in European Greece; Carthage ultimately failed at it in North Africa; Syracuse failed at it in Sicily; Tarentum failed at it in southern Italy. (Eckstein 2006, p. 244)

And the answer put forward by Eckstein is essentially equal to Cornell's:

> It is not stern militarism but Rome's ability to assimilate outsiders and to create a large and stable territorial hegemony that makes Rome stand out from other city-states. (Eckstein 2006, p. 245)

Rome was not alone in this liberal attitude toward outsiders. Eckstein (2006, 246f.) points out that all the Latin cities had a liberal policy in this respect, facilitating, for instance, commercial exchange and intermarriage between their citizens and allowing citizens of other cities to buy property and settle within their boundaries. The Greek *poleis*, on the other hand, tended toward "virulent exclusivity" and tried to restrict access to their citizenry as far as possible. For them, it would have been unthinkable to do what the Romans did, i.e., to extend their citizen rights not only to the Latins, who at least shared the common language and culture, but also to real aliens such as the Etruscans, who were not even native speakers of Latin.

Rome was particularly favored by a location that facilitated trade and economic growth, and apparently for this reason it had much better starting conditions compared to all other Latin cities, but the other crucial aspect of the rise of Rome, i.e., the ability to integrate outsiders, was common Latin heritage. So Rome could outgrow all competitors in Italy by absorbing ever more allies into her political system, but merely absorbing them would not have been sufficient. Decisive was that the system proved stable even during major crises, and this was due to a policy that maintained at least a minimum amount of consent among the allies.

The importance of this aspect can be seen by direct comparison with other powerful city-states which built alliance systems that were, in principle, comparable to the Roman one (see Baltrusch 2008 for an overview of recent scholarship on ancient alliances and empire formation; especially on early Rome pp. 9–14) but

which were plagued by dissent and separatism. The maritime republic Carthage had, quite like Venice much later, a "*terra ferma*" in North Africa, i.e., a dominion that formed the basis for her overseas adventures. Besides the territory of the city of Carthage proper, there were a number of allies, including other cites that shared the common Phoenician origin but also the autochthonous Numidian and Libyan tribes. But the tensions between Carthage and all these neighbors were a constant theme in Carthaginian politics, and the necessity to maintain a large force to defend the homeland posed a limit on the military resources that were available for overseas operations, for instance, in Sicily. And with respect to Athens, Russell Meiggs has noted:

> In the second year of the Peloponnesian war, according to Thucydides, Perikles could admit to the Athenian assembly that their empire was a tyranny. This language has shocked some modern scholars; it would not have shocked contemporaries. They knew that Athenian rule did not rest on the free consent of the allies, and I suspect that they had known this for a long time. (Meiggs 1963, p. 1)

The fact that alliance systems were vulnerable to tensions between the allies was of course common knowledge, and when Hannibal invaded Italy during the Second Punic War, part of his overall strategy was the assumption that he would be welcomed as a liberator and that the Roman alliance system would fall apart. Some communities actually defected, but the overall system remained intact, much to Hannibal's disappointment. In this respect, he was merely repeating the experience of Pyrrhus two generations earlier, who was also faced with an essentially stable Roman alliance upon his invasion in southern Italy.

It was a principle of Roman policy to require only military service from the Italian allies. Military service was seen as honorable and it included the attractive prospect of getting a share of the spoils of war. By contrast, the Athenians initially required either military service and ships or the payment of tribute from their allies, but in the long run the demand for monetary contributions became dominant, which was a serious bone of contention. In ancient political thinking, paying tribute to another state was a sign of lost independence; this explains why this Athenian habit was thoroughly unpopular among the other members of the Delian League. The Athenians paid the price in the form of riots and, finally, dissolution of the league. And the Carthaginians frequently overstretched the patience of their allies by requiring both military service and substantial payments (Huss 2004, pp. 339–343).

Now their system enabled the Romans in a first step to create a stable hegemony in Italy. But this is not yet world domination, and the second step was the involvement of Rome in the affairs of the eastern Mediterranean from the second century BC onward. This involvement in the east finally gave rise to a unified Mediterranean world dominated by Rome, either by direct territorial incorporation as a *provincia* or by treaties and alliances. We cannot trace the events, spanning more than 300 years, in detail here. But it is interesting to look at the basic mechanism, because there is actually such a mechanism to be identified.

According to Eckstein, the decisive triggers were Greek calls for help, which received a positive reply from the Romans:

> Our study will then conclude with an analysis of the decision by the Roman Senate and people in 200 B.C. to answer the Greek states' calls for help against Philip V of Macedon and Antiochus III of the Seleucid empire. The two wars that followed this decision shifted the balance of power in the Mediterranean decisively in Rome's favor and brought Roman influence and power permanently into the Greek world. In a real sense, it laid the foundation of Roman political preponderance throughout the entire Mediterranean. Yet the decision itself was of the type that we have seen throughout this study was normal (not exceptional) for a great ancient state to make when confronted by requests for help from lesser states. (Eckstein 2006, 244f.)

The network of treaties and alliances that finally led to the Second Macedonian War is too entangled to be discussed in detail, but a crucial driving force were the Aetolians, who had a long-standing conflict with Macedon and who had a major interest in getting the Romans involved.

When this happened and the Romans defeated Philip V, the Aetolians were not satisfied, however. They had hoped for a large territorial increase at the cost of Macedon, which the Romans refused to concede them. The Romans tried to establish some sort of peace order that essentially maintained the status quo before the war. The Aetolians therefore switched alliances and induced now Antiochus III to make war in mainland Greece.

So we have a basic mechanism of large networks of linked powers, linked by either long-term treaties or immediate calls for help in a situation of urgency. It was an interstate system where hostile diplomacy, armed conflict, and the switching of sides were frequent, and the growth of ever larger systems of alliances created the danger that any local conflict could easily become a major war. There is nothing specific Roman here, these are features of the ancient interstate system at large, and it could act as an amplifier of even the smallest internal conflicts within a single city.

And it is also not correct, although one encounters such views frequently, that the Romans happily took the first opportunity to impose their order, their will in the Greek east. Rather, they showed a remarkable adaptability and followed, to a large degree, the political concepts and traditions of the Greeks. Erich Gruen (1984) has shown this by detailed analyses of the arrangements made by the Romans, their treaties and alliances with Greek states, once they had become involved in eastern Mediterranean affairs.

The causes of the Persian Wars and the Peloponnesian War, as narrated by Herodotus and Thucydides, respectively, were mentioned above. Let us now look at these examples.

The Persian Wars started with the Ionian Revolt, which ultimately arose from internal dissent on the island of Naxos. The aristocratic party of Naxos appealed to Aristagoras, the ruler of Milet, for help, and he in turn asked his Persian overlords for an army. The Persians supplied this army, but Aristagoras managed to start an argument with Megabates, the commander of the Persian expedition forces. Aristagoras was in an awkward position; he switched sides and warned the people of Naxos of the imminent attack – the attack that was due to his initiative – and

searched for further allies among other Greek cities along the Ionian coast. Since he felt that this was still not enough to confront the Persians in a conflict that had, by then, become a general Ionian uprising, he sought further allies on the Greek mainland. Sparta refused, which is a rare exception, but the Athenian people's assembly were enthusiastic when they learned about the royal treasury of the Persian king, and thus the Greek mainland became involved in a major war with the Persian Empire.

The Peloponnesian War started with internal conflicts at the city of Epidamnos, and the aristocratic party appealed for help to the Illyrian natives in the area. The city thus came under pressure and asked for help at the mother city of Kerkyra, which Epidamnos had once founded. It would have been a moral obligation to provide such help, but for unknown reasons Kerkyra declined the request. So Epidamnos had to ask Corinth for help, which, in turn, was the mother city of Kerkyra. At this point, the people of Kerkyra suddenly remembered how important Epidamnos was for them, which they considered their own possession. So the intervention of Corinth in Epidamnos was seen as an insult that could only be answered by war. For this reason, they appealed to Athens for help, which was readily granted. Corinth in turn appealed to Sparta for help, and so finally the two largest powers of the Greek world – Athens with the Delian League and Sparta with the Peloponnesian League – became opponents in a war that originated in the small city of Epidamnos at the semi-barbarian fringe of the Greek world.

Seen in the light of these events, the outbreak of the Second Punic War is not the perfect illustration of Roman imperialism, in contrast to how it is usually presented. Rather, it shows just all the features that appear familiar from the examples above; in the Iberian city of Saguntum, there was internal dissent and one party appealed to Rome, making her the arbitrator and protector of the city. In addition, Saguntum had conflicts with surrounding tribes who were allied to Hannibal. When open war between Saguntum and the tribe of the Turdetani broke out, the latter appealed to Hannibal, while the Saguntines sent envoys to Rome. That Saguntum was located south of the river Ebro and thus in a region that had been defined as Carthaginian, rather than Roman, zone of influence did not help to deescalate the situation either, and now the whole system was ready for a major war.

But why then did the large states agree to be drawn into the messy affairs of minor powers? They knew that far-ranging and destructive wars could ensue, and they frequently also knew that the legal or moral justification for their intervention was weak, as in the case of Athenian help for Kerkyra as well as of Roman help for Saguntum. But the irresistible benefits from the perspective of the large powers were always the same:

Reputation – it conferred prestige to be a widely accepted helper and arbitrator, and international prestige is a value in itself (for ancient as well as for modern governments).

(continued)

(continued)

Increased radius of operations – these calls for help were an optimum pretext for promoting one's own interests in distant regions, spoils of war and other material benefits included.

Competition – if you decline this request for an alliance, someone else will accept it and reap the benefits in your place.

So in these respects, Rome is typical rather than exceptional. But it was the rise of Rome to the single dominant power in the Mediterranean that effectively ended this violent interstate system with its frequent wars and destabilizing alliances.

Without doubt, Roman domination had its own adverse effects. In particular, outside of Italy the Romans did not continue their system of alliances without tribute payments. Rather, they imposed taxation on their provinces abroad, and the tax burden was probably the single most important cause for the riots and separatist movements that occurred in some instances in the Roman Empire. But in the overall balance, the empire remained remarkably stable, and the benefits of the *Pax Romana* seem to have been real, rather than being perceived as mere propaganda. The reasons for the empire's stability are manifold, but one aspect stands out, by direct comparison with the violent interstate system that had prevailed previously: the removal of the background of constant warfare brought in itself substantial economic benefits and enabled a period of general prosperity. To conclude this section, I will present three illustrative examples from the Roman east.

In 167 BC the federation of Lycian cities (in the southeastern corner of modern Turkey) had been declared "free" by the Romans – meaning that Rome became the protector of their independence from the former overlord Rhodes, after a series of Lycian revolts. The result was an enormous building activity, the traces of which can still be seen today (Marek 2010, 291f.). Around the same time (ca. 170 BC), the Pergamon Altar was built, one of the most impressive extant monuments of antiquity. Again chronology reveals the causal nexus: after his defeat by the Romans and the peace treaty of Apameia in 188 BC, Antiochus III had to withdraw from *Asia Minor*. Thus, the kingdom of Pergamon, not yet a Roman province but under protection of Rome, could recover from the previous wars, and an ambitious building program was started at the capital. A strikingly similar pattern is observed one century later, when the Near East had also come under Roman influence:

> What is now certain, however, is that Petra, as a city with monumental architecture and rock-cut facades, belongs in that period in the history of the Near East when Roman domination was assured, but Roman direct rule was either absent or still relatively lightly imposed. The royal monuments of Commagene belong in this period, if in the earliest phase of it; but also do the temple of Bel at Palmyra, Herod's Temple in Jerusalem (as well as his other major monuments), the temple of Baalshamin at Sia' and, as will be seen, the temple of Zeus at Gerasa. The major constructions of this period were sometimes royal creations, as in Commagene or Judaea, but others were expressions of the culture of local communities, as in Palmyra, Sia', or Gerasa. (Millar 1993, 407f.)

8.4 How to Run an Empire?

It should have become clear that a single master plan for world conquest is not the explanation for the emergence of the Roman Empire. The power and influence of Rome grew in a piecemeal fashion, as the sum of many individual episodes. And the patchy nature of the Roman possessions was also continued in the imperial era, when territories with quite different legal status became collectively called the *Imperium Romanum*. Even the question what the empire really was defies a clear modern definition: it was not simply the personal possession of a king, like the Hellenistic monarchies. It was also not simply the territory of the Roman people because the latter, i.e., the *ager Romanus* resp. the legally defined Roman homeland in Italy, was never expanded beyond central Italy. The best approximation of a legal definition in modern terms describes the empire as an alliance of cities with Rome as the senior partner. So the administrative structure reflects the mode of growth of the empire, and it had three levels: the cities with considerable local autonomy, the provinces, and finally the emperor. The latter two represent the "imperial" or "Roman" administration of the empire, but it is uncertain what this really means. Our sources do not explicitly explain the workings of the imperial administration; we are lacking texts that would provide an organization chart or handbook of administrative procedures. But what we do know is that the bureaucratic apparatus was small by modern standards. There were no large bureaucracies, neither in Rome nor in the provinces. The emperor ruled essentially with the help of a limited number of friends, advisers, and secretaries. And the same system of minimal government was repeated by the individual governors in the provinces, who also had only a small staff. Some scholars even deny that the modern notion of an administration applies at all to the *Imperium Romanum* (especially on Roman *Asia Minor,* see Marek 2010, 453f.; for a general discussion of the emperor's role, Millar 1992 is essential).

The absence of a professional bureaucracy indicates that the empire cannot be understood in terms of a modern state. A further example which provides evidence for this is the issue of Roman strategy. The problem is that there were no general headquarters of the army, no ministry of defense or state department, no secret services, no permanent embassies or professional diplomats, no institutes for political science or international relations, and no think tanks or military academies. The whole institutional framework that is essential in a modern state in order to formulate the strategic aims is lacking.

Nevertheless, Edward Luttwak published his "Grand Strategy of the Roman Empire" in 1976, and the debate about Roman strategy is still influenced by this work (see Campbell 2010; Heather 2010 for recent discussions).

Luttwak has analyzed the military arrangements of the Roman Empire, in particular with respect to the borders, and he considers three distinct phases of Roman strategy:

Phase 1 in the time of Julio-Claudian emperors (27 BC–68 AD): The system is based on client states and mobile armies, a broad buffer zone provides security, and the mobile army is concentrated at several points (Luttwak 1976, pp. 7–50).

Phase 2 from the Flavians till the Severans (69 AD–235 AD): A system of "scientific frontiers," i.e., short frontiers selected for optimum defensive qualities, and preclusive defense. The empire becomes a sort of fortress with precisely defined perimeter and limes fortifications (Luttwak 1976, pp. 51–126).

Phase 3 in the later third century until Diocletian (235 AD–305 AD): Fixed frontiers are abandoned and replaced by defense in depth; the mobile armies are located in the interior and serve as a large strategic reserve that can operate wherever needed (Luttwak 1976, pp. 127–190).

Some of the facts are uncontroversial. For instance, a number of semi-independent kingdoms were successively transformed into regular provinces during the course of the first century. However, the majority of scholars have denied that one can identify clear and elaborated systems of Roman strategic overall planning.

Any discussion on the Grand Strategy of the Romans must start with the basic evidence: the number of troops, the borders of the empire, and the distribution of the troops within that territory. But the problems start here already: We do not know, for instance, which factors limited the size of the Roman army, since ancient sources do not discuss these matters. In general, it is assumed that financial, rather than demographic, reasons set the limit for a comparatively small army of about 300,000 men under Augustus and his successors for the next 250 years or so. But this is just conjecture, and there is another possible explanation; a large army posed a potential political threat. In the final decades of the Republic, soldiers had repeatedly played an active role in military and political matters, and generals, including Octavian, had been forced by mutinies to follow the wishes of their soldiers (Keaveney 2007; Kienast 2009, 320ff.). Whether these experiences played a part when Augustus finally designed the army of the principate is unknown but conceivable at least. What is known is that in any case the number of troops garrisoned *in one place*, under the command of single provincial governor, was restricted. Large troop concentrations were frequently the crystallization points of attempts to usurp the emperor's purple by ambitious generals. In short, domestic policy may have been at least as important as the strategic response to external threats for decisions on troop distribution.

The notion of a coherent Grand Strategy of the Romans suffers also from the problem that the required institutions were not there, as discussed already. Implementing a strategic doctrine in the modern sense would have required the collection, colligation, and evaluation of large amounts of information – it is unclear who should have done this.

Another problem is the issue of maps. Looking at the Roman Empire from a bird's eye perspective, as we can do using modern maps or satellite pictures, certain features appear rational, such as the use of deserts, mountain ranges, or large rivers for defining the borders. But it is unknown whether this way of looking at things was available to the Romans, since it is contentious whether they used maps comparable to modern ones. This problem obviously affects the question why they made particular border arrangements.

The Roman *agrimensores* made exact measurements of the terrain on a small scale. These techniques were used, for instance, for building roads, for laying down the city plan of a newly founded *colonia*, or for constructing a military camp with its regular structure. These techniques were obviously also used to produce the giant map of the city of Rome from the time of Severus (ca. 200 AD). It is striking how this graphic representation appears familiar to a modern observer, almost two millennia later (Koller et al. 2005). So when the Romans were able to do such things for public display, would they not have used the same techniques for strategic planning?

But getting the topography of a city right is one thing, getting the topography of a whole continent right quite another. Under Augustus, there was also the world map of Agrippa on display in Rome – this may have been what we are looking for, but some scholars have denied that this was a real map and suggested that it was an *itinerarium* (Kienast 2009, p. 264, fn. 187), i.e., a list of places and the distances between them. Such itineraries served to break down the two-dimensional topography of an area into linear, one-dimensional relations between places.

Since the sources do not tell explicitly if and how maps were used for strategic purposes, other approaches are needed. Christian Hänger (2001) has analyzed ancient sources with respect to the geographical knowledge of the acting persons, insofar as it can be inferred indirectly from the text. When Tacitus reports, for instance, on the campaigns of Tiberius, Drusus, or Germanicus in Germany, his own account of the landscape is fragmentary and superficial. But, as Hänger has shown, the Roman generals' geographical knowledge must have been much better than the scanty fragments that made their way into the historian's account. The choice of optimum marching routes and the clever use of rivers as supply lines betray familiarity with the area of operations (Hänger 2001, 180ff.). According to Hänger, the Roman commanders must have possessed at least very precise *mental* maps. This is not yet proving the existence of *physical* maps, but it makes them plausible in any case (see also Sheldon 2005, pp. 148–150 on Roman strategy and maps. According to a recent analysis by Fedi et al. 2010, the so-called Artemidorus Papyrus is probably authentic; it apparently contains a map of the Iberian peninsula from the first century AD).

Thus, it seems possible to cure one important moot point of Luttwak's theory, but this does not end the discussion. One of the most severe criticisms came from Benjamin Isaac (1990), who has offered an alternative explanation for the military arrangements of the Roman Empire. Isaac rejects a Grand Strategy and a defensive

organization of Roman troops altogether. But his major point is not the question of a feasibility of Roman overall strategy in a technical sense – the issue of maps and so on – rather, he sees a completely different motivation at work; according to him, the Roman army served mainly aggressive purposes to enable the emperors to celebrate themselves as triumphant conquerors and also as an instrument of extortion and oppression of the subjected peoples. The influence of Harris' views is obvious.

Luttwak's perspective was biased because he tried to transfer concepts from modern strategic studies to the ancients. But Isaac's perspective appears likewise biased, because his conclusions on the particular area, i.e., the Roman Near East, cannot provide a model for the interpretation of the empire as a whole, with borders in vastly differing regions of the world, facing a multitude of different local conditions and challenges. The Near East was the place of the Jewish War and other Jewish revolts, but these events are the exception rather than the norm. The revolt of Arminius, for instance, that gave rise to the famous battle of the Teutoburg Forest was not some kind of Pan-Germanic and anti-colonial liberation movement. Rather, it was a military coup of an ambitious leader who calculated that being a Roman officer was nice, but being a Germanic king even nicer. In general, the Roman Empire is characterized by the rarity of riots that are "nationalist" in modern parlance.

The allocation of the legions shows impressively that large interior areas of the empire were virtually military free, in particular the peninsular areas of modern Spain, Greece, or Turkey. The army appeared to be concentrated in two particular border regions, the northern one along Rhine and Danube and the eastern one. This allocation cannot be reconciled with suppression of internal riots being the main purpose of the army.

Leaving aside Isaac's extreme view, his denial of any defensive considerations on the part of the Romans, a consensus exists among the critics of Luttwak's theory. The attempt to transfer the modern concept of a coherent, systematic, and long-term strategic overall planning is generally regarded as failed.

That Roman arrangements may have been piecemeal rather than systematic can be seen, for instance, by comparing the *limes* systems of the neighboring provinces *Germania Superior* and *Raetia*. Both are located in the same type of Central European landscape, and both have faced essentially the same type of enemies. Nevertheless, the Germanic *limes* consisted in its final stage of a wooden palisade, a ditch, a rampart, a connecting road, and, finally, on the inner side of the whole arrangement, a line of watchtowers. By contrast, the adjacent Raetian *limes* was formed by a stone wall into which the towers were integrated, while the connecting road was behind them. No plausible explanation for these differences in terms of different tactical necessities exists, and it seems that there were simply different regional traditions at work (discussion of various *limes* systems from the perspective of intelligence and communication in Sheldon 2005, pp. 199–249).

In philosophical terms, one might describe Luttwak's approach as a logical reconstruction. He took apart pieces from Roman military history and reassembled them to fit a modern strategy analysis. But since he used modern concepts and

notions, his work yields no genuine explanations for the Romans' decisions because their original perspective is no longer considered in his reconstruction:

> When Edward Luttwak [. . .] analyzed the defense policy of the Roman Empire using the vocabulary of modern military structures, he produced an interesting conceptualization for readers at the Pentagon, but I am still not convinced he brought us any closer to an understanding of the Roman mentality. Roman activities were often messy, unprofessional, and even unsuccessful, but we do less damage to the historical record if we leave them that way. Trying to incorporate them into a grand strategy that may not even have existed may be more satisfying to us intellectually, but it is ultimately less accurate. (Sheldon 2005, p. xvi)

So rather than assuming long-term and "scientific" planning, we should see Roman decisions as frequently being ad hoc, opportunistic, and based on the personal idiosyncrasies of individual emperors. It was, in particular, Fergus Millar (1982) who stressed the highly personalized form of Roman government, the lack of large and inert institutions, where the individual preferences of emperors necessarily became a major determinant of Roman politics and warfare. And these preferences were not always of a "rational" or "objective" nature. Rather, ideology and tradition, thinking in terms of glory and precedent, played a great role. The Roman emperors' desire to emulate Alexander the Great defies the attempt to analyze it using the tools of modern strategic studies, but exactly this desire was one major reason for the repeated wars between Rome and her eastern neighbors.

This constant worship of the classical tradition, of the glorious examples of the past, does not only affect Roman texts with respect to their usefulness for us. It may have likewise affected the contemporaries' ability to comprehend their own time. Herwig Wolfram has identified an "incorrect theory" with respect to Germanic peoples as one of the major causes for the difficulties the Romans experienced when the Germanic world set itself in motion in late antiquity. When thousands of Lombards crossed the Danube under Marcus Aurelius in 167 AD, the Romans knew – at least since the beginning of the Christian era – that their original places of settlement were at the lower Elbe River, hundreds of kilometers to the north:

> But nobody considered to draw any conclusions from this, such as asking, for instance, whether movements within Germany might have been the cause for the outbreak of the terrible fights, and if perhaps even worse things were to come. [. . .]
>
> Such an approach would have required an understanding of the barbarian world that was not available, rather, one liked to believe that there were no and could be no new barbarians. [. . .] How should a government in Rome have taken the correct preemptive measures, when its experts on foreign matters were literates, who designated Goths, Vandals and Huns alike as 'Scythians', equaling them to that people of the southern Russian steppe which in reality had become extinct long ago. For the situation at the Rhine, one similarly used traditional categories: Alemanni and Franks continued to be seen as Germans of the type encountered in the early imperial era or even – in accordance with the pre-Caesarian custom of the Greeks – as 'Celts' or 'Celtoscythians'. (Wolfram 1990, 70f., transl. SB)

How far this literary and antiquarian approach actually affected Roman decision making is a matter of debate. Susan Mattern (1999, pp. 1–2) pointed out that the

literates, who produced such works, were frequently *amici* of the emperor, i.e., persons very close to the center of power. Thus, she concluded that the level of strategic thinking found (or rather not found) in their works reflects the debates among the emperor and his companions. However, as one reviewer of her book has argued (see Sidebottom 2003), exactly because the tendency of literary stylization in such texts is known, we cannot use them to make direct inferences about the debates that were actually led.

In any case, Kimberly Kagan has warned that one should not overstretch the criticism of Luttwak's book and she pointed out that when the definition of Grand Strategy becomes too narrow, it finally becomes useless even for discussing the behavior of modern states. In particular, she noted that the crucial aspect is not how far the Romans conformed to modern definitions of strategy. Rather, one should look at what they actually did:

> The patterns of troop movements also show clearly that imperial decision-making about grand strategic issues occurred even without visible long-term planning. Emperors restored distributions even after intervening events such as wars, rebellions, and imperial successions had disturbed them. [...] Emperors worried about the stability or security of provinces when they conducted major operations. The successive replacement of legions moving off to war shows that emperors thought about how their activities on one frontier (in crisis or for conquest) affected the whole empire. Emperors made decisions about how to allocate resources to meet objectives empire-wide, and thus definitely thought about grand-strategic issues. The grand strategy of the Roman Empire can be studied as long as we ask questions that the available sources support. (Kagan 2006, p. 362)

Besides troop movements, the issue of bridges is a further example: For long stretches of time, there were no bridges across the two major border rivers in Europe, the Rhine and the Danube. The Roman technical expertise to build such bridges is out of question, and the deliberate refusal to build them is even more astonishing given their potential benefits; they would not only have facilitated commercial exchange in itself, they would also have facilitated the collection of import and export taxes by channeling this traffic. Nevertheless, the function of large rivers as an obstacle for barbarian intruders was obviously ascribed a higher value than other criteria.

In conclusion, it is clear, on one hand, that the Romans were not completely blind to reality; they included issues of defense and security into their plans. On the other hand, the concept that the Romans followed coherent and rational strategic systems with long-term and centralized planning is rejected. In particular, the idea that there was a succession of three distinct, clearly defined military doctrines is generally regarded as the weakest point of Luttwak's analysis. The shape of the empire, as it emerged during its history, may appear rational to us, with respect to the choice of river borders and so on, but it is, after all, a product of chance, the accumulation of numerous decisions which were made for a whole range of different reasons. What we have here, in effect, is a beautiful analogy to results of biological evolution. It may look like a product of rational design, but it is a product of chance nevertheless.

8.5 Concluding Remarks

Two central questions in Roman history appear quite similar from a formal perspective:

1. How can we explain the rise of Rome, which policies and strategies were used by the Romans when they conquered their empire?
2. How can we explain the shape and structure of this empire, which policies and strategies were used by the Romans to defend and/or enlarge it?

Despite the formal similarity of both questions, there are deep differences with respect to the tasks that modern historians have to confront. In the case of the rise of Rome, there is abundant ancient evidence on the detailed course of events. One may even say that there is an overabundance of evidence because so many individual events are known that somehow may have contributed to this process. Thus, when a historian wants to explain the rise of Rome, a major part of the task is to select the pieces which are crucial, to form a coherent narrative from the available material by declaring some points as relevant, but others as irrelevant.

In contrast, the issue of Roman strategy once the empire had emerged is quite another story because no ancient sources allow for a direct insight into a possible overall strategy of the Romans. Therefore, in this case the historian's task is to assemble as many clues as possible from the fragmentary literary and material evidence. The various problems concerning ancient sources, as discussed above, are therefore more pertinent here.

One tool that plays a role in both fields is the comparative approach. This should be stressed, because comparative approaches, like the related issues of world history or universal history, are still a kind of fringe activity in modern historiography, which is still organized along traditional boundaries of eras and regions. To explain the rise of Rome, the direct comparison with other ancient states has destroyed the myth of a particular Roman attitude toward war. In addition, comparisons with the modern world occur in both fields of inquiry. For the issue of the rise of Rome, the application of the notion of an interstate system, borrowed from modern international studies, has proved fertile. In contrast, the transfer of the concept of a Grand Strategy to the Roman Empire, borrowed from modern strategic studies, was not successful in the eyes of a majority of scholars of the ancient world.

But even such failure can have its merits. The comparative method can either demonstrate some underlying common principles or serve to highlight the differences between apparently similar trajectories and phenomena of history.

To illustrate this point and to conclude this chapter, we will have a look at ancient China. A comprehensive general theory of ancient empires does not exist yet, since such broad perspectives are rather unusual and unpopular in historiography (Pomper 2005 is focusing on modern empires and does not even mention the

Imperium Romanum; Münkler 2005 has a few observations pertinent to the theme). Nevertheless, a number of scholars are working in this direction; Walter Scheidel, who is one of the pioneers in the field and has recently edited a whole volume dedicated to the detailed comparison of Rome and China (Scheidel 2009b, see also Morris and Scheidel 2009), notes:

> Two thousand years ago, perhaps half of the entire human species had come under the control of just two powers, the Roman and Han empires, at opposite ends of Eurasia. Both entities were broadly similar in terms of size. Both of them were run by god-like emperors residing in the largest cities the world had seen so far, were made up of some 1,500 to 2,000 administrative districts, and, at least at times, employed hundreds of thousands of soldiers. Both states laid claim to ruling the whole world, *orbis terrarum* and *tianxia*, while both encountered similar competition for surplus between central government and local elites and similar pressures generated by secondary state formation beyond their frontiers and subsequent 'barbarian' infiltration. (Scheidel 2009a, p. 11)

The point at the end of this quotation is a very interesting one: the process of secondary state formation in loosely integrated tribal societies, which arises from the contact of a large, centralized empire with its barbarian neighbors, is one of the most important recurrent features in the history of the premodern world. There is no doubt that both Roman and Chinese studies could benefit from a cross-cultural perspective that, in the case of Rome, may contribute to a causal explanation of the late-antiquity phenomena known as "Völkerwanderung" in German or "Dark Ages" in English.

But returning to the issue of formation of the Roman Empire, a comparison with the first unified Chinese Empire is also illuminating. Both empires arose within a violent interstate system of competitive states engaged in unstable alliances and constant warfare among each other. The Chinese era between the fifth and the third century BC is aptly called "Warring States," and incidentally, it covers a span about as long as the formative time between the First Punic War and the start of imperial government under Augustus. The state of Qin finally emerged as a superpower with a military potential that had no equals, and its ruler was from 221 BC onward the sole ruler of the Chinese world.

Despite these similarities, the means for maximizing the military potential were quite different in both cases. As we have seen, Rome grew in an *extensive* mode, by adding communities to the political system. There was no forced homogenization, communities retained a substantial degree of internal autonomy, and accordingly, the central administration was small during both republic and empire. It was small compared to a modern state, but also compared to the first Chinese Empire, because Qin became the most powerful state in its world by *intensive*, rather than extensive, growth:

> Central to the Qin reforms was the grouping of the population into units of five households that were each responsible not only for providing the squads of five recruits that formed the building blocks of Qin armies but also for mutual surveillance. [. . .] To ensure that the maximum amount of land was brought under cultivation, Qin also penalized households with adult sons living at home. These penalties forced sons to establish independent households and to cultivate their own allotments of land in order to support them. In tandem with this step, Qin also divided its territory into a grid of blocks, each of which was sufficient

> to support a family from the food produced on it. This reshaping of the countryside in order to ensure the maximum extraction of the resources for war was given physical expression through a system of paths forming a rectangular grid over the crop lands of the state. Finally, the government financed its war making through a head-tax imposed on the population.
>
> Qin carried out this vast effort at social and economic engineering through the creation of an equally extensive administrative apparatus. [. . .] To control this system, Qin established a bureaucracy capable of extending the central government's reach down to the local level. (Rosenstein 2009, 25f.)

So despite similarities in the ways to empire, it seems that some deep-rooted differences between Western and Chinese culture actually may be accounted for by explanations that date back 2,000 years. While in Western political discourse the ideal of plural, independent entities plays a central role (see, e.g., the hostile reactions in all European countries against the concept of further integration and unification), Chinese politics – imperial or communist – was always much more focused on the ideal of a strong and heavily centralized state. And it is certainly no coincidence that Fernand Braudel, who coined the expression *longue durée* for such long-term continuities, chose the Mediterranean world as a prime example in his studies.

References

Ancient Sources

Herodotus, Histories (Histories apodexis)
Livy, History of Rome (Ab urbe condita)
Polybius, Histories (Historiai)
Thucydides, Peloponnesian War (Ho polemos ton Peloponnesion kai Athenaion)

Modern Works

Ando, C. (2008). Decline, fall, and transformation. *Journal of Late Antiquity, 1*(1), 31–60.
Baltrusch, E. (2008). *Außenpolitik, Bünde und Reichsbildung in der Antike*. München: Oldenbourg.
Berry, S. (1999). On the problem of laws in nature and history: A comparison. *History and Theory, 38*(4), 121–137.
Berry, S. (2008). Laws in history. In A. Tucker (Ed.), *A companion to philosophy of history and historiography* (pp. 162–171). Malden: Blackwell.
Campbell, D. B. (2010). Did Rome have a grand strategy? *Ancient Warfare, 4*(1), 44–49.
Cornell, T. J. (1995). *The beginnings of Rome: Italy and Rome from the bronze age to the Punic wars (c. 1000–264 BC)*. London: Routledge.
Cowan, R. H. (2009). *Roman conquests: Italy*. Barnsley: Pen & Sword Military.
Demandt, A. (1989). *Die Spätantike: Römische Geschichte von Diocletian bis Justinian 284–565 n. Chr*. München: C.H. Beck.
Eckstein, A. M. (2006). *Mediterranean anarchy, interstate war, and the rise of Rome*. Berkeley: University of California Press.

Fedi, M. E., Carraresi, L., et al. (2010). The Artemidorus papyrus: Solving an ancient puzzle with radiocarbon and ion beam analysis measurements. *Radiocarbon, 52*(2), 356–363.
Fitzpatrick, M. P. (2010). Carneades and the conceit of Rome: Transhistorical approaches to imperialism. *Greece & Rome, 57*(1), 1–20.
Gruen, E. S. (1984). *The hellenistic world and the coming of Rome*. Berkeley: University of California Press.
Gruen, E. S. (2004). Material rewards and the drive for empire. In C. B. Champion (Ed.), *Roman imperialism: Readings and sources* (pp. 30–46). Malden: Blackwell Publishing . (Reprinted from *Imperialism of mid-republican Rome* (pp. 59–82). In W. V. Harris (Ed.), 1985, Rome: American Academy in Rome).
Hänger, C. (2001). *Die Welt im Kopf: Raumbilder und Strategie im Römischen Kaiserreich*. Göttingen: Vandenhoeck& Ruprecht.
Harris, W. V. (1979). *War and imperialism in republican Rome, 327–70 BC*. Oxford: Clarendon Press.
Heather, P. J. (2010). Holding the line: Frontier defense and the later Roman Empire. In V. D. Hanson (Ed.), *Makers of ancient strategy* (pp. 227–246). Princeton: Princeton University Press.
Huss, W. (2004). *Die Karthager*. München: C.H. Beck.
Isaac, B. H. (1990). *The limits of empire: The Roman army in the east*. Oxford: Clarendon Press.
Kagan, K. (2006). Redefining Roman grand strategy. *Journal of Military History, 70*(2), 333–362.
Keaveney, A. (2007). *The army in the Roman revolution*. London: Routledge.
Kienast, D. (2009). *Augustus: Prinzeps und Monarch*. Darmstadt: Primus.
Koller, D., Trimble, J. et al. (2005). Fragments of the city: Stanford's digital forma Urbis Romae project. *Journal of Roman Archaeology*, (Supplement 61), 237–252 (see also http://graphics.stanford.edu/projects/forma-urbis/)
Luttwak, E. N. (1976). *The grand strategy of the Roman Empire, from the first century A.D. to the third*. Baltimore: The Johns Hopkins University Press.
Marek, C. (2010). *Geschichte Kleinasiens in der Antike*. München: C.H. Beck.
Mattern, S. P. (1999). *Rome and the enemy: Imperial strategy in the principate*. Berkeley: University of California Press.
Meiggs, R. (1963). The crisis of Athenian imperialism. *Harvard Studies in Classical Philology, 67*, 1–36.
Millar, F. (1982). Emperors, frontiers and foreign relations: 31 B. C. to A. D. 378. *Britannia, 13*, 1–23.
Millar, F. (1992). *The emperor in the Roman world*. London: Duckworth.
Millar, F. (1993). *The Roman near east: 31 BC – AD 337*. Cambridge: Harvard University Press.
Morris, I., & Scheidel, W. (2009). *The dynamics of ancient empires: State power from Assyria to Byzantium*. Oxford: Oxford University Press.
Münkler, H. (2005). *Imperien: Die Logik der Weltherrschaft vom Alten Rom bis zu den Vereinigten Staaten*. Berlin: Rowohlt.
Pietrykowski, J. (2009). *Great battles of the hellenistic world*. Barnsley: Pen & Sword.
Pomper, P. (2005). The history and theory of empires. *History and Theory, 44*(4), 1–27.
Rich, J. (2004). Fear, greed, and glory: The causes of Roman war making in the middle republic. In C. B. Champion (Ed.), *Roman imperialism: Readings and sources* (pp. 46–67). Malden: Blackwell Publishing, (Reprinted from *War and society in the Roman world* (pp. 38–68). In J. Rich, & G. Shipley (Eds.), 1995, London: Routledge)
Rosenstein, N. (2009). War, state formation, and the evolution of military institutions in ancient China and Rome. In W. Scheidel (Ed.), *Rome and China: Comparative perspectives on ancient world empires* (pp. 24–51). Oxford: Oxford University Press.
Scheidel, W. (2009a). From the "great convergence" to the "first great divergence": Roman and Qin-Han state formation and its aftermath. In W. Scheidel (Ed.), *Rome and China: Comparative perspectives on ancient world empires* (pp. 11–23). Oxford: Oxford University Press.
Scheidel, W. (Ed.). (2009b). *Rome and China: Comparative perspectives on ancient world empires*. Oxford: Oxford University Press.

Sheldon, R. M. (2005). *Intelligence activities in ancient Rome: Trust in the gods, but verify*. London: Routledge.
Sidebottom, H. (2003). Review of "Rome and the enemy. Imperial strategy in the principate" by S. P. Mattern. *The Classical Review, 53*(1), 172–174.
Wolfram, H. (1990). *Das Reich und die Germanen: Zwischen Antike und Mittelalter*. Berlin: Siedler.

Chapter 9
Causal Explanation and Historical Meaning: How to Solve the Problem of the Specific Historical Relation Between Events

Doris Gerber

Abstract History is no mere chronicle of events. This insight of Arthur C. Danto's (often misunderstood) discussion of the concept of history implies that the historical meaning of a past event can change in the course of time – simply because of what happens afterwards. If we hold, however, that history has a real structure and that the historical meaning of past events is determined by the causal and temporal structure of these events, then we have to be able to show how the historical meaning of past events can be causally explained. And how can this be shown without presupposing the highly controversial thesis of backward causation? After discussing Danto's thesis at some length, I argue first very generally in favour of a counterfactual analysis of causality and, second, that an expansion or revision of this analysis can solve the problem of this specific historical relation between events.

Keywords Historical explanation • Historical meaning • Counterfactual causality

9.1 Introduction

Histories are not mere chronicles of events, or so emphasizes Arthur Danto in his book *Analytical Philosophy of History*. Even the so-called Ideal Chronicler who knows whatever happens the moment it happens, and has the gift of instantaneous transcription, would be unable to tell a history because he would be unable to construe the historically relevant relations between the events. Nevertheless, he can describe the course of each event's occurrence in full detail. The issue Danto is pointing out through his fictional Ideal Chronic and his concept of narrative sentences – this means sentences in which one event is described from the

D. Gerber (✉)
Universität Tübingen, Philosophisches Seminar, Bursagasse 1, 72070 Tübingen, Germany
e-mail: doris.gerber@uni-tuebingen.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_9,

perspective of another temporally later event – is obviously relevant to the problem of explanation in history in that past events have a property, which we can call their historical meaning, and this historical meaning can change over the course of events, simply because of what happens afterwards. And this fact, the fact that the historical meaning of past events can change over the course of time, challenges the thesis that historical events can have a causal explanation because if an event's historical meaning can change in virtue of what happens afterwards, then it seems to be that we have to accept the possibility of backward causation if we want to insist that this historical meaning is a real property, which is causally determined and therefore can be causally explained. Now, some philosophers are convinced that some kind of backward causation cannot be conceptually excluded; I think, however, that the relevance of causal explanations in history could not and should not depend on the controversial possibility of backward causation.

Therefore, my goal is to show that the historical meaning of past events can be causally explained without supposing backward causation, but instead by revising or expanding the concept of counterfactual causality. First, I will discuss Danto's well-known example of two scientists who supposedly formulated the same scientific theory independent of each other and with great temporal distance between their respective actions. Second, I attempt to clarify the concept of a historical meaning by stressing the underlying problem in Danto's discussion which, in my opinion, is the distinction between the historical meaning of events on the one hand and the semantic meaning of linguistic expressions and sentences on the other hand. In the third section of the discussion (Sect. 9.4), I argue for a counterfactual theory of causality assuming that these arguments are free of the particular problem of a specific historical connection between events that I am concerned with. Lastly, I will end by coming back to this problem and propose how it can be solved by revising in two respects the traditional counterfactual analysis of causality proposed and developed by David Lewis.

9.2 Danto's Scientists

Arthur Danto's example of the two scientists is set within the context of his discussion about the characteristics of an Ideal Chronic. An Ideal Chronic entails every possible piece of truth with regard to every event and all the information which can be transcribed in the moment it happens. This means that the Ideal Chronic describes every event in full detail but without reference to earlier or later events. It represents, as you may put it, the happenings one by one over the course of time, including only the information that is true for the events in the moment that they occur. Such a Chronic is both very rich and very poor, and it seems to be clear why a Chronicler's transcription of happenings cannot tell a history: Histories essentially represent the relations between events, describing events not one by one, but within their relations. It is exactly this essential property of histories that Danto's fictional Ideal Chronic cannot possess.

Danto's puzzling example of two scientists formulating the same theory independent of each other articulates these conceptual correlations: 'Suppose, for example, that a scientist S discovers a theory T at t-1. S perhaps does not publish T. At some later time t-2, a different scientist S* independently discovers T, which is now published and included into the body of accepted scientific theories. Historians of science subsequently find out that S really hit on T before S*. This need take away no credit from S*, but it allows us to say, not merely that S discovers T at t-1, but that S *anticipated* at t-1 the discovery by S* of T at t-2. This will indeed be a description of what S did at t-1, but it will be a description under which S's behaviour could not have been witnessed and it will be an important fact about the event which accordingly fails to get mentioned by the Ideal Chronic' (Danto 1965, pp. 155–156).

What is going on here? What is the problem and what has this problem to do with causality? The puzzling issue is the fact that the first event, the formulation of T by the first scientist, S, seems to acquire a new property, the property of being the anticipation of T, in virtue and only in virtue of the occurrence of a later event, namely, the formulation of T by the second scientist, S*, at t-2. At t-1, when S discovers T, this act of discovering *is* no anticipation *yet*. It only *becomes* an anticipation when S* rediscovers T. It is not an anticipation at t-1 because it also would not have been an anticipation at t-2 if S* had not rediscovered T at t-2. Because and only because S* rediscovered T at t-2, the first event becomes an anticipation, and therefore, it cannot be an anticipation at t-1.

Does all this mean, however, that the past can actually change? And does all this mean that the second temporally later event is a cause or a kind of cause of the former event? Danto confesses that there is a sense in which we could say that the past is changing. However, what Danto explicitly wants to exclude is backward causation: '... there is a sense in which we may speak of the past as changing; that sense in which an event at t-1 acquires new properties not because we (or anything) causally operate on that event, nor because something goes on happening at t-1 after t-1 ceases, but because the event at t-1 comes to stand in different relationships to events that occur later' (Danto 1965, p. 155).

Now, Danto's discussion, as far as I understand it, starts getting rather complicated and very unclear. Danto formulates that there is no sense in which anything can in any way causally operate on past events. Yet he also says that it is possible that these past events form different relationships with events that occur later. How shall we understand this last assertion? What could these 'different relationships' be unless causal relationships if the past could change in virtue of these different relationships? Although Danto rejects the possibility of backward causation, he nevertheless introduces the distinction between necessary and sufficient conditions for events and contends that if a former event, E-1, at t-1 is a necessary condition for a later event, E-2, at t-2, then it follows that E-2 at t-2 is a sufficient condition for E-1 at t-1. However, in so far as these so-called conditions are really conditions *for events*, we have to understand them as factual conditions and that means we have to accept them as causal conditions. But this seems to suppose that we have two different concepts of causality in the discussion, namely, causal conditions and

proper causes. Now, the question would surely be: What is the criterion to make this distinction? Danto does not formulate and therefore does not answer this question. Instead, he emphasizes the connection between such conditions and the level of description. And it is exactly this shift in Danto's discussion, the shift from the factual level and the question of whether the past itself can change, to the level of description which is, in my opinion, not coherent. To illustrate the relevant quotation again in full detail: 'A sufficient condition for an event may thus occur later in time than the event. We cannot readily assimilate the concept of cause to the concept of necessary and sufficient conditions unless we are prepared to say that causes may succeed effects. So it is difficult to suppose that E-2 *makes* E-1 happen. But at the very least it permits a *description* of E-1 under which E-1 could not have been witnessed and which, accordingly, could not have appeared in the Ideal Chronic' (ibid).

Danto is surely right to say that our descriptions of past events are becoming richer and richer over the course of time simply because of what happened afterwards. But the crucial question in his puzzling example of the two scientists is whether the earlier event, E-1, can really acquire new properties in virtue of the occurrence of E-2 at t-2. It is unquestionable and therefore not very interesting that the truth of our description of E-1 as an anticipation of T depends on the occurrence of E-2 at t-2. It would simply be false to describe E-1 as an anticipation of T if E-2 never happens. However, the interesting question is whether E-1 really gets into, as Danto himself puts it, different relations to later events, that is, whether E-1 really acquires new relational properties at the time of the occurrence of E-2.

It might be a bit unfair to accuse Danto of having confused the factual level with the level of description because it seems that all Danto wants to show with his puzzling example is that the Ideal Chronicler cannot use words that express causal relations. Causes, as he emphasizes, 'cannot be witnessed *as* causes' (Danto 1965, p. 157). Danto mentioned that David Hume pointed this out long ago. However, Hume's argument for this contention is very different from the reason why the Ideal Chronicler is unable to use the word 'cause' or other synonymous expressions. Hume insisted that all we can really observe are mere regularities; but the Ideal Chronicler who transcribes the occurring events instantaneously is even unable to describe regularities, whatever sorts of regularities there may be. And my crucial point is that all this leaves the question open as to how we can conceptualize the fact that past events can change their relational properties over the course of time and in virtue of the occurrence of later events.

9.3 Historical and Semantic Meaning

Concerning histories, the aforementioned distinction between the factual level and the level of description refers to the difference between historical and semantic meaning. In its broadest sense, the concept of historical meaning expresses a property that every event that is part of a distinctive history possesses. That means

that every historical event possesses any historical meaning, simply by virtue of being a historical event. Historiography, however, is interested especially in such events that are endowed with a historical meaning which is outstanding in some respect. 'Being the anticipation of a later famous theory' is, in my opinion, a typical example of the historical meaning of an event. Other examples are 'being the final trigger of the war', 'being the first democratic election in this country', 'being the beginning of political disturbances' or 'being a great discovery'. I accept and want to defend the thesis that such historical meanings are real properties of events or are real properties of, more or less, complex connections of events. I also want to argue for the thesis that the historical meaning of an event is determined by the causal role that this event occupies. The causal role in turn is determined by the totality of the causal relations this event holds to other events, that is, by the totality of causes and effects concerning this event. Every event stands in at least some causal relation to other events. Thus, one can roughly say that the event's historical meaning is especially ample and important if this event is causally related to many other events and if these or some of these connections are temporally and spatially rather far-reaching. For example, the shooting of Archduke Franz Ferdinand in Sarajevo, considered as one trigger of the beginning of the First World War, surely has an important and decisive historical meaning exactly because its causal scope was so varied and far-reaching. If these shots can indeed be justified as a necessary but not solely sufficient cause of the First World War, then this single event is causally responsible for a war that lasted four years and was characterized by a hitherto unknown extent of cruelty in warfare. Whether this particular event, the shooting of Archduke Franz Ferdinand, was actually a cause of the First World War is no easy question. It is, however, surely right that the answer to this question does not depend on our descriptions but on the real properties of this event. And in this context, the crucial properties are the causal properties. That is, the event's historical meaning simply consists in the event's causal relations.

This realistic thesis concerning the historical meaning of past events stands in sharp opposition to narrative constructions of the concept of history. Arthur Danto is sometimes considered to be a kind of mentor of such narrative constructions which Hayden White and Frank Ankersmit prominently hold. In my opinion, however, the metaphysical consequences of Danto's discussions about the concept of history and the problems of explanation in the science of history are far from being clearly antirealistic. The realistic picture I want to defend is at least compatible with Danto's view of a history.

Although Danto speaks of necessary and sufficient conditions for events itself on the one hand and at the same time of necessary conditions for events being correctly describable as causes on the other hand, he is, as I understand him, very conscious of the fact that descriptions depend on the occurrence of the events they are describing and not vice versa. He explicitly emphasizes that only the occurrence of E-2 from our example permits a description of E-1 as an anticipation of T. But what does this 'permission' of the description imply? Is it also adequate to say that the occurrence of E-2 itself makes the description of E-1 as an anticipation of T *true*? Nothing that can be observed or witnessed during the occurrence of E-2 would

show that this event is a *rediscovery* of T. However, to describe E-2 as a rediscovery of T seems to be a precondition for describing E-1 as an anticipation of T. E-1 is an anticipation of T only in relation to E-2 and vice versa: E-2 a rediscovery of T only in relation to E-1. This is the case because 'being an anticipation' and 'being a rediscovery' are relational properties that imply causal relations, even if, as it is supposed in Danto's fictional example, the respective scientists do not know anything about each other and their respective theories.[1] This means that the truth of the description of E-1 as an anticipation of T depends not only on the occurrence of E-2 but also on the relation held between E-1 and E-2. 'Being an anticipation' is a property that is determined by the relational, i.e. by the causal properties of the event possessing such a property. For 'being an anticipation' necessarily implies that there is a connection to a different event. And how can we conceptualize this connection as anything other than a causal relation?

At this point, one may object that I am simply stipulating that there is a real relation between E-1 and E-2 at all which is established by the occurrence of E-2. Was not this exactly the questionable issue in Danto's example? Narrativists would certainly contend that there is no *real* connection but that we, as historians, are only construing such a relation by describing the first event as an anticipation and the temporally later event as a rediscovery. I would contradict this. Suppose, for example, that no one ever observed or witnessed the first scientist formulating a theory at t-1, which some hundred years later, after E-2 at t-2, becomes a published and famous theory. The first scientist's detailed notes lay for years undiscovered in a shed which, unfortunately, burns down many years before the second scientist formulated his theory. Nobody knows and nobody could ever come to know anything about the first scientist's pioneering work. It would nevertheless be perfectly true that he achieved this pioneering work. It is true that the theory's first formulation was an anticipation and that the second formulation of the same theory was a rediscovery, independent of what we or anyone else know or could know about the two events. This means that the historical meaning of past events is independent of our descriptions or interpretations. Our descriptions do not construe any historical relations, but they refer to such relations, which are determined by the causal relations of the respective events and exist independently of what we know or assert about these events or their relations. To reject this thesis is, in my opinion, tantamount to confusing the property of historical meaning, which is a property of events, with the property of semantic meaning, which is, of course, a property of linguistic expressions.

Until now, I have said nothing at all about the concept of causality that I hold and want to defend. But this question certainly needs some clarification, although it cannot be discussed in any detail. Therefore, I will now address this issue before I return to the special problem of the connection between causal and historical relations.

[1]For the sake of historical truth, it may be adequate to mention that concerning the real protagonists of Danto's example, namely, Aristarchus and Kopernikus, this condition is not met. Kopernikus was acquainted with Aristarchus' work.

9.4 The Concept of Causality

The question of whether and in which sense causal explanations are relevant in the human and social sciences has evoked controversial debates since the first theories in these sciences were developed. I have the impression that in the last years, the significance of causal explanations has been gaining ground. Much of the former or still existing scepticism against the importance of causality in the human and social sciences is justified by the characteristic that these sciences are mostly concerned with the explanation of human actions and that actions have special features, which leads to the consequence that they cannot be causally explained. Of course, the events in Danto's example of the two scientists also consist in actions, namely, the respective intentional formulation of a theory by two rational persons. I would contend that all historical events are action events because the concept of history is essentially connected with real possibilities, and this in turn presupposes that historical events are essentially connected to the phenomenon of intentionality. That is, because histories essentially imply possibilities, only events which have intentional properties and likewise the capacity to be causally efficacious can be historical events. And only action events can fulfil both conditions, or so this line of argumentation contends.[2] Before turning to the general problems concerning the concept of causality, it is therefore worthwhile to briefly discuss some of the main suspicions against the importance of causal action explanations, which are provoked by the supposed characteristics of intentional actions. The first of these suspicions refers to the problem of regularity, the second refers to the question of whether causality consists in a kind of causal mechanism, the third is represented by the so-called logical connection argument, and the fourth concerns the problem of mental causation.

The problem of regularity has an overwhelming significance in the debate about the possibility of causal action explanations. Here, the objection which is often emphasized is that human actions may show some kind of regularities, but certainly not strict and lawlike regularities. It is said that the behaviour of rational persons can be prognosed at least with some probability, but there is no possibility of a definite prediction. This objection, however, presupposes a specific concept of causality, namely, David Hume's view of causality as strict regularity. Hume has argued that causality is nothing more than regularity because if we are trying to observe causal relations, all that we can really observe are mere regularities between types of events. And these regularities must be strict or lawlike regularities because the criterion to distinguish between causal and, for example, temporal regularities in Hume's opinion is necessity. However, even in the contemporary natural sciences, it is widely admitted that, as regards natural events, strict regularity also is a requirement which cannot be met by all types of events. In the philosophy of natural sciences, this admission does not imply rejecting the concept of causality altogether.

[2]For a more detailed version of the argument, see Gerber (2012), Chap. VII.

Instead, it provides a platform from which to develop new approaches that lay beyond Hume's contentions. This means that the discussions in the philosophy of natural sciences show that the problem of regularities is a general problem, which does not impose any *special conceptual* problems on the explanation of actions.

The second objection is rooted in the intuition that causality is or represents a kind of blind mechanism being located on the presumably deepest level of reality, namely, merely on the physical level. The causal course of events is understood to be a mere course of unconscious happenings, whereas actions have reasons and are performed by persons who have desires, wishes, and intentions. A. I. Melden, for example, has expressed this intuition by saying: 'The agent confronting the causal nexus in which such happenings occur is a helpless victim of all that occurs in and to him' (Melden 1961, p. 129). Donald Davidson responded to this claim somehow desperately: 'Why on earth should a cause turn an action into a mere happening and a person into a helpless victim?' (Davidson 1980, p. 19). Davidson suspects that Melden's view implies a kind of doubling of the agent. He argues that although agency surely requires an agent, there are agentless causes and that the states and changes of states in persons are exactly such causes. Melden, however, would not have been convinced by this critique. He would have insisted that precisely these states and changes in persons, which are causes, transform the agent into a helpless victim. I think that the only way that Melden's concern can be rejected is by arguing that causality is no blind mechanism because it is no mechanism at all. What should a *general causal mechanism* consist of? To suppose the existence of such a mechanism is identical to the senseless attempt to search a cause for a cause. Of course, there are various kinds of 'mechanisms', i.e. causally efficacious properties which operate or function in various types of events at various levels of natural and mental phenomena. However, to describe such mechanisms in more or less full detail is nothing more than to redescribe the event itself and to describe it as a cause.

The third objection is also connected particularly with A. I. Melden's name, but others have also supported it, for example, Georg Henrik von Wright.[3] The so-called logical connection argument asserts that there cannot be a causal relation between actions and their reasons because there is a logical connection between them and the existence of a causal relation presupposes that the relata of such a relation are logically independent from each other. It was often emphasized in the discussions about this argument that it is far from clear how we should understand the respective claims of necessary logical connectedness or independence. I think the underlying fault in this argument concerns the distinction between logical relations of concepts and essential relations of events. If there is a logical connection or interdependence between concepts, then it is nevertheless not the case, as the argument tacitly implies, that the essential connections between the respective events or states covered by these concepts cannot be distinct from each other. I confess that actions are essentially connected with their reasons; moreover, I would say that actions are essentially connected with their intentions, which means that every action is caused

[3]See von Wright (1971), 93ff.

by a proper intention which has to be conceptualized as a distinct mental state. Essentially connected states or events, however, can nevertheless be temporally and spatially distinct and can therefore occupy the roles of causes and effects very well. That the concept of action implies that every action has a reason is only meant to say that there can be no action without any reason. From this it does not follow at all that the reasons of actions cannot be causes.

The fourth objection represents the most serious challenge to causal action explanations: How can it be that mental events cause physical or biological events? The problem of mental causation emerges because of the thesis that the physical world is causally complete, i.e. that every physical effect has a sufficient physical cause. I cannot discuss this serious problem in any detail here. However, I only want to hint at a possible solution, which consists in a combination of two different but related theses, namely, the thesis of explanatory dualism and the thesis of property dualism. If we do not understand the assertion that the physical world is causally complete or closed as an ontological thesis, i.e. that every physical event has a physical cause, but as a more modest thesis, i.e. that every physical event has a physical explanation, it is possible that a physical event has a physical and a mental explanation at the same time. And if we confess that mental events, although they are necessarily physically realized, have mental properties that cannot be reduced onto their physical realizers, it is possible that mental events cause physical events. Both theses are expressing the contention that only the mental properties can really explain why we are doing what we are doing. However, from the fact that the world of causes is also a world of reasons, as Fred Dretske puts the relevant phenomenon, it does not follow that reasons cannot be causes.

I will now turn to the question which view or theory of causality should convince us. And I want to stress one aspect of this question which, as far as I can see, is often underestimated in the debate: What is our intuition concerning causality? What is the commonly supposed sense of the concept of causality? One may think that this approach to the problem is not very original or witty. However, I have the impression that the scientific discussions about the concept of causality are too much influenced by the special problems, efforts or requirements within the different sciences. The reason for my approach is not really that philosophy often starts with intuitions. The reason is that one can easily realize that our ordinary thinking as well as our ordinary language is overwhelmingly characterized by causal considerations and explicit or implicit causal expressions. The philosophy of science should take this fact earnestly. This does not mean that we should reanimate an old-fashioned ordinary language philosophy, but instead means that we first of all have to understand the general and common sense of our concept of causality. The discussions on special scientific problems should draw on such a common understanding instead of ignoring it. We have to understand each other, not only as ordinary people but also as philosophers and scientists. That means that we need a common concept that is broad enough in order to meet the different requirements in different sciences and that is specific enough in order to represent a scientific concept at all.

If the question is put in this way, there are two main competitors for an answer, namely, the regularity thesis and the counterfactual theory of causality. I think that other theoretical approaches, for example, probabilistic causality, the manipulation theory or the dispositional account, are all different forms of either the regularity or the counterfactual sense of causality. David Hume unintentionally pointed out these two possible senses in his famous definition of a cause: '. . . we may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed' (Hume 1902, p. 79). Nowadays, there is agreement on the point that Hume's 'other words' actually did not introduce any synonymous formulation to the first-mentioned regularity thesis but instead defined a very different concept, that is, the counterfactual concept of causality. I want to propose two arguments in favour of the counterfactual conception.

The first argument revolves around the question of whether the regularity thesis can provide any coherent sense of causality at all. This question seems to be surprising in view of the triumphal march of the regularity thesis, especially within the natural sciences. However, if we remind ourselves that in his deductive-nomological model of explanation Carl Hempel converted the causal explanation to be a case of a nomological explanation understood more broadly, then the relevance of this question is more obvious. The most urgent problem for the regularity theory of causality, which simply reduces the sense of causality to the sense of regularity, is to find a convincing criterion for drawing a distinction between causal regularities and other regularities, for example, mere temporal regularities. Hume himself was, of course, very conscious of this challenge for his approach. His proposal was to suppose that only causal regularities are necessary regularities. But is this proposal convincing? Can the modal category of necessity make a real difference? This would only be the case if necessity always consisted of nomological necessity. Hume's answer would only be satisfactory if it were correct to say that necessity necessarily implies regularity. But this is obviously false. It would conceptually exclude the possibility of singular relations, which are nevertheless necessary, and this corollary is untenable. I can see no other possible criterion to distinguish between causal and other regularities unless we turn to Hume's 'other words', i.e. to the counterfactual view of causality.

The second argument therefore stresses the point that the counterfactual view can represent our intuitions concerning causality well. David Lewis emphasized this in his argumentation in favour of the counterfactual analysis: 'We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it' (Lewis 1986, pp. 160–161). In fact, it is essential to our understanding of causality that causes are responsible for real differences and changes in the course of events and, moreover, that they are responsible for the fact that there is a course of different and distinct happenings at all. The concept of causality is essentially connected to the concept of change; change is its crucial point. And the concept of regularity misses this point entirely. That something happens regularly is no explanation for the fact that it happens at all, that is, that something occurs and makes a difference. Conversely, regularity

presupposes change and therefore cannot explain it. And if we want to know whether a certain event A is a cause of another event B, we are actually asking whether B would also have occurred if A had not existed. So, as Lewis says, 'We do know that causation has something or other to do with counterfactuals' (Lewis 1986, p. 160). However, if it is correct that causation has something to do with counterfactuals, why should not we take the bull by the horns and simply take the route to reduce causal relations between events to counterfactual relations between statements? To say that A is a cause of B simply means that the corresponding counterfactual 'If A had not occurred, then B would not have occurred' is true.

The reason for some theorists' reluctance towards this solution is well known: It is difficult to formulate a satisfactory and convincing semantics for counterfactuals and for subjunctive conditionals because this attempt implies a paradoxical task. We have to find a criterion for the truth conditions of counterfactuals although their antecedent assertion is or could be false. However, the truth conditions we are longing for should be, of course, truth conditions in our actual world. That means that we have to define actual truth conditions for non-actual situations. The solution for this, at first glance, impossible task is to take possibilities seriously. 'If A had not occurred, then B would not have occurred' is true if and only if a possible world where A has not occurred and B also has not been the case is more similar to our actual world than another possible world where A has not occurred but B nevertheless has. The assertion 'If Barack Obama had not been elected as president, there would be no Tea Party movement in the U.S. today' is true in our actual world if and only if a possible world where Barack Obama has not been elected and no Tea Party movement exists is more similar to our actual world than another possible world where such a movement does exist although Obama has not been elected.

Let us grant for the sake of argument that Lewis' semantics or some other version of a possible world semantics is convincing. We should grant this, in my opinion, not because of my uneasiness concerning possibilities and possible worlds perhaps not being as great as yours, but because we understand counterfactuals in our ordinary communication very well. We should have one semantics or other that provides a theory for this actual linguistic ability. Nobody would respond to the assertion 'If Barack Obama had not been elected, the Tea Party movement would not exist' with the words: 'What? I don't understand what you're saying!' On the contrary, everybody would understand what this assertion means, namely, that the election of Barack Obama as president was a cause for the formation of the Tea Party movement.

The counterfactual analysis for being a cause can be summarized as follows:

A is a cause of B iff:

1. A occurred and B occurred.
2. If A had not occurred, but everything else being equal, then B would not have occurred.

9.5 Danto's Scientists Revisited

The proposed analysis implies that the existence of regularities between types of A and types of B is, of course, not excluded, but not presupposed, either. Whether regularities can be observed or not depends on the kinds of events. We have to take into consideration the difference between causes and causal reasoning. Causal and therefore counterfactual reasoning imply generalizations of some sort or other. However, counterfactual causation does not presuppose that the causal relation is a relation that holds only between types of events. The proposed analysis also implies that singular causes are necessary but not necessarily sufficient causes. If a historian contends that the shots in Sarajevo were a cause of the First World War, then she is asserting that this event was counterfactually necessary for the First World War. This means that the shots were certainly not the only cause of the war, but if the Serbian assassin had not murdered the Austrian heir to the throne, then this war would not have occurred.

Nevertheless, this would be a rather strong historical assertion. Additionally, this rather simple analysis does not help us at all with regard to Danto's puzzling example and the problem of the specific historical relation between past events. To repeat, this problem consists in the fact that the historical meaning of an event is a relational property, which is essentially influenced by events happening afterwards. At the time of Obama's election as president, no one could foresee that his election and, of course, his subsequent policy would provoke something like the Tea Party movement. One day, maybe, historiography will come to the conclusion that the election of the first black president had the consequence of dividing the American people rather than bringing them together. History is related to its respective future; moreover, one can say that history depends on its respective future. The German historian Reinhart Koselleck expressed this connection by calling history a 'Past Future'.

Danto's example, however, is more puzzling than the consequences of Obama's election. On the day the president was elected, it was at least possible to speculate about the question of whether this event could really reconcile the American people or would, on the contrary, deepen the rift between the political camps. This means that it is very natural to suppose that there must exist a causal connection between Obama's election and the subsequent events, although the historical meaning of his election is not determined on the day he was elected. But if we accept the supposition in Danto's fictional example that the two scientists do not know anything about each other and have formulated the very same theory independent of each other, then the case seems to be that there cannot be a causal relation. However, how can we understand and explain that the occurrence of the second, temporally later event is responsible for the fact that the earlier event has the property of being an anticipation?

I have already argued that rejecting the realistic thesis that the historical meaning is a real property of events is not a possible way out nor would it be a possible solution to suppose that causes can temporally follow their effects. Instead, I suggest revising the counterfactual analysis of causality in two respects. First, the time of the

occurrence of the respective events should be/is to be mentioned in the formulation of the conditions. This revision shall exclude backward causation and make the entire proposal more subtle and more adequate in regard to historical explanations because in history the time of an event's occurrence can be a very important fact. Second, an event can be referred to as a cause if its efficacious force only concerns particular properties of the effected event and not the occurrence of the other event itself. The consequence of this second revision is that the temporally first event in Danto's example is a cause of the temporally later event and not vice versa. The conditions are formulated as follows:

A is a cause of B iff:

1. A occurred and B occurred.
2. A occurred at time t-1 and B occurred at time t-2, i.e. A and B are temporally related to each other and A occurred earlier than B.
3. If A had not occurred at time t-1, but everything else being equal, then the following holds: either (a) B would not have occurred at time t-2 or (b) there is at least one essential property of B, which B would not have possessed, that is, C would have occurred.
4. If (b) in condition (3) is the case, then it also holds that A and C would be temporally related to each other in the same way as A and B.

According to this analysis, the earlier event in Danto's example can be seen as a cause of the later event because condition (b) in (3) is met. The later event would not be a rediscovery of a theory if the earlier event had not happened. The earlier event is causally responsible for the later event having a particular essential property. In this sense, and only in this sense, the earlier event actually changes its causal properties at time t-2. This means that the later event is causally dependent on the earlier event because the following counterfactual conditional is true: If E-1 at time t-1 had not occurred, then E-2 at time t-2 would not have had the property of being a rediscovery. In this sense, and only in this sense, E-1 is a cause of E-2.

This analysis, which manifests a version of the well-known counterfactual account of causality, means to solve a special problem emerging in the science of history, namely, the problem of the specific historical relation. To take this problem seriously is tantamount to taking Danto's original insight into the structure of history seriously. History as a science is no mere chronic, and the real histories are no mere temporal successions of events. The temporally related events are also causally structured. However, as historical events, they have a peculiar property, namely, a historical meaning that can change in the course of time, simply by virtue of what happens afterwards. I have tried to reconcile this original insight into the structure of history with a realistic picture of history. In my opinion, this means that we have to show how it can be conceptually possible that the historical meaning of a past event is nevertheless determined by the causal role which this event occupies.

References

Danto, A. C. (1965). *Analytical philosophy of history*. Cambridge: Cambridge University Press.
Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.
Gerber, D. (2012). *Analytische Metaphysik der Geschichte*. Berlin: Suhrkamp.
Hume, D. (1902). *An enquiry concerning human understanding*. Chicago: Open Court Publishing.
Lewis, D. (1986). *Philosophical papers* (Vol. 2). Oxford: Oxford University Press.
Melden, A. I. (1961). *Free action*. London: Routledge.
Von Wright, G. H. (1971). *Explanation and understanding*. Ithaca: Cornell University Press.

Chapter 10
Do Historians Study the Mechanisms of History? A Sketch

Daniel Plenge

Abstract In this exploratory sketch, I move across the boundaries of philosophy of historiography to social science and its philosophy. If we want to answer the central question of this chapter, we need to know what types of scientific problems historians are interested in, what history is, and what mechanisms are. I sketch the most prominent theories of social mechanisms in the context of wider ontological approaches. I investigate Mario Bunge's "Emergentist Systemism," "Critical Realism" in the tradition of Roy Bhaskar's influential philosophy, and Daniel Little's "Methodological Localism." Since it turns out that mechanisms are taken to be rather different entities, the question is only answered trivially, but some problems are suggested that need to be separated if the debate shall not end up in "mechanism talk." It is also suggested that philosophers of historiography can find in these debates what they are normally not interested in, that is, science-oriented philosophy of history.

Keywords History • Social mechanism • Social system • Social structure • Social causation

10.1 The Question: Historians, Mechanisms, and Histories

If we want to approach an answer to the central question raised in the title of this chapter we need to deliver a lot that we are currently incapable of providing. Our question presupposes answers to some of the little but pertinent and notoriously unsolved problems of philosophy of so-called history.

D. Plenge (✉)
Philosophisches Seminar, Westfälische Wilhelms-Universität Münster,
Domplatz 23, 48143 Münster, Germany
e-mail: daniel.plenge@uni-muenster.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_10,

First, we need to know what so-called historians do, that is, what types of problems they claim to solve, try to solve, or even solve in their research before they produce reports on former research, the products of scientific "history" (Topolski 1976). Second, it would be of interest to answer the strange or in these days even seemingly ridiculous question "What is history?" before it perhaps finally makes sense to have a look at whether at least some of the people that are called historians study the mechanisms of history. At this point, some account of what (social) mechanisms are would be of interest.

Unfortunately, the bigger part of philosophy of historiography can help us neither in solving the puzzle of what historians do nor of what it is that they perhaps study beyond the so-called historical sources, if we still dare to commit ourselves to realist assumptions at all. Fortunately, there is much and growing literature on social mechanisms in historico-social science discourses and the realist part of its philosophy that might help us in advancing towards an answer.[1]

What I will try to do in the following is to present some of the independent theories regarding the ontology of (social) mechanisms proposed by philosophers of social science over the past 40 years and to sketch some implications these philosophies have with regard to answering the unpopular question "What is history?" or the rediscovery of that question. Since all these philosophies claim that hinting at some mechanism is central to (scientific) explanation of societal change or stasis, we get to the kernel of these issues as well, since one objection to realist models of mechanismic explanation in the sciences of history/historiography might be that there are no such "things."

We start, in a rough chronological order, with "Emergentist Systemism" (ES), established by Mario Bunge,[2] and the role of mechanisms therein (Sect. 10.2) before

[1]Since the above is as much to be taken literally as it is polemic and, as I have been told, easily misunderstood, I should make explicit the following claims about the academic game around philosophy of historiography. First, philosophy of historiography cannot help us in answering the first part of the question because it is not interested in what historians do, that is, it does not care at all about research conducted by historians. Most philosophy of historiography is about "narratives." Although there are more concepts of narrative around than there are narrativist philosophers, from the view taken here, these approaches are mostly irrelevant to a philosophy of *Geschichtswissenschaft*. (If you have no qualms about doing so, call it "historiography.") Scientific historians simply do not get degrees for writing pleasant narratives but for solving scientific problems, although they might gain a Nobel Prize and the attention of philosophers of historiography by painting such "narratives." Second, philosophers of historiography cannot tell us anything about history because the ontology of history was famously buried as speculative already in the 1950s and the concept was completely moved to methodology or exchanged with "the past." See as a paradigm Marrou (1975 [1954], p. 29): "L'histoire est la connaissance du passé humain (. . .)." As anybody knows, this is no accident but the result of speculative metaphysics of the one history and its course. Put in a memorable yet unclear slogan, we can thus say that official philosophy of history is not about history. A presupposition of this paper is, to the contrary of the tradition in philosophy of *historiography*, that the concept of history belongs to ontology anyhow. If this presupposition is wrong, the question of this chapter does not make any sense. As we will see, it is doubtful that it does.

[2]I use the label coined by Wan (2011a) to refer to Bunge's system.

we have a look at some aspects of "Critical Realism" (CR), initiated by Roy Bhaskar (Sect. 10.3). Then we approach the ontology called "Methodological Localism" (ML) by Daniel Little (Sect. 10.4). This selection is justified by the observation that these philosophers are prominently discussed in social scientific journals and their contributions are thus believed to be relevant to social scientific practice. Furthermore, they are all realists about mechanisms.

Since the common use of the term "mechanism" in philosophical and sociological literature in recent times suggests a degree of convergence in the discourse that might turn out to be misleading, I will situate the respective theories of mechanisms in the context of wider ontological systems, that is, some exegesis is unavoidable and necessary. At this point, however, I will deliberately ignore theories that derive more directly from within the social sciences.[3] The resulting and admittedly wordy exegesis is necessary because due to conceptual ambiguities in the wide discourse around the notion of a "social mechanism," it is quite unclear (i) what such a social mechanism might be, (ii) which problems are to be solved in these debates, (iii) what their solution is, and whether (iv) this solution is necessary. I am going to sketch issues around (i) and (ii). The biggest problem lurking in the background is, of course, that sociologists and historians are in no agreement about the objects of research for social science, about what is to be explained if anything, and how such explanations can and should be approached (see, e.g., Blaikie 1993). Because this seems to be so, my strategy in this sketch is basically to take nothing for granted, not even a single concept such as "history" that appears to be innocent. My faint hope is that thereby possible problems become clearer that are perhaps even worth solving and that scholars who share the experience of fearing to drown in these debates might find some rescue in the following lines.

The primary result will be meager. Given the following reconstruction, there is no obvious reason to believe that historians are not interested in mechanisms. The secondary result is exactly that it is by now unclear what a social mechanism is, and the differences in theories of social mechanisms are made explicit. The tertiary result is that if there is a problem about social mechanisms, this problem comes in a bundle with others fairly familiar from social theory. The question "What is a history?" might express one of those problems. This result is achieved by attending in detail to the differences in theories of social mechanisms.

Anyway, a word of caution seems to be in order at this point. I do not claim to be an expert on any of these philosophical systems. The apology I offer for discussing them nevertheless is that this literature has hardly ever been discussed together in a comparative fashion, although people rather frequently quote from each of these positions and others, as if it were unquestionable that those positions

[3]For further recent literature and different traditions of thinking about social mechanisms that I will not discuss directly on this occasion although they have in part intersections with the theories that I sketch and are equally relevant, see, for example, Lawson (1997), Hedström and Swedberg (1998), Tilly et al. (2001), Barberi (2004), Bennett and George (2005), Cherkaoui (2005), Manicas (2006), Pickel (2006), Schmid (2006), Wight (2006), Elster (2007), Glynos and Howarth (2007), Kurki (2008), Moessinger (2008), Elder-Vass (2010), Demeulenaere (2011), Wan (2011a).

treat the same stuff behind the veil of the term "social mechanism" and, furthermore, as if it were unproblematic if they were not. My hunch is that this is questionable and problematic. This chapter might therefore be of interest to social theorists and researchers, including historians, who do not believe that this all just amounts to "mechanism talk."[4]

10.2 Mechanisms and Emergentist Systemism (Mario Bunge)

What does history consist of? According to Bunge, the world is a world of concrete, that is, material things and consists of nothing else. Thus, the so-called social world is equally supposed to be a world of things. At its heart, the project of scientific historians, once called historiologists (1985, p. 193), and sociologists consists in the study of such "things" or "social matter" (1974, p. 445, 1981, p. 5).[5]

As it is well known or at least claimed quite often, many contemporary historians and perhaps all sociologists seem to be hardly interested in individual persons and their actions but rather in so-called social facts. But what is such a social fact? According to Bunge, a fact is the being in a state or a change in the state of any material thing, so that all facts are moreover singular and positive (1996, p. 17, 2006, p. 17). A nonsocial, for example, a mental fact, then, is the being in a state of a brain or a change in the state of a brain of a higher-order animal, given Bunge's psycho-neuronal emergentist monism (1984). But what are social facts or the analogue to brains?

The basic pillars of the "systemic approach" (2006, p. 128) are the concepts of a system and of emergence (2001a, 2003a). According to Bunge, the world is a world of systems (1979a). Everything that is not a system is at least a part of one, or if it is not yet, or merely not for the moment, such a part, it will become a component of a larger whole.[6] Whereas some physicists might study elementary particles that are things but not systems, the rest of the scientific community, including historiologists, studies complex concrete systems (1979b, 1992b, 1993b).

---

[4]This chapter might not be without interest to historians because these philosophies have hardly received attention in philosophy of historiography, which is my point of departure here, and they have not been discussed among historians themselves, although they deal with questions permanently discussed in their circles. Exceptions are to be found in (McLennan 1981; Gibbon 1989; Lloyd 1986) in the case of Roy Bhaskar's work. Bunge's work has been ignored so far, perhaps because of his claim that historiology is the most rigorous of all the social sciences and due to his robust ontological and epistemological realism; see, for example, Bunge (1985, 1988, 1998a).

[5]On Bunge's materialism, see Bunge (1977, 1981, 2002), Mahner (2001), Bunge and Mahner (2004). If not indicated otherwise, references in the text concern the work of the author mainly discussed in the respective section.

[6]The basic slogan of this ontology is therefore (Bunge 2004a, p. 191): "*everything in the universe is, was, or will be a system or a component of one*."

> A concrete system is a bundle of real things held together by some bonds or forces, behaving as a unit in some respects and (except for the universe as a whole) embedded in some environment. (Bunge 1997, p. 415)

Human social systems are obviously composed of persons and the artifacts they have built (1996, p. 21). And they are constructed, maintained, altered, and primarily destructed by persons. Thus, to qualify as such, human social systems have to be composed of at least two persons engaging in some social relation or interaction, but they can be as big as is compatible with the laws of social systems, that is, laws that are assumed to exist by Bunge in contrast to many contemporary sociologists (2007). Such social systems can be short living as well as long lasting, either rather local or spread all over the globe. To give some examples, "families, afternoon coffee parties, football teams, school classes, working groups in organizations (Institutsarbeitsgruppen)" (Bunge and Mahner 2004, p. 86; any translation D.P.) or even the world economic system, if it qualifies as such, are social systems on Bunge's account.

The traditional and controversial question concerning the existence of social wholes is therefore affirmed in ES. It seems to be equally important that there are two types of wholes according to Bunge. A system is not a mere aggregate or heap because it has emergent properties: most importantly the system structure that determines that the parts of the system are at least minimally integrated or cohesive, so that the system "behaves" as a whole in some respect.[7] Due to this real or ontic integration or connection, for example, "a labor union is a social system and therefore just as concrete and real as its members" (1998a, p. 69). Mere aggregates or heaps as, for instance, a bunch of people accidentally attending a football (soccer) match "together" or a selection of people sharing some more or less relevant property (e.g., bourgeoisie) do not constitute systems, although they might qualify as wholes.[8]

Disentangling the short definition quoted above and thereby summarizing the former, every (social) system at any point in time is supposed to be minimally characterized by a definite composition, a definite structure, and a definite environment. This yields in scientific reconstruction what Bunge formerly called the minimal CESM-Model of a system (e.g., 1995, 14f.). Whereas the *composition* of a system is the set of its components that are as concrete and as real as the system as a whole, the *structure* (sometimes also called "organization" or "architecture"; 2004a, p. 188) of the system is the set of all the relations holding among the system components. The *environment* is said to be the set of relations the system components hold with things (systems) that are not part of the system. For example, in an army relations of command are part of the endostructure of the system, and relations of combat or supply belong to its exostructure (2003c, p. 277).

---

[7]Cf. (Bunge 1996, p. 21): "The structure of a system is its key emergent property." See also (Bunge 2010, p. 379).

[8]Mere heaps are sometimes called "statistical wholes," for example crowds, classes and institutions. Systems like gangs or firms are called "ontic wholes." For this distinction and the examples, see (Bunge 2004b, p. 372). For Bunge's concepts of group and class, see (Bunge 1995, Chap. 3).

Skipping some interesting problems concerning systems ontology relevant to historico-social scientific practice,[9] subsets of the overall structure of the system are what Bunge once called the "system structure (or bondage)" and the "spatial structure (or configuration)" of any system (1979a, p. 11). Whereas spatial and temporal relations are often of minor interest to historians and social scientists because they presumably make not much of a difference to the whole or its parts, social relations make, by definition, a difference to the relata because they are bonds, ties, couplings, or links.[10]

At this point we only have to mention that these relations are in part dynamic and causal. They are interactions if they are "root (or basic) social relations" (1979a, p. 222). But they might also consist merely in being a part of a whole, as in being a citizen or being married to someone, the latter being perhaps the prime example of a social relation that does not necessarily imply interaction. Other psychosocial bonds as affection and interest (1979b, p. 20), loyalty (2003a, p. 20), love (2004a, p. 189), or the rules inherent in or constitutive of social relationships (1979a, p. 196) hardly count as causal as such in Bunge's system, in contrast to CR and ML.

Furthermore, different social systems might be related among themselves as wholes to constitute higher-level systems. If one system acts on another system, this is said to be one case of "social power." Whole social systems might also be dynamically related to persons; this is the second type of social power in ES (2009b, p. 189). But barring miracles, every relation between social systems, that is every non-basic social relation, is executed by a person-to-person or basic social relation, that is, if it happens to be causal. A further reason to call interactions the basic social relations in ES is that social systems and the more permanent bonds that hold them together diachronically emerge from personal interaction (2001b, p. 134).[11]

Thus, structures of systems in general and social structures specifically are not quite easy to grasp in social science or philosophy. According to ES, the latter are mediated ultimately by the brains of actors because social relations depend on them. In Bunge's view, this is the truth of ontological individualism (1997, p. 453, 2008, p. 61). However, social relations are claimed to be emergent from interaction of people that are embedded in a social and natural environment, not on free floating ideas. Moreover according to ES, rather obvious but often ignored in traditional philosophy of historiography, social "systems have no brains, hence they have no intentions" (2004b, p. 376), although all their social properties and most of their changes are due to people acting and interacting.

[9] On the problem concerning the boundaries of social systems, see (Bunge 1992a).

[10] For the concept of a bond or link and similar notions, see Bunge (1977, p. 261, 1979a, p. 225, 1992a, b, 1993b); Bunge and Mahner (2004, p. 73).

[11] We should note here that there is a family of concepts around the notion of emergence in ES that should be distinguished at this point. Emergent properties are, of course, properties (see below). Emergence is a process in which an emergent property comes into existence. An emergent is a thing possessing some emergent property.

Another important point for our purpose in addition to the reality of (social) wholes or "totalities" (1999, p. 8) as such is that they are claimed to possess properties of their own, that is, as wholes, some of which we did already encounter above, that is, societal or systemic properties, which are the reason to grant these totalities an ontological status of their own in the first place. In Bungean ontology, there are again two types of holistic properties, namely, emergent and resultant properties. What is an emergent property according to ES?

> P is an *emergent* property of a thing *b* if and only if *b* is a complex thing (system) no component of which possesses P, or *b* is an individual that possesses P by virtue of being a component of a system (i.e., *b* would not possess P if it were independent or isolated). (Bunge 1996, p. 20, cf. 2003a, p. 14; Mahner 2001, p. 79)

Thus, there are two types of emergent properties. The former are called global or intrinsic emergent properties, the latter relational or structural emergent properties (Bunge and Mahner 2004, p. 79). If the property P is a property of a whole, but not emergent, it is called resultant. It is already possessed by the system components and merely aggregates to an unstructured whole.

Let us furnish a bunch of arbitrarily selected examples for these types of social properties. Global emergent properties are supposed to be social stratification, cohesion, mobility, stability, economic growth, form of government, political stability, mode of production, undergoing a social revolution, size of territory, or population. Examples of structural emergent properties are role, civil right, scarcity, price,[12] and every form of being coupled to other system components at all, so that every time a social system is built, at least two people are supposed to acquire structurally emergent properties (2003a, p. 78). In contrast to resultant properties (e.g., the total consumption of bubble gum in a society), emergent properties are ontologically irreducible to (though explainable by) properties of the system components, as is the total production of the economy of a society or being the goalkeeper in a football (soccer) team (1979b, p. 22; Mahner 2001, pp. 170–294).

Framed in a few fancy slogans reminiscent of the philosophical tradition, we perhaps get near the core of systemic ontology: There are not just capitalists but capitalist economies. There are not just protestants but people organized in churches. There are not only soldiers, but soldiers organized in nested military organizations with military subunits allegedly characterized by an emergent firing power manifesting in the field.

This is the static view of the world in Bungean ontology. Since we started out this paragraph with an utterly vague question about the constituents of history, we should not finish it without saying something about changes occurring in the world, since once upon a time the "historically minded" thought "history" in the potentially ontic and epistemic meaning to have something to do with change (Topolski 1976).

[12]These examples are all taken from Bunge's work; see Bunge (1977, p. 97; 1979b, p. 20; 1996, 19f.; 1999, p. 8; 2003a, p. 13), Mahner (2001, p. 298).

The static CES-Model of a system presupposes, of course, a version of ontological realism, the thesis that the (social) world does not care about it being studied – a thesis rather uncommon in contemporary sociohistorical academia (1993a). Given this theory, every concrete (social) system at every point in time possesses a definite number of properties. This is the *state* of the system at a given moment.

The dynamics of the system consist then, according to Bunge, in the changes of state of the system in the epoch of its existence. In ES a change of state of a thing is an *event.* A *process* is a sequence of changes of state or events. And last but not least, the *history* of the thing or system is supposed to be nothing but the total sequence of its states during its journey through the "course of history," the totality of its changes in properties, including the acquisition (or processual emergence) of relatively or absolutely new properties (2003a, p. 13):

> Whereas a process (or partial history) of a thing is any (ordered) sequence of states of (or, alternatively, events in) a thing, the (total) *history* of a thing is the ordered set of *all* its successive states (or events). (Bunge and Mahner 1997, p. 19, cf. 2004, p. 58; Bunge 2003c, p. 128)

At least some interesting implications of ES for any philosophy of history are noteworthy, although they are inconspicuous. Although there are quite many, we will list only three of them. First, every history is the history of a thing and every event occurs in some thing. In contrast to what can frequently be found in philosophy of historiography, social theory, and historiology, there are no processes in themselves, properties of events, histories of events, or evolving histories, and society is not a process.[13] Furthermore, history is not "a tale told by an idiot" (Sewell 2005, p. 102), but strictly speaking nonexistent for being a summative concept representing all histories of concrete things (1996, p. 26), which are on their part as real as anything could possibly be. For not being things, histories of things do not have properties and therefore do not change or do anything at all (Bunge and Mahner 2004, p. 68). Second, according to this explicit ontology, most of the talk about history implicit in philosophy of historiography is either nonsensical or false. The latter holds for "history was in the past" (Tucker 2012, p. 277) or "All history is the history of thought" (Collingwood 1994, p. 215), the former for common slogans as "the end of history," "the return of history," or "the course of history."[14] Third, since societies are considered to be real in ES and to have genuine properties that might change or even be newly acquired, there is a real history of every society or of any social system of whatever scale that can in principle become an object of study. This might seem trivial but it is not given that this is impossible in ontological

[13] On this account, expressions such as "historical events," "historical processes," "historical facts," or "historical societies" are pleonasms and talk of "historicity" is trivial; cf., Bunge and Mahner (1997, p. 20).

[14] Of course, in ES everything whatsoever has a history, whether it happens to be a boot maker or a coffee maker; cf., Bunge (1977, p. 255).

individualism or "historical individualism" (2000b) because on that account there is no fall of any empire since there are no empires in the first place.[15]

Accordingly, we can finally say what social facts are in Bunge's philosophy. They are states of social systems or their changes. Social change is simply defined as change in societal properties (1979a, p. 235). And although social states and their changes occur outside any individual's mind, they depend on social interaction and so finally on the activation of the furniture of the minding brains, to put it in Bungean terms, of concrete persons (2008, 60f.) or the "individual-in-society" (1979b, p. 17). This individual is on its part characterized by structural emergent properties and reacts furthermore to social facts as she perceives them, although these are not reducible to her interpretations of reality (1997, p. 453).

In order to provide us with further material to contrast ES with the other ontologies involving mechanisms, two further implications should be noted. Since societies do, according to Bunge, not hover above individuals that compose them and, moreover, social relations are constituted by interacting persons, whereas the set of all relations among the system components represents its structure, all of these are no candidates for being causes. Especially, "[t]here is no action of the whole on its parts (...)" (1979a, p. 39). More concisely, because in ES events (changes) are the relata in causal relations, neither properties nor states or conditions are considered as causes. The expression "a causes b" is therefore a short version of "a change in the state of thing *a* produces a change in the state of thing *b*" (Bunge and Mahner 2004, p. 95). Since Bunge conceptualizes the productive or generative aspect of causation in terms of different forms of energy transfer (1996, p. 31, 1999, p. 27), it seems to follow that only concrete persons as they are working with, talking to, or shooting each other cause something in the so-called social world, alongside changes in technological and natural systems, of course. The former quote thus continues: "rather, there are actions of some components on others" (1979a, p. 39). In a nutshell, social properties are real, given ES, but do not cause social or psychic events.[16]

[15] See Veyne (1996, 153f.), who happens to be a historian: "La France ne fait pas la guerre, car elle n'existe pas réellement; seul existent des Français (...). Pour un historien comme pour tout homme, ce qui est proprement réel, ce sont les individus." Sztompka (1991, 188), who is a sociologist, does not nod: "[T]he army is more than soldiers, a corporation more than all those employed, and Poland more than all Poles."

[16] For reasons of space and complexity we cannot here discuss the whole story and this should remind us of the circumstance that my reading of Bunge is far from infallible. But first of all, Bunge has suggested for a long time that not every determinant of change is a causal determinant (Bunge 1982, 2009 [1959]). The structure of a social system (macro property) and even its subset of spatial relations, for instance, in a production line as a sub-system of a factory, might determine the output of the system (macro property). But in ES this is far from stating that these relations cause actions or changes in the properties of social systems, though they determine the possible state of the system before and while causation is going on through people's hands. In some examples Bunge (1996, p. 280; 2000a) talks of macro-causation in terms which might turn out to be problematic, for example, when it is suggested that actions are causally stimulated or constrained by the place the individual holds in a system. Of course, this is not problematic if one remembers that such

After all these seemingly just preliminary remarks, what is a social mechanism according to ES? Is it a more or less stable complex thing (e.g., a family, the German economy, a state, a university, a mafia family) or a process (e.g., child rearing, production, growth, innovation, social control, academic tenure, dealing drugs) or a property (e.g., cohesion, productivity, structure)?[17] Though we have to shorten the story considerably, in this ontology the concept of a mechanism as a structured object is unnecessary due to the concept of a concrete complex system, its components, and its structure.[18] Consequently, "mechanisms are not things but processes in systems" (2009, p. 19). Moreover, "[e]very mechanism is a process, but the converse is false" (1999, p. 24), since not every systemic change is a mechanismic process, for example, economic growth, the spread of brilliant ideas among social systems or dancing. Therefore, mechanisms are not identical to the history of a system, the boundary concept of change (of a thing) in Bunge's system. Mechanisms are ultimately supposed to be a subset of the totality of processes occurring in a concrete complex system, namely those that involve the properties that are essential to the kind of system we are concerned with, their specific processes or functions that keep the system alive or "running", "make it work" the way it "normally" does or "what it is" (e.g., 2003b, p. 146).[19] Accordingly, further basic assumptions of this ontology are that every (social) system is endowed with at least one specific function, mechanismic or essential process (1995, p. 43; 2010, p. 376), that there are mechanisms of change and mechanisms that prevent or control change or keep the system in a state. That is to say, there are mechanisms for "either the emergence of a property or another process" (2003a, 20f.; 2010, p. 379). A social mechanism is, of course, just a mechanism in a social system, that is "a process involving at least two agents engaged in forming, maintaining, transforming or dismantling a social system" (1999, p. 57).[20]

Where are mechanisms when it comes to analyze the so-called social world? Are they to be found on the side of agency or structure? In ES they are neither to be found in the heads of people, that is why they cannot be *reenacted* or *verstanden*,

---

descriptions are most often short for complex interactions and their patterns, though the place or role an individual holds in a system is in ES emergent and systemic.

[17] The examples are taken from Bunge's work.

[18] A quick look at the development of Bunge's thinking on mechanisms and mechanismic explanation suggests that he started off with a theory that tended tacitly to conflate these categories, whereas his long-lived project of systems ontology lead to their strict separation yet systematization. See Bunge (2009a [1959], 1965, 1967, 1968, 1983, 1998b). If one is to believe Wan (2011a), it seems that the current literature on mechanisms moves in the opposite direction.

[19] On Bunge's concept of function, see Bunge and Mahner (2001).

[20] Or more formally (Bunge 2006, p. 131; cf. 2004a; 2010): "Definition 1: If σ denotes a system of kind Σ, then (1) the *totality of processes* (or *functions*) in σ over the period T is $\pi(\sigma) =$ The ordered sequence of states of σ over T; (2) the *essential mechanism* (or *specific function*) in σ over the period T, that is, $M(\sigma) = \pi_s(s) \subseteq \pi(s)$, is the totality of processes that occur exclusively in σ and its conspecifics during T. Definition 2: A *social mechanism* is a mechanism of a social system or part of it."

nor in a mysterious whole above people or social structure, but "in or among" social systems and a part of the systemic processes unfolding therein (1999, p. 57). But social interaction is not only said to be the "source of system," that is, of their diachronic emergence under the condition that one essential process gets running, but also the "fuel of mechanism" (2001b, p. 134). More interesting and explanatory accounts of social systems or their histories thus also have to include mechanisms in CESM-models (1998a; 2002).[21]

Let us finally deliver some suggestive examples that might help us in answering the main question. The specific function or part of the mechanism of the postal system is to distribute the mail (1995, p. 43). The finance authorities collect taxes or fail therein, and in schools, pupils are taught and teachers managed (Bunge and Mahner 2004, p. 75). And as any enthusiast for movies believes to know, blackmail, drug dealing, and intimidation of judges are specialties of mafia families and, according to Bunge, mechanisms or part of the total mechanism that keeps these systems running (2008, p. 53).

10.3 Mechanisms and Critical Realism (Roy Bhaskar)

What does history consist of in CR, that is, the philosophy that gained wide influence in social theoretical discourses and philosophy of social science over the past decades?[22] At first glance, there seem to be some similarities to what we found in Bungean ES. The world is said to consist of "things" or "mechanisms."[23] The overall conception of this ontology is also materialist but not reductionist, because psychological and social levels or the properties of its members are supposed to be genuinely real, that is, "emergent" (1986, p. 104, 1989, p. 91) from the preceding levels.[24] As in the case of Bunge, emergence is an ontological, not an epistemological, concept (1978, p. 113). The resulting ontology is therefore called "synchronic emergent powers materialism" (SEPM).[25] Moreover, Bhaskar at times equally distinguishes between mere aggregates (heaps) and what he calls

[21] Perfect knowledge of a system would also include its history and its laws; see Bunge (1979a, p. 8). The reader will have noticed that mechanisms have been included in the ideal model of a system in Bunge's philosophy fairly recently, although he is thinking about mechanisms since the 1950s.

[22] See, for example, Benton (1977), Outhwaite (1987a), Archer (1995), Danermark et al. (2002), Groff (2004), Manicas (2006), Frauley and Pearce (2007), Elder-Vass (2010), Sayer (2010a), Wan (2011a). In order to keep track of history, Bhaskar's work is cited by the date of the original publication.

[23] See Bhaskar (1978, p. 51): "The world consists of things, not events." See also Bhaskar (1978, 47): "The world consists of mechanisms not events."

[24] See Bhaskar (1994, p. 74): "The human world is an irreducible and causally efficacious dependent mode of matter."

[25] On Bhaskar's emergentism in comparison to that of Bunge, see Kaidesoja (2009).

"totalities," which are said to be "characterized by an emergent principle of structure" (1994, p. 80). Although this suggests that structures are properties of totalities, on the same occasion totalities are said to be structures.[26]

In SEPM "people and society" are accordingly supposed to be "radically different kinds of thing" (1979, p. 42), or put in slightly different terms, "while the properties and powers of individuals and societies are *necessary* for one another, they are *irreducible* to one another" (1989, p. 63). However, whether ES and CR are compatible is yet an open question, given the first tendentious quotes that seemingly equate things with, in Bungean parlance, "their" mechanisms, while the second equate totalities with structures, whereas the concept of a power is furthermore absent from ES.[27]

The things that are supposed to constitute the world according to Bhaskarian CR are "causal agents." Causal agents are those entities endowed with causal powers.

> To say that x has the power to do φ is to say that it will do φ in the appropriate circumstances in virtue of its nature (e.g. structure or constitution); that is to say it will do it in virtue of its being the kind of thing that it is. (Bhaskar 1978, p. 237)

These natures are on other occasions also called "essences" or straightforwardly "structure." Thus, to ascribe a power (to some thing) amounts to distinguishing accidental from essential properties of the thing. Only if the "intrinsic structure or essential nature of a thing" changes that the powers and tendencies of the thing change. This would not be the case if only some conditions for its manifestation or relations to other things changed (1978, p. 97).

A causal agent is then nothing more, but also nothing less, than "anything which is capable of bringing about a change in something (including itself)" (Bhaskar 1978, p. 109). These things or agents are the bearers of at least two types of causal power, namely, (i) powers and (ii) liabilities. The former are held to be capacities to produce changes actively, whereas the latter are conceived to be capacities to suffer or passive powers (1978, p. 87). To use a common example, a fire is supposed to have the power to burn people that are liable to be burned.[28]

Causal powers, in turn, are the foundation of tendencies. There were supposed to be two types of tendencies in the 1970s (1978, p. 230); later on Bhaskar distinguished seven types (1994, p. 83). Tendencies, if exercised, are said to ground the normic behavior of things (1978, p. 106) and are among others the referents of normic law statements. As it seems, Bhaskar shares Bunge's belief that there are "causal laws, generalities, at work in social life" (1979, p. 27). But what are

[26] See also Bhaskar (1978, p. 85): "Societies, people and machines are not collectivities, wholes or aggregates of simpler or smaller constituents."

[27] Sometimes similarities between both ontologies are noted though the differences are seldom made explicit. For comparisons see Kaidesoja (2007, 2009), Wan (2011a, 2011b).

[28] Harré and Madden (1975, p. 47). I will here not address the problem of the relationship of Harré's work to that of Bhaskar. But it is worth reminding that the concept of mechanism as used in CR has its basis in Harré's work of the 1960s and early 1970s. See Harré (1961, 1970, 1972).

tendencies more exactly? We find that tendencies "are roughly powers which may be exercised unfulfilled" (1978, p. 98).

> Tendencies may be possessed unexercised, exercised unrealized, and realized unperceived (or) undetected by men; they may also be transformed. (Bhaskar 1978, p. 18)

If tendencies are possessed unexercised, they seem to be mere powers. The thing is said to possess the tendency even though it is not yet tending to do anything. Tendencies are "dynamized" powers or powers "set in motion," although exercised powers (tendencies) need not manifest themselves in open systems (1978, p. 50), given that other tendencies might counteract. If a power is triggered, it (or the thing bearing it) is claimed to tend towards its manifestation. It acts "transfactually" and would actualize if it is not counteracted by other actualized powers. In other words and terminology that is hardly used by critical realists, powers seem to be dispositions and tendencies are dispositions that are triggered or released and manifest themselves only ceteris paribus.[29]

In summary, we can say that whereas the basis or foundation of tendencies seem to be powers or that tendencies are thought to be a mode of being of powers, these on their part seem to have a basis from which they are supposed to emerge synchronically. The bases of powers are the "natures" of things (1978, p. 178) or their "real essences" that are supposed to be their "intrinsic structures" (1978, p. 174).

If our reading is not structurally beside the point, we can preliminarily picture the basic outline of this ontology in the following way:

Structures (or natures or essences) → powers → tendencies (normic behavior or laws)

that is, roughly, structures ground powers which are the foundation of tendencies.[30]

Even at this point, before we even got near the social ontology of CR, we can sketch some hypotheses of CR's philosophy of history. First, history is not just a sequence of some such events, but something in "the course of" which something real persists or even radically changes or is transformed as far as to eventually produce qualitative novelty.[31] Secondly, given that Bhaskar believes that something persisting is the basis for a "genuine concept of *change*, and hence *history*" (1979,

[29] For a discussion of problems around these central notions in CR, see Fleetwood (2009, 2011).

[30] The background of this ontology is of course an anti-positivist stance in form of the hypothesis that "the real" is not exhausted by perceptions of events or events, especially "[s]ociety is not a mass of separable events and sequences" (1979, p. 68). These assumptions are at the heart of the three ontological domains of CR (1978): "the real" (structures, powers, totalities etc.), "the actual" (events), and "the empirical" (observed events).

[31] See Bhaskar (1989, p. 10) reminiscent of Marx: "In the constant conjunction form history grinds to a halt in the eternalized present. History is what there has been or is elsewhere but is no longer here now."

p. 47), the question arises what this something is that endures, changes, or is transformed. In his later work, causal powers are said to be "processes-entified-in-products" (1993, p. 52) and Bhaskar even goes as far as to write about the "presence of the past," its causal efficacy and of the "presence of the future" (1993, 140ff.). Accordingly, one answer to the question which stuff is transformed during the "course of history" and even determines in some sense "the future" that we can extract from CR is that this stuff is causal powers. A "historical event," then, is not any event whatsoever as in ES, but an event that significantly changes or transforms historical things and their powers (1979, p. 24).

Where do mechanisms enter the ontological picture? A "generative mechanism is nothing other than a way of acting of a thing. It endures and under appropriate circumstances is exercised as long as the properties that account for it persist" (1978, 51f.), that is, as long as the natures, essences, or intrinsic structures do not change significantly. "Mechanisms are enduring; they are nothing but the powers of things. Things, unlike events (which are changes in them), persist" (1978, p. 221).[32]

On this reading we can substitute the former summary of basic CR ontology by the following schema so that mechanisms take the place of powers:

> Structures (or natures or essences) → mechanisms → tendencies (normic behavior or laws)

that is, roughly, structures are the foundations of mechanisms which ground tendencies.[33]

Whereas in Bungean systemism mechanisms are actual or manifest and furthermore processes, in Bhaskarian realism they are, or seem to be, dispositional properties or powers. Yet on another reading, they might be something in between. In *Plato* Etc. we find the statement that structures possess causal powers, "which, when triggered or released, act, as generative mechanisms" (1994, p. 23). Given this reading, mechanisms are powers triggered or dynamized, the role taken above by tendencies. Accordingly, the concepts of power and tendency are said to "come together in the concept of *generative mechanism,* which may be either or both" in standard CR literature (Hartwig 2007, p. 57).

> When stimulated, released or enabled, then, powers and generative mechanisms are tendencies (. . .). Where a thing just is its powers and tendencies (mechanisms), these are the same as structure. (. . .) Otherwise mechanisms and structures are distinct, i.e., mechanisms (powers and tendencies) are *of* (instantiated in) structured things. (Hartwig 2007, p. 57)

[32] See also another classic formulation by Bhaskar (1978, p. 50): "[T]he generative mechanisms of nature exist as the causal powers of things."

[33] In Sayer (2010a, p. 15; 2010b, p. 117) we find a slightly shorter schema: structures → mechanisms → events.

On the second reading, then, we get something like this schema:

Structures (or natures or essences) → powers → mechanisms

that is, roughly, that structures ground powers, which, when triggered, transfactually act as mechanisms or are mechanisms if triggered.[34] In the case of people, who, in the later work (1993, p. 165), are said to be an example of things that just are powers, we get:

Structure (thing) = mechanism

that is, some structures are mechanisms (or ensembles of powers). With a little slip of the pen, we might summarize the forgoing in the following schema:

Mechanism (thing, structured thing, structure) → mechanism (power) → mechanism (tendency).[35]

Given the foregoing, what are mechanisms in CR? Are they complex things, properties, or processes? As far as I can see, this happens to be rather unclear. They sometimes appear to be complex "things," sometimes events, whereas the primary referents are non-manifest properties capable of "doing" something or "bringing about" changes in things. Given this multiplicity of meanings of the mechanism concept in CR, it is not surprising that one gets the impression that much of the literature interprets these CR mechanisms implicitly as complex objects (mechanisms as systems)[36] characterized by recurrent processes (mechanisms as processes)[37] due to the properties of the component things (mechanisms as powers) and their relation, organization (mechanisms as structures or structural powers), or interaction.

[34] If we read "act as generative mechanisms" as "resulting in actual processes," then we might already here get the hypothesis that mechanisms are processes, though this seems to be against the spirit of the letter. We get that result in the next footnote.

[35] To round up the story, we have to add here that according to Bhaskar (1994, 257f.), processes or rhythms also have powers, and according to Hartwig (2007, p. 189), events might also "function as mechanisms," which seems to amount to the claim that events possess powers of their own beyond the powers that are grounded in or emergent from the structure or essence of the thing that undergoes a change in the event. For short, events and processes might also be powerful dispositional mechanisms.

[36] For example, Wight (2006, p. 31), affiliated to the tradition of CR, writes of "the causal power of mechanisms."

[37] Cf. Kurki (2008, p. 233).

But this reading seems to be slightly beside the point, because given the former interpretation and the strict distinction between the domains of "the real" and "the actual," the former including mechanisms (powers), the latter encompassing events as changes in things, hardly anything is happening in our world yet, given that even triggered powers (mechanisms or tendencies), though pinched towards moving, might, by definition, not end up in actual changes in the world, even when they are said to be acting in some sense (transfactually), since they might remain unrealized though exercised, if they happen to be exercised at all. Bungean history is, as it seems, still dormant in the story told until now.

Given that social mechanisms were in ES said to be neither social things nor persons but systemic processes, where are mechanisms to be found in CR? Since they are (primarily) powers, mechanisms are where powers are to be found. And according to CR's *social ontology*, there are social structures or societies and persons, which are both claimed to be ensembles of powers. These types of powers finally meet each other in "processes" or "rhythms."[38] In the social sphere

> *process* [is] where structure meets events; that is, in the study of the mode of becoming, bestaying and begoing of a structure or thing, i.e. of its genesis in, distantiation over and transformation across space-time. Process is not an ontological category apart from structure and event; it just is a structure (or thing), considered under the aspect of its story (sic!) – or formation, reformation and transformation – in time. (Bhaskar 1986, p. 215)

In ways similar to Little's ML framework, in the social sphere "social structures" or "social forms" (1983, p. 85) fuse in actions with agential causal powers (mechanisms) and natural causal mechanisms (powers) to lead to changes in the maintenance or transformation of "social structures" (mechanisms). Societal change is thus a change in or a transformation of societal powers. A society, a causally inert system of systems in ES, is said to be "a complex and causally efficacious whole – a totality (. . .) which is being continually transformed in practice" (1989, 87f.). Although:

> Society (. . .) is an articulated ensemble of tendencies and powers which, unlike natural ones, exist only as long as they (or at least some of them) are being exercised; are exercised in the last instance via the intentional activity of human beings; and are not necessarily space-time invariant. (Bhaskar 1989, p. 79, cf. 1978, p. 196)[39]

It is noteworthy that the powers in the social case are dispositions that are always actual or at least acting transfactually as tendencies or exercised powers, but are never purely dispositional, given the former quote.[40]

---

[38] On "rhythms" see Bhaskar (1993, 1994).

[39] On the more narrow CR conception of society, it does not consist of individuals or groups or some such circumstances but of internal relations: "A relation aRb is internal if and only if a would not be what it is *essentially* unless it were related to b in the way that it is." See Bhaskar (1993, 10); see, also Bhaskar (1994, 75; 1979, 32, 54). This theory has implications for the philosophy of social change (Bhaskar 1979, 52): "In social life only relations endure."

[40] Here we also have to admit that the story is far more complex. There has been a discussion about this point in CR that resulted in the acceptance of social powers or dispositions that do not just exist as exercised or actualized powers. Cf., Porpora 2007.

In the famous "Transformational Model of Social Activity" (TMSA) or the later "Social Cube," social life is conceptualized as work on a preexisting social world, that is, as work on "social structure." This move yields the hypothesis that "in every process of productive activity a material as well as an efficient cause is necessary" (1979, p. 43, 1986, p. 119). Accordingly, in classical CR "social structures" are supposed to sponsor the world with "social material causes" (1993, p. 155) that "govern, enable and constrain" (1986, p. 130) individual action. Accordingly, these powers are, first of all, believed to be predating these actions, whereas persons change, reproduce, and transform "social structures" that enable and constrain their actions. The truth of social individualism according to Bhaskarian CR is that "people are the only moving forces in history" (1989, p. 81).

Since nothing makes things clearer than examples, let us have a look at some arbitrarily selected instances of social structures that carry powers or simply are powers. Examples for "social entities" are "institutions, traditions, networks of relations and the like" (1989, p. 175); the former two are also on occasion called "emergent social things" (1993, p. 54). Structures are, furthermore, said to be "religious rites established by the practices of the long dead" (1994, p. 95); "the economy, the state, the family" (1989, p. 4); "Nazism, bureaucracy and (. . .) capitalism"; and "buildings we have, the stock-market, the whole financial economic system"; they are claimed to be "everything that is there before any given voluntaristic act" (2001, 28f.), as are "languages," "systems of belief, cultural and ethical norms" (1978, p. 196). Anything that constrains or enables individual actions is structural and powerful: even "stories are social structures" (2001, p. 36) as well as "the age structure of a population, or the occupational structure of a population, or the academic status of a population or perhaps the class structure of a population" (ibid. p. 37). Furthermore, "social structures and their generative mechanisms" are said to be exemplified by "ways of cooking, making micro-chips or production generally" (1993, p. 155).

The heterogeneity of these examples of social structures would not be problematic if these structures were not believed to be mechanisms or powers, which are what distinguishes causal agents, namely, "anything that is capable of bringing about a change in something" (1978, p. 109), from non-agents. If our story above is correct, in order to get powers in CR, we need "the key concept of a causal agent" (1978, p. 77). The question is whether we find those agents with essences and emergent powers in these allegedly social examples, for example, traditions, stories, norms, and social relations.

The problem seems rather obvious. First, if there are no such complex objects to be found whereof those powers are properties, CR faces the problem that nothing seems to justify the assumption of "social causal powers" any more.[41] If we do not need agents as structured objects or systems in Bungean terms to get powers, then we seem to face in social ontology a different concept of powers than in the

[41] See Bhaskar (1978, p. 51): "Most things are complex objects, in virtue of which they possess an ensemble of tendencies, liabilities and powers."

materialist ontology for the natural sciences.[42] Second, it seems to be controversial to frame "social material causality" in terms of the original causal powers account, since this seems to be in slight conflict with the productive or generative account of causation.[43] Anyway, it might be worth believing that a fistful of dynamite has the power to blow one's brains out and actually does so if it is triggered and, finally, acts undisturbed. But, as any realist knows, this seems to be something different from the case in which a fistful of dollars results in my buying a cuckoo clock if it is handed over the counter. One of the powers of a fistful of dollars might be to be liable to be burned, but that one of its powers is to buy a clock might be controversial, though everybody seems to understand what is going on in this social episode in common sense terms.[44] Third, if we stick to the belief that powers have to confront triggers or releasers or what have you, we might ask whether people are the conditions for the manifestation of a social causal power (mechanism), for example, of a norm. Or we can ask whether social causal powers (mechanisms) are merely the conditions for the manifestation of individual powers. Since a condition for the manifestation of a disposition of a thing is another powerful thing that is disposed to trigger the disposition of the former if they happen to join one another in an event, we might say that personal and social powers have somehow been made for one another. But however we twist and turn, in most examples of social causal powers, we lack the second and furthermore social "agent" and a property that might be considered as a candidate for a disposition. What is, say, the disposition of, or dispositional about, "the age structure of a population"? Of course, if it would turn out that there are no plausible candidates for social powers, there would be no social mechanisms, given CR ontology. But, of course, this is an open question.

10.4 Mechanisms and Methodological Localism (Daniel Little)

Daniel Little is another outstanding philosopher who has worked on the problem of sociohistorical explanation, mechanisms, and social ontology fairly independently of the authors discussed so far since the 1980s.[45] Methodological Localism revolves

[42] To grant "unobservables – such as ideas, rules and discourses," a causal role, which seem to be "non-agent-like factors," Kurki (2008, pp. 170–174) frames the concept of an "ontological object" that is not supposed to be a "'thing'" (ibid. p. 169). Contrary to this, Kaidesoja (2007) argues that something like a Bungean complex thing is necessary to ascribe something a power and wants to correct CR in this direction.

[43] See the criticism by Harré (2002), Harré and Varela (1996). Famously Lloyd (1993, p. 46) already distinguished two types of powers: "Persons have agential power, structures have conditioning power."

[44] For the claim that money has an essence, see Bhaskar (1978, p. 88).

[45] He seems to have been influenced by the work of Rom Harré; see Little (1989; 1991). For the precursor of mechanisms, see Little (1986) and the "logic of an institution."

around three ontological hypotheses: (i) the social causation thesis, (ii) the micro-foundations thesis, and (iii) the agency-structure thesis. The basic methodological thesis is that sociohistorical scientific explanations do actually invoke "mechanisms" and also have to do so in order to be appropriate. This is the epistemic counterpart to the ontological hypothesis that "social causation" works somehow but only and solely through individual agency or action. A microfoundational explanatory account then basically provides an answer to the question how macro-powers get to manifest their dispositional natures in micro-action. The main assumptions are the following:

> Social structures and institutions have causal properties and effects that play an important role within historical change (the social causation thesis). They exercise their causal powers through their influence on individual actions, beliefs, values and choices (the microfoundations thesis). Structures are themselves influenced by individuals, so social causation and agency represent an ongoing iterative process (the agency-structure thesis). (Little 2010, p. 97; cf. 2007, p. 358; 2009, p. 169)

Methodological Localism is in somewhat more detail the hybrid out of the theses that "[s]ocial entities supervene upon individuals" (2007, p. 367) or "upon individuals and institutions" (2010, p. 56). Those social entities are "the sum of the constellation of socially situated individuals and institutions" (2007, p. 354) so "that all social facts are carried by socially constructed individuals in action" (2009, p. 159), which leads to "the idea that the causal nexus of the social world is the behaviours of socially situated and socially constructed individuals" (2011, p. 293).

Obviously, at least at first glance, this is rather uncontroversial. And equally salient, this ontological model is almost a reinvention of Bhaskar's TMSA. Accordingly, "influence" should here strictly be read in terms of causation.[46] We should consequently emphasize again the hypothesis that there is social stuff and that this stuff is furthermore supposed to be a bearer of "causal powers" and a producer or generator of individual action and derivative social facts.[47] Given the problems inherent in CR ontology, much depends on the ML theory of social structures and their powers, the mechanisms of CR.

In accordance with Bhaskar, who claimed that "social mechanisms and structures generating social phenomena" (1986, p. 122) are only relatively enduring, Little claims in accordance with most historians that "all social structures are historically rooted; so there is no fixed 'essential' nature of a state or economy" (2010, p. 75). That is, "historical individuals" (2010, p. 42) or "historical entities" (ibid. p. 47) always "'morph' over time" (ibid. p. 62). The central problem that guides the search for a theory of social mechanisms and mechanismic explanation is therefore the

[46] As far as I can see, Little quotes only Bhaskar's "Realist Theory of Science," in which the TMSA was not developed; cf. Little (2011, p. 278).

[47] To avoid misunderstanding, one should distinguish two claims under the heading social causation. The first is the claim that social macro stuff causes individual action. The second is that there is social macro-macro-causation whether through action or not; cf., already Sztompka (1991, p. 58). Of course, one could deny both claims. The easiest way to deny CR and ML styled social causation is to claim that "there are no structures" (Harré 2009, p. 138).

same as in Bhaskar's TMSA framework, "Agents constitute structures; and agents are in turn constituted by structures," which is then interpreted as some form of "ongoing mutual influence" or causation "within and across generations" (2007, p. 356).[48] The three main assumptions quoted above say in a nutshell that "[m]acro entities exercise causal properties through the individuals who constitute them at a given time" (2007, p. 366). Given that Little shares with Bhaskar and Bunge the aim of finding a middle way between ontological and methodological individualism and holism (2007, p. 346), we stick to the tradition established above by asking what does the so-called social world and perhaps history consist of and what are mechanisms? More to the point, what are social facts, macro-entities, or social structures in ML? What is it for such a "social thing" (2010, p. 72) to possess a causal property? And how do they cause individual actions and derivative social effects?

First of all, according to ML, there are "social things," for example, "relations, institutions, practices, organizations" (2010, p. 72), "historical individuals" or the "concrete social formation" (2010, p. 47), and furthermore "things as revolutions or capitalist economies" (2010, p. 42). At times, Little seems to suggest that "things" are that type of entity that bears "causal powers".

Second, in ML society, that is, a system of systems in ES and an ensemble of powers and internally related social positions in CR is thought to "consist of specific social, economic, and political institutions, mentalities and systems of beliefs and values, and higher-level structures that are composed of these institutions, practices, and mentalities." Agents are moreover claimed to "constitute" or "populate" these "social factors" and to act "within the context of these structures." Thus they "affect the future states of the system while being prompted or constrained by existing structures and mentalities" (2007, p. 353). This, then, amounts to the hypothesis that individuals are always "socially situated" in the sense that their "domain of choice" is restricted by existing "social institutions," that is, these decisions are caused by the given circumstances and their powers; in contrast to ES and in accordance with CR. Agents are furthermore said to possess "social properties," that is, they "exist in social relations and social institutions."[49] They are, moreover, "socially constructed" in the sense that their furniture of mind is acquired through interaction (2007, 353f., 2009, p. 174). Thus, agents are sometimes claimed to be in "social states" exemplified by "beliefs, intentions, reasoning, dispositions and histories" (2010, p. 59; 2007, p. 352).[50]

[48] As is well known in CR, the sociologist Margaret Archer (1995) formulated a similar theory.

[49] Although, as far as I can see, there is no concept analogous to structurally emergent properties in ES that accounts theoretically for this claim in ML. But this concept would not fit in here anyway because ML institutions or structures are, as it seems, not Bungean things or systems.

[50] A difference to Bunge's ontology is remarkable at this point, given that social states can be found neither in individuals' brains nor in individual actions according to ES (Bunge 1996, p. 45). Whereas in ES poking one's nose is not a social fact but an individual one, though poking another's nose or each other's noses are social facts, in ML the former is a social fact and a social action,

In summary, "the social" consists, according to ML, of individual agents that constitute, compose, or "embody" institutions. "Social institutions and organizations come together to constitute complexes of institutions" (2007, 354f.). Such complexes of institutions are called structures, which are the constituents of social formations, which are said to be the "comprehensive social entity at the macro-level" or "large systems" (2010, pp. 56–58).

What is a social structure, given its centrality as a causal agent and patient in "the agency-structure" thesis? In his early work, Little (1989, p. 24) believed a social structure to be "a set of constraints and incentives imposed on individual conduct and embodied in patterns of individual behavior." Examples of social or historical structures in his recent work are human organizations as, for example, a "rail system" (2011, p. 286) and "the fiscal system of the *ancient regime*" (2010, p. 3). A "fascist movement" and a "market" (2010, p. 88) are also called structures as are "the revolution of 1848" (2010, p. 81) and "large complexes of rules and practices" (2010, p. 75). Although it might seem that structures in ML sometimes are complex concrete things (formal organizations or systems in Bungean systemism), they sometimes resemble events or processes or even sets of rules or patterns of rule-governed individual actions. The last paradigm comes near the theory advanced in the earlier writing:

> a social structure is a system of geographically dispersed rules and practices that influence the actions and outcomes of large numbers of social actors. (Little 2010, p. 73)

Because anything else would probably amount to a harsh form of holistic idealism, it is claimed that each social entity is "constituted by the socially constructed individuals who make it up, through their beliefs, values, interests, actions, prohibitions, and powers" (2010, p. 56). As it seems, on the one hand, social entities are not necessarily concrete systems as in Bunge's systemism, in which systems are straightforwardly composed of men and women of flesh and blood. They are not composed of or constituted by actors' beliefs or values.

On the other hand, Little claims that agents "populate" "social factors," "institutions," and "higher-level structures"; what seems to make some sense if these are concrete objects. But those structures then are said to consist of "institutions, practices, and mentalities" (2007, p. 353), which can hardly be said to be populated or to be constituted by people. In any case, it remains unclear what this could mean. In order to clarify this and since formations are also said to be constituted by structures and these by institutions, it is worth to get a clearer picture of what institutions are according to ML:

> An institution (. . .) is an embodied set of rules, incentives, and opportunities that have the potential of influencing agents' choices and behavior. An institution is a complex of socially embodied powers, limitations, and opportunities within which individuals pursue their lives and goals. A property system, a legal system, and a professional baseball league

---

as is eating breakfast cereals or smoking for oneself in private, since we have somehow acquired every taste or preference by someone; cf. Little (2007, p. 351f.).

> all represent examples of institutions. Institutions have effects that are in varying degrees independent from the individual and "larger" than the individual. (Little 2007, p. 352)

Institutions, then, seem to be almost the same as what Little once called social structures, which was in part to be expected given that above structures were said to be "complexes" of such smaller institutions.

Although individuals in the end "embody" the whole social world, his position is intended not to be the same as ontological individualism, since (i) the individual is by itself social or socially constructed (2010, p. 58) and (ii) "social arrangements and circumstances affect individual action," that is, through the social situation (2007, p. 360). Because of this allegedly causal influence, social stuff is believed to exist in the first place.[51] Even more, large social stuff such as villages are supposed to be "lodged with a larger political, economic, and natural environment" that "influence and constrain" what is going on in the village. Thus, Little does not subscribe to what he calls "ultra-localism," the theory that the social world is exhausted by face-to-face-interaction (2007, p. 349), which, we should add, would considerably limit our possible thoughts about real histories.[52]

But how do such social facts or powers cause individual action and social facts? Little subscribes to the view he terms "causal realism":

> a thesis about the reality of causal mechanisms or causal powers. (2010, p. 101; 2011, p. 275)

Given the former section, the question arises how causal powers and mechanisms are related, given that in CR they are roughly the same and such passages in Little's work suggest a similar reading. Little is not very explicit about his theory of causal powers, though it is remarkably different from Bhaskar's due to an absence:

> What is it to attribute a causal power to an entity? It is to assert that the entity has a dispositional capacity to bring about specific types of outcomes in a range of causal fields. To have a causal power is to have a capacity to produce a certain kind of outcome in the presence of appropriate antecedent conditions. (Little 1998, p. 205)

What are the "entities" endowed with such causal powers, given that these, in early CR, were supposed to be complex objects or concrete particulars: "Only things and materials and people have 'powers'" (Bhaskar 1978, p. 78).

On Little's account "events, conditions, structures"; "institutions, ideologies, technological revolutions, communications, and transportation systems"; "properties, conditions, and events" (1998, 198f.); and "various social forces" (2011,

---

[51] To be more exact, Little writes (2007, p. 360, emphasis added): ML "is not equivalent to methodological individualism or reductionism because it admits that social arrangements and circumstances affect individual action. For it is entirely likely that a microfoundational *account* of the determinants of individual action will include reference to social relations, norms, structures, cognitive frameworks, etc."

[52] In ES this would be expressed by the claim that social systems (e.g., families) are as real as their members, face a social environment and might be the components of higher-level systems (e.g., villages), whereas it is apparently unclear what a village is in ML, given that it seems to be rather odd to say that it is a set of opportunities or a system of rules.

p. 275) as "social classes" (2010, p. 83) are supposed to be, as well as incentives and opportunities (1998, p. 206), and "rumors" (1991, p. 19) are possessors of causal powers. In short, every necessary condition is in accordance with early CR in the end a cause (1998, p. 200) and, by definition, a bearer of dispositional capacities.

This account and many examples given lack something which resembles a concrete social thing or system that bears the powers and produces something, so that there must be another trick than from some emergence basis in a complex thing to endow "social factors" with causal powers that somehow produce actions of individuals. How do such "things" as rules, opportunities, incentives, or mentalities come to possess powers to cause changes in concrete things according to ML?

The first theory embodied in ML might be named the theory of instantaneous power acquisition:

> Institutions and other aspects of social organization acquire their causal powers through their effects on the actions and intentions of the individuals involved in them – and *only* from those effects. (Little 1991, p. 19)

A rather strange but perhaps suggestive analogy might help here. For example, when the sun, although perhaps no social thing, causes modern people to go "sunbathing" by actually shining, because people decide to do so, it acquires the dispositional capacity to cause people to go sunbathing in the right and a variety of circumstances and instantaneously exercises it. This seems to be incoherent if causal powers are supposed to be intrinsic, dispositional, and therefore possessed even if not actualized and possessed before the thing causes anything or is triggered by some event. If this is not to be presupposed, we face a somewhat different concept of a power.

The second theory about the way the ascription of social causal powers can be justified might be termed the explanatory account of social dispositions. On this reading, "the causal capacities of social entities are to be explained in terms of the structuring of preferences, world views, information, incentives, and opportunities for agents" (2009, p. 170; 2007, p. 361; 2010, p. 106). If we can somehow explain microfoundationally how the shining sun provides modern people with an opportunity for sunbathing, that is, if we can sketch the "pathway" of how it manages to do that through people's heads, we are justified to claim that it simply possesses the causal capacity to bring about sunbathing people. But, at least originally, the "powers" of "things" were to be explained by their composition, structural organization, or their intrinsic natures, not by cultural accidents. Again, here we do not seem to face the concept of a causal power that grounded the realist tradition in philosophy of social science, but something else.

In a third reading we find what might be termed the relational or plausibility account of social causal powers or the theory of their rational calculation dependence. In this theory, that is similar to the former and perhaps explicates what "structuration" means, it is maintained that the causal powers of social stuff "derive from the incentives, powers and knowledge that these institutions provide for

participants" (2010, p. 106), that is, they "derive from the structured circumstances of the individuals who make up those entities, and from nothing else" (2007, p. 358). The result of the plausibility account is that social things "possess causal powers in a derivative sense: they possess characteristics that affect individuals' behavior in simple, widespread ways" (2007, p. 362; 2010, p. 106). According to this account, for example, the protestant ethic as a supposedly supervenient social entity, or perhaps an embodied mentality, causes a whole bunch of actions in or through individuals who happen to believe in protestant doctrines and are therefore incited by these. Furthermore, this has "historical changes" as unintended effects, if these people actually happen to act on those doctrines and thus instantiate the powers of Protestantism.

The problem with this account is that it is somehow plausible but quite fluffy, if we stick to the idea that also in the historico-social sciences or their ontology something explicit and perhaps exciting is associated with social causal powers or properties of some "thing."[53] But such social powers reside somewhere between the circumstances of acting people; the incentives, constraints, or opportunities; and their relation to the agents themselves, that is, their perception or interpretation of "social things." Virtually lost is what was thought to be in need of justification, given the realist tradition in philosophy of social science, namely, social things, their powers, and thereby productive social causation.

Put in different words, we might get the impression that the social powers metaphysics in philosophy of social science is grounded in theories about how "the social" constrains and enables actions. Though this might be perfectly alright in explanatory contexts, it seems to be questionable if this is enough to populate the world with social powers and causes.[54] Given that Little states that social causal capacities "are entirely defined by the current states of psychology, norm, and action of the individuals who currently exist" (Little 2007, p. 347; 2009, p. 166; 2010, p. 61), critics of realist social ontologies might simply want to expel the powerful ghost without a machine altogether or assign him a powerless status. Reification

[53]Let us take the risk to pose some naïve questions: Which is the disposition or power of, say, a mentality? Is a mentality, or a norm (or what have you), a property? If yes, of what? If it is not a property of something, where is it floating? In a different context, Sztompka (1991, p. 23) has seen clearly the problem we seem to face: "In modern sociology one may find such fashionable and influential notions as 'habitus' (Bourdieu), 'historicity' (Touraine), 'figurations' (Elias), 'mobilization' (Etzioni), 'anomie' (Merton), 'duality of structure' (Giddens), 'agency' (Archer) – and many others. It is not easy to say what exactly the referents of these concepts are, what kinds of objects are described, because clearly they are neither people nor systems."

[54]See the quote in note 51. Causation might be one thing, explanation quite another. To say the same more carefully, one should be careful not to slide into an ontological misinterpretation of the famous "Thomas Theorem," which says "If men define situations as real they are real in *their* consequences" (quoted in Sztompka 1991, 83, emphasis added). Of course, there might be nothing social beyond or behind the heads of people that has "powers" or the former consequences, that is, that causes actions or social changes. Because of such worries Boudon (2010, 23) calls powers or causes such as mentalities or social structures "forces fantomatiques." Again, this is only supposed to indicate that there is something problematic about social causation or social powers.

of "the social" might not be "the attribution of causal powers to entities without an understanding of the mechanisms through which those powers are expressed" (2007, p. 350), but the attribution of causal powers to entities that are not concrete systems or what was formerly called a powerful particular.[55]

After all these preliminaries, what are mechanisms and what is their function in ML? Are mechanisms complex things, processes, properties, or do they embody the whole spectrum? The question suggests itself, given that in the quotation above it seemed as if powers were roughly mechanisms, whereas in the ending of the last paragraph, mechanisms are said to be something where powers are "expressed," which suggests that mechanisms are events or processes, that is, they are supposed to be exactly what they are explicitly not supposed to be in CR.

The role for mechanisms in ML is first of all to "mediate" social causation (1998, p. 203), that is, to be the medium of the "expression" of macro-powers and to be the connection between macro-cause and macro-effect (2011, 278f.). Although it is slightly misleading, we can picture this the following way:

Macro → Mechanism → Macro.

Accordingly, the first theory of mechanisms in ML was straightforwardly that of a causal chain.[56] This line of thought seems to persist in Little's recent writing: "A causal mechanism is a series of events or processes that lead from the explanans to the explanandum" (2007, p. 357). In the case of social causal mechanisms, it is "a set of social conditions, constraints, or circumstances combined to bring about a given outcome" (2009, p. 168).

Given the ontological microfoundations thesis and causal realism, "individual actors embody this causal process" (2007, p. 347; 2010, p. 61), that is, the process that is central to the agency-structure thesis.

Social structure → Individual actions → Social structure.

Personal and social powers are also said to be part of the metaphysical "substrate" of the mechanism that brings about social processes. They provide those processes with fuel that is burnt only and solely in actions. The story can be abridged insofar as in the social world "causal mechanisms are constituted by the purposive actions of

[55] For such criticisms see again (Lewis 2000; Harré 2002; Manicas 2006; Kaidesoja 2007).

[56] Little (1991, p. 15): "A causal mechanism (. . .) is a series of events governed by lawlike regularities that lead from the explanans to the explanandum." That mechanisms are chains of events is still suggested in his recent work when he writes that mechanisms have two ends; cf. Little (2011, 278).

agents within constraints" (2011, p. 273). But finally, we get Little's recent account of causal mechanisms:

> A causal mechanism is (i) a particular configuration of conditions and processes that (ii) always or normally leads from one set of conditions to an outcome (iii) through the properties and powers of the events and entities in the domain of concern. (Little 2010, p. 102; 2011, p. 277)

Accordingly, a mechanisms is in ML roughly what is called a process or an event in CR, the place where in CR individual and social mechanisms (powers) get actualized, which are here mainly to be found under clause (iii). At the same time they resemble slightly what in ES is called a concrete system, because what is there called a component of a system (and perhaps also what figures as the environment of a system and as its structure in ES) is in ML a part of a mechanism. But a clear notion of a social thing seems to be absent from ML as it is from CR, although examples for mechanism such as "the feudal manor, the collective farm, the Wall Street law firm" (2011, p. 284) are straightforwardly systems in Bunge's rather clear sense, not ES mechanisms. Therefore, the little process schemas above are slightly misleading because in the end social structures and social institutions are as well part of a social causal mechanism in ML as are persons that constitute structures, institutions, and formations, which together bring about the "behavior" of the social mechanism (ii). Yet, similar to Bunge, ML social mechanisms happen neither to be running in persons nor in "structures" but are somehow the resultant of both, that is, of actions within "constraints." The difference is that ML mechanisms might also be said to be constituted by or composed of persons and social structures, which is, of course, impossible in ES.

And, finally, what is history? Little clothes his philosophy of history into the metaphor of a pathway: "history is an accumulation of pathways and roadways that embed human action over time" (2010, p. 9). But isn't that, roughly, a social structure or a social mechanism?

10.5 Do They or Do They Not?

At the workshop that anteceded the publication of this volume, the question arose how central mechanisms are to the historian or history. If we want to refrain from answering this question by the traditional stories of philosophy of historiography that tell us that historians are by their essence, that is, by definition, interested in individual actions, singular or unique "big events," the aesthetics of narratives, or what have you, we would need to know what people that are sometimes called historians do, what mechanisms are, and what history is.

Since I do not claim to know what so-called typical historians do, we have to speculate about this aspect of our question. If we follow Bunge's materialism, we even have to discard the whole question. Why? Simply because it does not make any sense to ask about the mechanisms of history because the latter is conceptual

according to ES. No single overarching history exists, although it is perfectly plausible to investigate the history of any thing or system or to hypothesize about the mechanism for an aspect of a history of such a thing.[57] We also saw that the leading philosophers of social mechanisms disagree remarkably on what mechanisms are and on further basic ontological assumptions, so that it is not by any means clear what a social mechanism is supposed to be.

Social mechanisms turn out to be processes occurring essentially in a kind of social system (ES), social dispositional properties or powers (CR), or configurations of processes and events, conditions, powerful people, and powerful social entities that regularly bring about social processes (ML) (Problem I: What is a social mechanism?). We furthermore witnessed notorious differences about the place of causation and determination in ontologies of the so-called social world, which looks rather different through these philosophies. For instance, "social entities" (Problem II: What is the furniture of the "social world"?) turn out to be causal agents or factors for some, not for others (Problem III: What is causation?). Thereby we also should have seen that the merits of philosophical theories of social mechanisms, causation, and social explanation are in the end only assessable if they are analyzed in their respective philosophical or social theoretical environments.[58] In any case and in any of these philosophies, central notions such as "constraint" and "enablement" or the thought that the "environment does not act *on* a person, but rather *through* a person" (Bunge 2009a [1959], p. 181), what Marx famously called *Zwang der Verhältnisse,* remain worth disputing, although these are classic questions of social theory and philosophy (Problem IV: What is determining or causing change or stasis in "the social world"?).

It is hardly worth mentioning that the "structure vs. agency" issue is the most central problem in social theory and social research, and it figures prominently in mechanism discourses. More basically, this amounts also to the question what does change in "the social world," if anything (Problem V: What is a history?). If there are no "social entities" or "social structures," there is no social change, if change presupposes something that changes. Of course, different configurations of responses to these problems lead also to different opinions about the possible explananda of social or historiographical (*geschichtswissenschaftlich*) research or

[57] Nota bene, philosophers of historiography, many historians and sociologists constantly talk about such an overarching history or they never make explicit what they believe they are talking about while writing about history in a realist or ontic sense.

[58] This also holds for formerly notorious questions about the role of "laws" in historico-social science. Whereas for Bunge "mechanisms without conceivable laws are called 'miracles'" (2006, p. 135), Little's ML claims to be something like a counterprogram to the usefulness of social "laws," whereas critical realists seem to accept restricted ("historical") tendencies as such "laws." For problems critical realist have with the notion of a law, see Outhwaite (1987b). Recently, there has emerged a powerful tendency towards an affirmative consensus in twenty-first century philosophy of historiography concerning the claim that the people that are called historians constantly invoke "laws"; see, for example, Klinger (1997), Di Nuoscio (2004), Antiseri (2005), Frings (2007), Berry (2008), Leuridan and Froeyman (2012).

about the stuff merely to be described and, finally, to different norms of how sociological or historiographical explanations have to be framed, for example, whether those explanations are causal explanations, and if the answer is yes, in which way. (Problems VI: What is to be understood and how is it accordingly to be explained? See also Blaikie 1993). In a nutshell, given the former sketch of a comparison of some of the main positions on social mechanisms, it seems to be so that the problem of social mechanisms comes with a whole bunch of others, and the sketch suggests that they should perhaps be more clearly separated or related in future debates. Otherwise, it might be that those scholars who believe we are dealing here with fruitless "mechanism talk" are right.

But let us answer the main question: If we presuppose the ontology of mechanisms in ES and therefore the whole framework, it is arguably beyond doubt that many historians study mechanisms that make past or present systems tick and made them what they were, became, or still are.

If we presuppose the ontology of mechanisms in CR, it is arguably beyond doubt that many historians study mechanisms (social structures) or invoke individual or social powers (mechanisms) in explanations of individual or social events. Though we equally found out that "[i]t is not by any means obvious what the concept of mechanism refers to in philosophical and critical realist frameworks (. . .)" (Kurki 2008, p. 177). This diagnosis is mirrored in Bhaskar's statement (1978, p. 49) that a "generative mechanism" is "a 'real something' over and above and independent of patterns of events." Moreover, ascribing powers to "emergent" social entities seemed to be problematic.[59] Though it was not the central point in our discussion, CR clearly implicates some philosophy of history.

If we presuppose the ontology of mechanisms in ML, it is arguably beyond doubt that many historians study mechanisms,[60] although the same problems that were diagnosed in connection with CR concern the ML framework. Perhaps because of the affiliations to realist traditions, the concept of mechanisms as powerful properties and mechanisms as processes is often not clearly drawn. In summary, there is no reason whatsoever to believe that historians do not study mechanisms. This is the meager but positive result of this sketch.

But these answers are obviously utterly unsatisfying because the following question is which of these ontological frameworks or philosophies of history, if any, is adequate and why, given that they disagree on almost any central category, although we did only visit the main realist positions in social philosophy.[61] This is the negative result of this sketch. While their merits can only be further evaluated in

[59] Bhaskar, of course, saw himself the problems that can occur in realist social ontologies (1982, 283): "Talk of 'emergence' can easily become vague and general, if not indeed laced with frankly idealist or romantic overtones."

[60] For lots of examples, see (Little 1989, 2010).

[61] Since we cannot discuss all the differences in these frameworks and I do not claim to be a metaphysician anyhow, let us list which notions are at stake in this debate: thing, property, types of properties, social property, change and transformation, event (historical), process, history, mechanism, structure, system, society, organization, institution, fact, social fact, causation and

comparison with accounts of mechanisms coming from other parts of social science and philosophy as well as in confrontation with historiological practice, to borrow Bunge's term, the merits of this exercise might be found in the simple observation that we did something resembling an ontology of history without engaging in speculation about the "course of history" and whether it moves in circles or squares, that is, we did not set out "historiosophical schemes" (Sztompka 1991, p. 182). Thus, our discussion suggests, among other "things," the need for an ontology of history, *if* the main question shall be answered on occasion. For short, there seem to be not many reasons to bury any philosophy of history from the start.

Finally, let us dwell a moment upon the philosophical problems that result from the claim, explicit in at least two of the sketched ontologies and furthermore as often supposed to be trivially true as it is straightforwardly denied, that (all) things have histories, that is, that they have properties that change. Those problems basically are which are those (social) things that have histories? What are their (social) properties that change? Why do they change or remain constant? Those questions do not seem to be far away from some of the questions some scientific historians set out to answer: What exists or existed? How does it change or how did it change? Why did it change or remain what it was or even still is?[62] And if I am not totally wrong, those are the more basic questions that are at issue in the debate over social mechanisms, although they are often hidden.

Obviously, the formulation above presupposes that things have histories. But isn't it a central implicit assumption of much philosophy of historiography that events have histories? Perhaps this might be one puzzle for philosophy of history, although it might also turn out to be ill posed.

Acknowledgments This study has been made possible by a grant of the Deutsche Forschungsgemeinschaft (DFG) and was realized in the project *Explanation, Laws and Causality in Historical Science,* a subunit of the research group *Causation, Laws, Dispositions and Explanation at the Intersection of Science and Metaphysics*. I thank Oliver R. Scholz for saving me from the biggest nonsense in history. The flowers go, as always, to Eileen.

References

Antiseri, D. (2005). Epistemologia e didattica della storia. Le ragioni della storiografia locale. In E. Di Nuoscio (Ed.), *Conoscere per tracce. Epistemologia e storiografia*. Milano: Edizioni Unicopli.

Archer, M. (1995). *Realist social theory: The morphogenetic approach*. Cambridge: Cambridge University Press.

---

determination, energy and causal power, laws, agency, agents and action, levels, micro-x vs. macro-x, emergence, etc.

[62] Because this variety of ontology of history would take its stock of questions from debates among historians and social scientists, on the one hand, and the implications or presuppositions of their research, on the other hand, I call it loosely "science-oriented".

Barberi, F. (2004). *Meccanismi sociali. Elementi di sociologia analitica*. Bologna: Il Mulino.
Bennett, A., & George, A. L. (2005). *Case studies and theory development in the social sciences*. Cambridge: MIT Press.
Benton, T. (1977). *Philosophical foundations of the three sociologies*. London: Routledge.
Berry, S. (2008). Laws in history. In A. Tucker (Ed.), *A companion to philosophy of history and historiography* (pp. 162–171). Malden: Blackwell.
Bhaskar, R. (1979). *The possibility of naturalism. A philosophical critique of the contemporary human sciences*. Brighton: The Harvester Press.
Bhaskar, R. (1982). Emergence, explanation, and emancipation. In P. F. Secord (Ed.), *Explaining human behavior. Consciousness, human action and social structure* (pp. 275–310). Beverly Hills: SAGE.
Bhaskar, R. (1983). Beef, structure and place: Notes from a critical naturalist perspective. *Journal for the Theory of Social Behaviour, 13*, 81–96.
Bhaskar, R. (1994). *Plato etc. The problems of philosophy and their resolution*. London: Verso.
Bhaskar, R. (2008 [1978]). A realist theory of science. London: Verso.
Bhaskar, R. (2008 [1993]). Dialectic. The pulse of freedom. London: Routledge.
Bhaskar, R. (2009 [1986]). Scientific realism and human emancipation. London: Routledge.
Bhaskar, R. (2011 [1989]) Reclaiming reality. A critical introduction to contemporary philosophy. London: Routledge.
Bhaskar, R. (2001). How to change reality: Story v. structure – a debate between Rom Harré and Roy Bhaskar. In J. Lopez & G. Potter (Eds.), *After postmodernism. An introduction to critical realism* (pp. 22–39). London: The Athlone Press.
Blaikie, N. (1993). *Approaches to social enquiry*. Cambridge: Polity Press.
Boudon, R. (2010). *La sociologie comme science*. Paris: La Découverte.
Bunge, M. A. (1965). Phenomenological theories. In M. Bunge (Ed.), *The critical approach to science and philosophy* (pp. 234–254). London: The Free Press of Glencoe.
Bunge, M. A. (1967). *Scientific research* (Vol. 2). Heidelberg: Springer.
Bunge, M. A. (1968). The maturation of science. In I. Lakatos & A. Musgrave (Eds.), *Problems in the philosophy of science* (pp. 120–137). Amsterdam: North-Holland.
Bunge, M. A. (1974). Les présupposés et les produits métaphysiques de la science et de la technique contemporaines. *Dialogue, 13*, 443–453.
Bunge, M. A. (1977). *The furniture of the world. Treatise on basic philosophy, Ontology I* (Vol. 3). Dordrecht: D Reidel.
Bunge, M. A. (1979a). *A world of systems. Treatise on basic philosophy, Ontology II* (Vol. 4). Dordrecht: D Reidel.
Bunge, M. A. (1979b). A systems concept of society: Beyond individualism and holism. *Theory and Decision, 10*, 13–30.
Bunge, M. A. (1981). *Scientific materialism*. Dordrecht: D Reidel.
Bunge, M. A. (1982). The revival of causality. In G. Floistad (Ed.), *Contemporary philosophy. A new survey, philosophy of science* (Vol. 2, pp. 133–155). The Hague: Martinus Nijhoff Publishers.
Bunge, M. A. (1983). *Understanding the world. Treatise on basic philosophy, epistemology & methodology II* (Vol. 6). Dordrecht: D Reidel.
Bunge, M. A. (1984). *Das Leib-Seele-Problem. Ein psychobiologischer Versuch*. Tübingen: J C B Mohr.
Bunge, M. A. (1985). *Life science, social science and technology. Treatise on basic philosophy, epistemology & methodology III: Philosophy of science and technology* (Vol. 7). Dordrecht: D Reidel.
Bunge, M. A. (1988). The scientific status of history. In U. Hinke-Dörnemann (Ed.), *Die Philosophie in der modernen Welt. Gedenkschrift für Prof. Dr. med. Dr. phil. Alwin Diemer, Teil 1* (pp. 592–602). Frankfurt: Peter Lang.
Bunge, M. A. (1992a). System boundary. *International Journal of General Systems, 20*, 215–219.
Bunge, M. A. (1992b). Systems everywhere. In C. V. Negoita (Ed.), *Cybernetics and applied systems* (pp. 23–41). New York: M Dekker.

Bunge, M. A. (1993a). Realism and antirealism in social science. *Theory and Decision, 35*, 207–235.
Bunge, M. A. (1993b). Social systems. In R. R. Delgado & B. H. Banathy (Eds.), *International systems science handbook* (pp. 210–221). Madrid: Systemic Publications.
Bunge, M. A. (1995). *Sistemas sociales y filosofía*. Buenos Aires: Editorial Sudamericana.
Bunge, M. A. (1996). *Finding philosophy in social science*. New Haven: Yale University Press.
Bunge, M. A. (1997). Mechanism and explanation. *Philosophy of the Social Sciences, 27*, 410–465.
Bunge, M. A. (1998a). *Social science under debate: A philosophical perspective*. Toronto: University of Toronto Press.
Bunge, M. A. (1998b). *Philosophy of science* (Vol. 2). New Brunswick: Transaction.
Bunge, M. A. (1999). *The sociology-philosophy connection*. New Brunswick: Transaction.
Bunge, M. A. (2000a). Systemism: The alternative to individualism and holism. *Journal of Socio-Economics, 29*, 147–157.
Bunge, M. A. (2000b). Ten modes of individualism – none of which works – and their alternatives. *Philosophy of the Social Sciences, 30*, 384–406.
Bunge, M. A. (2001a). Systems and emergence, rationality and imprecision, free-wheeling and evidence, science and ideology: Social science and its philosophy according to van den Berg. *Philosophy of the Social Sciences, 31*, 404–423.
Bunge, M. A. (2001b). *Philosophy in crisis. The need for reconstruction*. Amherst: Prometheus.
Bunge, M. A. (2002). *Ser, saber, hacer*. México: Paidós.
Bunge, M. A. (2003a). *Emergence and convergence. Qualitative novelty and the unity of knowledge*. Toronto: University of Toronto Press.
Bunge, M. A. (2003b). *Cápsulas*. Barcelona: Gedisa.
Bunge, M. A. (2003c). *Philosophical dictionary*. Amherst: Prometheus.
Bunge, M. A. (2004a). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences, 34*, 182–210.
Bunge, M. A. (2004b). Clarifying some misunderstandings about social systems and their mechanisms. *Philosophy of the Social Sciences, 34*, 371–381.
Bunge, M. A. (2006). *Chasing reality. Strife over realism*. Toronto: University of Toronto Press.
Bunge, M. A. (2007). Review of *Dissecting the social. On the principles of analytical sociology*, by Peter Hedström. *American Journal for Sociology, 113*, 258–260.
Bunge, M. A. (2008). *Filosofía y sociedad*. México: Siglo XXI.
Bunge, MA. (2009a [1959]). Causality and modern science. New Brunswick: Transaction Publishers.
Bunge, M. A. (2009b). *Political philosophy. Fact, fiction and vision*. New Brunswick: Transaction.
Bunge, M. A. (2010). Soziale Mechanismen und mechanismische Erklärungen. *Berliner Journal für Soziologie, 20*, 371–381.
Bunge, M. A., & Mahner, M. (1997). *Foundations of biophilosophy*. Heidelberg: Springer.
Bunge, M. A., & Mahner, M. (2001). Function and functionalism. A synthetic perspective. *Philosophy of Science, 68*, 75–94.
Bunge, M. A., & Mahner, M. (2004). *Über die Natur der Dinge. Materialismus und Wissenschaft*. Stuttgart: S Hirzel.
Cherkaoui, M. (2005). *Invisible codes. Essays on generative mechanisms*. Oxford: The Bardwell Press.
Collingwood, RG. (1994 [1946]). The idea of history. Oxford: Oxford University Press.
Danermark, B., Ekström, M., Jokobson, L., & Karlsson, J. C. (2002). *Explaining society. Critical realism in the social sciences*. London: Routledge.
Demeulenaere, P. (Ed.). (2011). *Analytical sociology and social mechanisms*. Cambridge: Cambridge University Press.
Di Nuoscio, E. (2004). *Tucidide come Einstein? La spiegazione scientifica in storiografia*. Soveria Mannelli: Rubbettino Editore.
Elder-Vass, D. (2010). *The causal power of social structures. Emergence, structure and agency*. Cambridge: Cambridge University Press.

Elster, J. (2007). *Explaining social behavior. More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
Fleetwood, S. (2009). The ontology of things, properties and powers. *Journal of Critical Realism, 8*, 343–366.
Fleetwood, S. (2011). Powers and tendencies revisited. *Journal of Critical Realism, 10*, 80–99.
Frauley, J., & Pearce, F. (Eds.). (2007). *Critical realism and the social sciences. Heterodox elaborations*. Toronto: University of Toronto Press.
Frings, A. (2007). Rationales Handeln und historisches Erklären. *Journal for General Philosophy of Science, 38*, 31–56.
Gibbon, G. (1989). *Explanation in archaeology*. Oxford: Blackwell.
Glynos, J., & Howarth, D. (2007). *Logics of critical explanation in social and political theory*. London: Routledge.
Groff, R. (2004). *Critical realism, post-positivism and the possibility of knowledge*. London: Routledge.
Harré, R. (1961). *Theories and things. A brief study in prescriptive metaphysics*. London: Sheed and Ward.
Harré, R. (1970). *The principles of scientific thinking*. London: Macmillan.
Harré, R. (1972). *The philosophies of science. An introductory survey*. Oxford: Oxford University Press.
Harré, R. (2002). Social reality and the myth of social structure. *European Journal of Social Theory, 5*, 111–123.
Harré, R. (2009). Saving critical realism. *Journal for the Theory of Social Behaviour, 39*, 129–143.
Harré, R., & Madden, E. H. (1975). *Causal powers. A theory of natural necessity*. Oxford: Blackwell.
Harré, R., & Varela, C. M. (1996). Conflicting varieties of realism: Causal powers and the problems of social structure. *Journal for the Theory of Social Behaviour, 26*, 313–325.
Hartwig, M. (Ed.). (2007). *Dictionary of critical realism*. London: Routledge.
Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.
Kaidesoja, T. (2007). Exploring the concept of causal power in a critical realist tradition. *Journal for the Theory of Social Behaviour, 37*, 63–87.
Kaidesoja, T. (2009). Bhaskar and Bunge on social emergence. *Journal for the Theory of Social Behaviour, 39*, 300–322.
Klinger, W. (1997). Legge e spiegazione in storia: un approccio naturalistico, Dissertation, Università degli Studi di Trieste.
Kurki, M. (2008). *Causation in international relations. Reclaiming causal analysis*. Cambridge: Cambridge University Press.
Lawson, T. (1997). *Economics and reality*. London: Routledge.
Leuridan, B., & Froeyman, A. (2012). On lawfulness in history and historiography. *History and Theory, 51*, 172–192.
Lewis, P. (2000). Realism, causality and the problem of social structure. *Journal for the Theory of Social Behaviour, 30*, 249–268.
Little, D. (1986). *The scientific Marx*. Minneapolis: University of Minnesota Press.
Little, D. (1989). *Understanding peasant China. Case studies in the philosophy of social science*. New Haven: Yale University Press.
Little, D. (1991). *Varieties of social explanation. An introduction to the philosophy of social science*. Boulder: Westview.
Little, D. (1998). *Microfoundations, method, and causation: on the philosophy of the social sciences*. New Brunswick: Transaction.
Little, D. (2007). Levels of the social. In S. P. Turner & M. W. Risjord (Eds.), *Philosophy of anthropology and sociology* (pp. 343–371). Amsterdam: Elsevier.
Little, D. (2009). The heterogenous social: new thinking about the foundations of the social sciences. In C. Mantzavinos (Ed.), *Philosophy of the social sciences. Philosophical theory and scientific practice* (pp. 154–178). Cambridge: Cambridge University Press.

Little, D. (2010). *New contributions to the philosophy of history*. Dordrecht: Springer.
Little, D. (2011). Causal mechanisms in the social realm. In P. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 273–295). Oxford: Oxford University Press.
Lloyd, C. (1986). *Explanation in social history*. Oxford: Blackwell.
Lloyd, C. (1993). *The structures of history*. Oxford: Blackwell.
Mahner, M. (Ed.). (2001). *Scientific realism. Selected essays of Mario Bunge*. Amherst: Prometheus.
Manicas, P. T. (2006). *A realist philosophy of social science. Explanation and understanding*. Cambridge: Cambridge University Press.
Marrou, HI. (1975 [1954]). De la connaissance historique. Paris: Éditions du Seuil.
McLennan, G. (1981). *Marxism and the methodologies of history*. London: Verso.
Moessinger, P. (2008). *Voir la sociéte. Le micro et le macro*. Paris: Hermann Éditeurs.
Outhwaite, W. (1987a). *New philosophy of social science. Realism, hermeneutics and critical theory*. Houndmills: Macmillan.
Outhwaite, W. (1987b). Laws and explanation in sociology. In R. J. Anderson, J. A. Hughes, & W. W. Sharrock (Eds.), *Classic disputes in sociology* (pp. 157–183). London: Allen & Unwin.
Pickel, A. (2006). *The problem of order in the global age. Systems and mechanisms*. New York: Palgrave.
Porpora, D. V. (2007). Social structure. In M. Hartwig (Ed.), *Dictionary of critical realism* (pp. 422–425). London: Routledge.
Sayer, A. (2010a [2000]). Realism and social science. London: SAGE.
Sayer, A. (2010b). *Method in social science. A realist approach*. London: Routledge.
Schmid, M. (2006). *Die Logik mechanismischer Erklärungen*. Wiesbaden: VS-Verlag.
Sewell, W. H. (2005). *Logics of history. Social theory and social transformation*. Chicago: The University of Chicago Press.
Sztompka, P. (1991). *Society in action. The theory of social becoming*. Chicago: The University of Chicago Press.
Tilly, C., McAdam, D., & Tarrow, S. (2001). *Dynamics of contention*. Cambridge: Cambridge University Press.
Topolski, J. (1976). *Methodology of history*. Dordrecht: D Reidel.
Tucker, A. (2012). Sciences of historical tokens and theoretical types: History and the social sciences. In H. Kincaid (Ed.), *The oxford handbook of the philosophy of social science* (pp. 274–297). Oxford: Oxford University Press.
Veyne, P. (1996 [1975]). Comment on écrit l'histoire. Paris: Éditions du Seuil.
Wan, P. Y.-z. (2011a). *Reframing the social. Emergentist systemism and social theory*. Farnham: Ashgate.
Wan, P. Y.-z. (2011b). Analytical sociology: A Bungean appreciation. *Science and Education*. doi:10.1007/s11191-011-9427-3.
Wight, C. (2006). *Agents, structures and international relations. Politics as ontology*. Cambridge: Cambrige University Press.

Chapter 11
Philosophy of History: Metaphysics *and* Epistemology

Oliver R. Scholz

Abstract Some of the most important questions historians have to answer are "What happened in the past?" and "Why did it happen?" and the epistemological question "How do we know?" or, more modestly, "How are our historical hypotheses epistemically justified?" It is important to note that answers to these questions require not only epistemological but also metaphysical, especially ontological, investigations. Due to the failures of speculative metaphysics of history (in the style of Augustine, Hegel, and Marx), metaphysical questions were frowned upon by recent philosophy of history. Thus, the focus has been on questions of logical form, conceptual analysis, and methodology (analytical philosophy of history) on the one hand and on questions of the literary and rhetorical forms of historical representations on the other hand (narrativism). In both research programs, the reality of history is in danger of disappearing. By discussing recent attempts to reduce the philosophy of history to the epistemology of historiography, I will argue that philosophy of history and scientific historiography are in need of metaphysical, especially ontological, investigations without falling back into the fallacies of a speculative metaphysics of history. Finally, the fertility of such enquiries shall be illustrated by raising an important question, namely, "How close can the contact with the historical past be?" and by attempting an answer.

Keywords Philosophy of history • Metaphysics • Ontology

O.R. Scholz (✉)
Philosophisches Seminar, Westfälische Wilhelms-Universität Münster, Domplatz 23, 48143 Münster, Germany
e-mail: oscholz@uni-muenster.de

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_11,

11.1 Philosophy of History

Nowadays, old-style philosophy of history (in the manner of, say, Augustine, Vico, Herder, Hegel, Marx, or Spengler) is not of high repute. The fanciful vision of history inevitably approaching a final destination and the pretension of knowing this necessary historical progression a priori are rightly considered as discredited. This form of philosophy of history rested both on bad metaphysics and on the neglect of epistemological investigations.

Currently, a different mistake is about to impend: partly because of the failure of speculative philosophy of history, partly due to prejudices against metaphysics in general, metaphysical questions are dismissed altogether in contemporary philosophy of history.[1] Thus, philosophers confine themselves either to epistemological and methodological questions (analytical philosophy of history) or to questions about the literary and rhetorical forms of historical representations (so-called narrativism). Put another way, a reduction of philosophy of history impends, either to epistemology or methodology of historical science (analytical philosophy of history) or to rhetorical or literary studies on the form of historical representations (narrativism), with a neglect of all metaphysical issues.

Since the narrativists frequently exaggerate the importance of the "reshaping" of data or even consider it to amount to a fictionalization, they typically end up in endorsing extremely anti-realist theories of history. Ontological anti-realism about history is usually combined with a radical relativism, preferably in the form of social constructivism stating that every society constructs its own past in accordance with its prevailing (non-epistemic) needs and interests. Anti-realism concerning the historical past is not only poorly justified,[2] it is also cynical toward the victims of history. All the atrocities and all the suffering history is filled with did not originate from intellectual constructs but rather from concrete persons, actions, and events.

Analytical philosophers of history are inclined to confine philosophy of history to the epistemology and methodology of historiography, that is, to the philosophy of scientific historiography. Thus, in his important book *Our Knowledge of the Past*, Aviezer Tucker decisively does not develop a philosophy of history but only a philosophy of historiography.[3] In a similar vein, Peter Kosso has emphasized that "[t]he philosophical issues in the analysis of historiography are almost entirely epistemological."[4]

---

[1] Similar problems can be found in natural philosophy. Here, too, speculative flights of fancy were followed by a total renunciation of metaphysics.

[2] For pertinent arguments against various forms of relativism, anti-realism, and constructivism, see Boghossian (2006).

[3] Tucker (2004).

[4] Kosso (2009), p. 9. Nota bene, I am only complaining about the neglect of ontology of history. To be sure, with regard to the epistemology and methodology of the historical sciences, Kosso (2001, 2009) and Tucker (see especially Chap. 3 of Tucker 2004) have made very important contributions.

In both of these research programs, historical reality is in the danger of being moved to the background or even to disappear. Within narrativism, it mutates into an aesthetic artifact, an artwork produced by a historiographer kissed by the muse Clio. In some trends of analytical philosophy of history, it wastes away into an aggregate of past events cut off from us by an unbridgeable chasm, a past which can at most be reached via hazardous causal inferences. Whereas early analytical philosophy of history was at least worrying about the explanation of historical *events*, more recent authors, such as Tucker, restrict their investigations to the question "How is the historical *evidence* explained?"

In the following, it will be shown at which points philosophy of history as well as scientific historiography relies on answers to metaphysical, in particular ontological, questions, without relapsing into the errors of speculative metaphysics of history. The brand of metaphysics I am suggesting is not fanciful, but disenchanted and analytical. It stands in the tradition of Aristotelian metaphysics and its renaissances.[5] Especially the project of a category theory or categorial ontology shall be made fertile for an ontology of history.

The principal thesis of this chapter is that *ontological* inquiries belong to the philosophy of history no less than *epistemological* investigations. This becomes plain when you (a) consider the historical sciences in their whole extent, (b) take seriously the most important questions historians should answer, and (c) keep in view the sources of knowledge that are available to them. In the following, I advocate a comprehensive ontology of history and begin to sketch some of its questions. Finally, the fertility of such enquiries shall be illustrated by raising an important question, namely, "How close can the contact with the historical past be?" and by attempting to answer this question.

11.2 The Spectrum of the Historical Sciences

Let us start with the concept "historical sciences." By "historical sciences" (in a broad sense), I mean all sciences whose inquiries are directed at the past (including, of course, its effects on the present). These sciences include, inter alia, cosmology, geology, and evolutionary biology. When talking about history and historical sciences within the context of philosophy of history, people typically think of history in a narrow sense: roughly, as all the sciences asking questions about that part of space-time that has been influenced by individual and collective actions of human beings (or that could at least have been influenced thus). These comprise, inter alia, political history, economic history, church history, and military history, as well as comparative linguistics, literary history, art history, history of science, and history of philosophy.

[5] See, for instance, Loux (2006) and Schaffer (2009).

11.3 Questions for the Historical Sciences

Scientific disciplines are often characterized by their respective objects and the methods to be applied. I prefer to focus on the pivotal questions that characterize the respective discipline. Which questions do historians wish to answer and which questions should they answer? Which questions are requested to be answered by the consumers of historiography? First of all, there are two main questions:

1. What happened in the past?
2. Why did it happen?

With each of these questions, we face the epistemological question:

3. How do we know? Or, more modestly, how are the respective hypotheses epistemically justified?

From an ontological point of view, an improvement of question (1) becomes already apparent, since (1) was tailored one-sidedly to happenings or events. The more comprehensive questions read as follows:

(1.1*) What was the world (or a certain part of it) like at time t_i? Which properties did the world (or a certain part of it) possess at time t_i?

(1.2*) How did the world (or a certain part of it) change between t_i and some later time t_{i+n}?

(2.1*) Why was the world (or a certain part of it) at t_i the way it was? How did it come about that the world (or a certain part of it) at t_i possessed those properties?

(2.2*) Why did the world (or a certain part of it) change between t_i and the later time t_{i+n}?

11.4 From Epistemology to Metaphysics: Objects and Sources of Historical Knowledge

That a metaphysics of history is needed already ensues from epistemological considerations. To see this, let us consider (a) the objects of historical knowledge, and (b) the sources of knowledge that are at our disposal.

(a.1) The historical sciences investigate the past and the development from the past to the present. Therefore, a metaphysics of time and change is required. (a.2) The historical sciences deal with human beings and their deeds and omissions. Human beings are persons and agents; thus, we further need a metaphysics of personhood and personal identity as well as a metaphysics of actions and omissions. (a.3) Human beings are cultural beings; they develop and pass on their culture. In addition to an ontology of the natural world, we therefore need an ontology of the cultural world. (a.4) Human beings are social beings; they build communities, societies, and institutions. Correspondingly, we are also in need of a social ontology.

(b) Let us now consider the sources of knowledge that are available to the historian. Epistemologists and historians classify the sources from which we may obtain justified beliefs about the past in different ways. Epistemologists normally list perception (or observation), introspection, memory, testimony, and reason as types of sources. Three of these sources[6] are available for answering the historical questions mentioned above:

(i) Memory
(ii) Testimony
(iii) Inferences from the present to the past (in which reason and experience work together)[7]

Historians mention:

(A) Memory
(B) The oral, written, and pictorial tradition
(C) Remains, including (C.1) unintended remnants and (C.2) intentional monuments[8]

[6]As we will see below, on closer examination, perception or observation (as supported by proper background information) is to be added.

[7]For example, inferences from properties of preserved testimonies and material remains to properties of objects, actions, and events of the past.

[8]Cf. Droysen (1882), §§20sqq.; Bernheim (1908), pp. 255sq.; for more elaborate classifications, see Feder (1924), pp. 84–105 and Howell and Prevenier (2001), Chap. I.

11.5 Metaphysics of History: Tasks and Projects

If historians and philosophers of history make self-conscious use of ontological categories at all, they mostly talk of *events*.[9] At times, there is also talk of historical *facts*. However, historians and philosophers of history hardly ever dwell on clarifying those categories.[10] In any case, the confinement to or fixation on very few ontological categories is unfortunate and misleading. In a categorial ontology of history, at least the following categories should be taken into account:

- Individuals (Aristotelian substances)
- Persons
- Individual actions and omissions
- Artifacts
- Properties (universals) and individualized properties (so-called tropes)
- Relations, in particular causal relations
- Events
- States of affairs and facts
- Groups/communities
- Collective actions and omissions
- Institutions, organizations, or the like

Among the tasks of a categorial ontology of history are the following: an analysis of the constitution of historical reality, an analysis of each particular category of historical reality, and an analysis of the relations holding between these categories. With regard to fundamental postulates, it is important to emphasize the following:

(OH 1) The ontology of history has to include all categories of physical reality.
(OH 2) The ontology of history has to include all categories of mental reality.

In addition:

(OH 3) The ontology of history has to include categories of cultural reality.
(OH 4) The ontology of history has to include categories of social reality.

[9] According to an influential current, historians are instead concerned with structures. Sometimes, it is absurdly suggested to students of the historical sciences that they have to choose between an investigation of events and an investigation of structures.

[10] To be sure, there are some exceptions: for example, Gruner (1969), Walsh (1969), and Pachter (1974) put some effort in clarifying the concept of an historical event.

Justification of these postulates is straightforward: The physical, mental, cultural, and social realities do not constitute separate worlds or separate strata of being. On the contrary, they constitute a unity. This holds in particular for the possibility of causal interactions.

Whether an ontology of history also has to include categories sui generis, that is, categories of historical reality as such, still requires investigation. (Possible candidates might be the historical situation, epoch, crisis, social movement, an alternative course of the world, etc.).

11.6 Seeing, Hearing, and Feeling the Past: Perception as a Historical Source of Knowledge

According to many philosophers of history, the past is absolutely inaccessible. It is separated from us by an unbridgeable chasm. This chasm is supposed to be not merely temporal but also epistemic and ontological. The prevalent opinion is that the past is gone for good. The only thing we might be able to do about it is to make hazardous conjectures, or we are even bound to construct it all by ourselves in an act of free creation.

Of course, I cannot investigate this misleading picture in every respect here. But I at least want to question it. With the background of our reflections on ontology of history at hand, I invite you to consider the question: How close can the contact with history get? How close can we get to the historical past? The hypothesis I am going to defend is the following:

> (OP) In some cases, it is possible to perceive the past.

And this does not only hold for that part of the past that belongs to space-time regions that cannot be influenced by human beings. (As is well known, it is possible to see stars that do not exist any longer.) We can also observe some parts of history made by humans.

In this context, it is important to remember that not only reports and other oral or written testimonies are preserved but also concrete objects which exhibit many of their original properties: fossils, skeletons, bone fragments, food remains, etc.; pyramids, cathedrals, town walls, etc.; photographs and movies; and sound recordings. While some of the original properties have to be inferred, others can be observed directly.[11] Indeed, a whole spectrum of cases has to be taken into account:

[11] At this point, one must be careful to distinguish between causal and epistemic intermediaries. Of course, in any causal process many intermediate causal links can be distinguished. What I want to dispute is that our access to the past is mediated by epistemic intermediaries (in the form of

(a) Entities whose properties are preserved unchanged – which can therefore be observed and measured
(b) Entities whose properties have changed but which can nevertheless be inferred reliably
(c) Entities whose properties have changed and which can no longer be inferred reliably
(d) Entities that are completely past and gone, in the sense that there are no discernible traces left

The clearest examples of cases of type (a) are material remains. Napoleon is claimed to have said to his soldiers: "Be aware that forty centuries look down upon you." Less metaphorically speaking, it can be said that every single observer of the pyramids sees the past, or put more exactly certain properties of past reality.

To take a more controversial example, when we regard Nadar's photographs of Charles Baudelaire, we gain knowledge of some of the poet's properties. If we listen to audio recordings of Winston Churchill's speeches, we hear the statesman's voice and words.

To be sure, observation of the past is only possible if we possess the requisite concepts and background beliefs, and in order to have them, a lot has to be learned. This, however, holds for perception in general. Thus, it remains true that we directly perceive some aspects of the past instead of having to infer them in a rather roundabout way.

Certainly, only a small part of the past is accessible to us in this way. Nevertheless, being able to perceive past reality is very important for the phenomenology of historical experience and its institutionalization in museums, memorials, and other places of remembrance. Moreover, this access to the past provides a point of departure for the rejection of radically skeptical and anti-realist views of history and historical science.

Acknowledgments This chapter is part of the project "*Explanations, Causality and Laws in Historical Science*" of the research group "*Causation, Laws, Dispositions and Explanation at the Intersection of Science and Metaphysics*," funded by the Deutsche Forschungsgemeinschaft (DFG). For helpful comments, I want to thank Eva-Maria Jung, Benedikt Kahmen, Martin Kusch, Daniel Plenge, Peter Rohs, Ansgar Seide, Markus Seidel, and Aviezer Tucker.

inferred beliefs) in every case. Some aspects of past reality can be perceived without such epistemic intermediaries, and in this sense directly.

References

Bernheim, E. (1908). *Lehrbuch der Historischen Methode und der Geschichtsphilosophie. Fünfte und sechste, neu bearbeitete und vermehrte Auflage*. Leipzig: Duncker & Humblot.

Boghossian, P. A. (2006). *Fear of knowledge. Against relativism and constructivism*. Oxford: Clarendon Press.

Droysen, J. G. (1882). *Grundriss der Historik*. Leipzig: Veit.

Feder, A. (1924). *Lehrbuch der geschichtlichen Methode. Dritte, umgearbeitete und verbesserte Auflage*. Regensburg: Kösel & Pustet.

Gruner, R. (1969). The notion of an historical event. *Proceedings of the Aristotelian Society, Supplementary Volume, 43*, 141–152.

Howell, M., & Prevenier, W. (2001). *From reliable sources. An introduction to historical method*. Ithaca and London: Cornell University Press.

Kosso, P. (2001). *Knowing the past*. Amherst: Humanity Books.

Kosso, P. (2009). Philosophy of historiography. In A. Tucker (Ed.), *A companion to the philosophy of history and historiography* (pp. 9–25). Malden: Blackwell.

Loux, M. J. (2006). *Metaphysics: A contemporary introduction* (3rd ed.). New York: Routledge.

Pachter, H. M. (1974). Defining an event: Prolegomenon to any future philosophy of history. *Social Research, 41*, 439–466.

Schaffer, J. (2009). On what grounds what. In D. J. Chalmers, D. Manley, & R. Wasserman (Eds.), *Metametaphysics: New essays on the foundations of ontology* (pp. 347–383). Oxford: Clarendon Press.

Tucker, A. (2004). *Our knowledge of the past. A philosophy of historiography*. Cambridge: Cambridge University Press.

Walsh, W. H. (1969). The notion of an historical event. *Proceedings of the Aristotelian Society, Supplementary Volume, 43*, 153–166.

Chapter 12
Causal Explanations of Historical Trends

Derek D. Turner

Abstract Philosophers interested in historical explanation have tended to focus on causal explanations of particular events, states of affairs, or observable traces. Yet researchers working in fields ranging from paleontology and evolutionary biology to climate science and economics seek to document and explain the occurrence of historical, population-level trends, where a trend is a persistent, directional change in some state variable. Examples of such trends include everything from evolutionary size increase (Cope's rule), to changes of gene frequencies in an evolving population, to global warming. This chapter explores the idea that explanations of historical trends are typically causal explanations. Woodward's interventionist theory treats causation as a relation between variables and so lends itself readily to the idea that trends can be causes and effects. A small extension of Woodward's theory can help illuminate cases in which scientists talk of one trend as being the cause of another. However, paleontologists often explain trends by claiming that they are passive, that is, that they involve a random walk away from a fixed boundary in the state space. This type of explanation of an historical trend does not invoke any causes in the interventionist sense, which suggests that some explanations of historical trends might not be causal explanations.

Keywords Causation • Historical trends • Interventionism • Paleontology

D.D. Turner (✉)
Department of Philosophy, Connecticut College, 270 Mohegan Avenue,
New London, CT 06320, USA
e-mail: Derek.turner@conncoll.edu

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_12,

12.1 Introduction

Researchers in a variety of different fields – from paleontology to climate science – investigate historical trends. A trend is any persistent, directional change in some interesting variable or measure. The first challenge of such research is to document a trend, or to establish that it is real. The next step is then to try to explain it. In practice, scientists often treat historical trends as standing in causal relationships, and they often try to develop and test causal explanations of those trends. In what follows, I begin (in Sect. 12.2) by describing one example of this kind of scientific research. The example, drawn from vertebrate paleontology, is one in which scientists seem to treat one morphological trend as the cause of another. Section 12.3 sets the stage for further discussion by offering a brief exploration of the metaphysics of historical trends. Section 12.4 develops an initially promising interventionist proposal for making sense of causal claims about historical trends, based largely on the work of Woodward (2003). However, in Sect. 12.5, I turn to examine a problem case. Paleontologists distinguish between active and passive evolutionary trends (McShea 1994), and passive trends pose a special challenge for the interventionist approach.

In what follows, I focus mainly on examples from paleontology. Paleontology makes for an especially rich supply of case studies, in part because paleontologists have focused so explicitly in recent decades on questions about the relationship between patterns and trends, on the one hand, and on underlying processes, on the other (see Kemp 1999, Sepkoski 2012, and Turner 2011 for introductions to this work.) However, it is worth noting that researchers working in many different fields study historical trends, and the three basic questions they ask are usually the same:

1. Which trends are the "real" ones?
2. What are the causes of those trends?
3. What are their effects?

Consider the following list of putative historical trends:

- Macroevolutionary body size increase (Cope's rule)
- Directional changes in gene (or trait) frequencies in a population
- Grade inflation
- Falling housing prices
- Rising unemployment
- Increasing economic inequality
- Global warming
- Rising sea levels
- Increasing atmospheric CO_2, in ppm
- Human population growth

- Declining fertility rates
- Biodiversity loss
- Increasing scientific knowledge
- Etc.

This list makes it clear, I hope, that historical trends loom large in many different fields and in many aspects of life. The study of historical trends is one distinctively historical variety of science and one that gets left out of some recent characterizations of historical science (e.g., Cleland 2002, 2011). Researchers in both the natural and social sciences often work on historical trends, and their work is sometimes highly relevant to policy. What's more, normative judgments about progress are typically anchored to empirical claims about directional historical trends. As Elliott Sober (1994) once put it: "progress = directional change + values." In some cases (e.g., with respect to global climate change), claims about the causes and effects of historical trends are politically controversial.

I mention these other cases only to underscore the importance of the scientific study of historical trends and to suggest that there are some interesting methodological commonalities between paleontology and other seemingly unrelated areas of empirical research, from economics to climate science. This raises the stakes somewhat with respect to the guiding question of this chapter – namely, how should we understand the meaning of causal claims about trends?

The spirit of this chapter is largely exploratory. The plan is to begin with Woodward's interventionism and see whether it can help us to think through some examples from paleontology. The results of this exploration are mixed (with some success, but also some problems), but they are also preliminary. The question how to understand causal claims about trends is one that needs further work.

12.2 Morphological Trends in the Fossil Record

Paleontologists often begin their work by seeking to document trends and patterns in the fossil record. For example, O'Keefe and Carrano (2005) identified two trends in the morphology of plesiosaurs, one group of marine reptiles that lived from about 220 until 65 million years ago. The scientists looked at 41 plesiosaur specimens representing a total of 28 distinct taxa. Working with a phylogeny for plesiosaurs that O'Keefe (2001) had developed a few years earlier, they focused on two interesting measures: body size and the ratio of head to neck length (hereafter, the HN ratio). They documented two noteworthy trends in plesiosaur morphology. The first of these is a trend toward larger overall body size, consistent with the idea that paleontologists have come to know as "Cope's rule," after the nineteenth-century American paleontologist E. D. Cope.

The reason for focusing on the HN ratio is that it may serve as a good proxy variable for trophic specialization. Scientists have long known that plesiosaurs evolved two different body types. The first of these (which was for a long time

classified as the *Pliosauridae*) had relatively long heads and short necks. The second (traditionally classified as the *Plesiosauridae*) had long, snakelike necks and relatively short heads. The HN ratio presumably tells us something about the size and type of prey that the animals could have eaten, and it is plausible to think that the two body types represent adaptations to different kinds of food sources. When O'Keefe and Carrano looked for trends in HN ratio in the plesiosaurs, they found what they characterized as a trend toward "increasing trophic specialization." Over time, the HN ratios tended to get more extreme.

Rather than investigating the causes of size increase and trophic specialization in plesiosaurs, O'Keefe and Carrano focus more on the relationship between these trends and other quantifiable morphological trends. They looked at six other features that have to do with the plesiosaur locomotor system:

- Scapula length
- Coracoid length
- Ischium length
- Pubis length
- Humerus length
- Femur length

The scapula and coracoid bones are associated with the shoulder joint, whereas the pubis and ischium bones are associated with the hip joint. Studying the relationships among these variables can tell us something about how the plesiosaurs might have used their four flippers to propel themselves through the water. Any changes in the relationships among these variables would reflect changes in the biomechanics of plesiosaur swimming. For example, O'Keefe and Carrano note that over time, the girdle components (including both the shoulder girdle, consisting of the scapula and coracoid bones, and the pelvic girdle, consisting of the pubis and ischium) tended to get relatively longer, while the bones in the limbs tended to get relatively shorter. They interpret this change as an allometric consequence of the evolution of larger body size. Setting aside the biomechanical details (see their 2005, 666ff.), the rough idea is that as the body size increases, the propulsion system must also change in order to generate sufficient force to move the larger animal through the water. So there is a morphological trend in the plesiosaurs – a change in the ratio of girdle length (i.e., the length of the pelvic and shoulder girdles) to propoidal length (i.e., the length of the femur and humerus, respectively) – and that trend is explained by reference to body size increase plus biomechanical constraints. In other words, O'Keefe and Carrano are *using one morphological trend to explain another*.

Figure 12.1 provides a rough illustration of the causal story that one can glean from O'Keefe and Carrano's (2005) paper.

I want to be cautious here in attributing causal claims to O'Keefe and Carrano because they themselves seem to shy away from talk of causation, preferring instead to say that trends are "correlated." Their work does at least suggest, however, the following two causal claims (represented in Fig. 12.1):

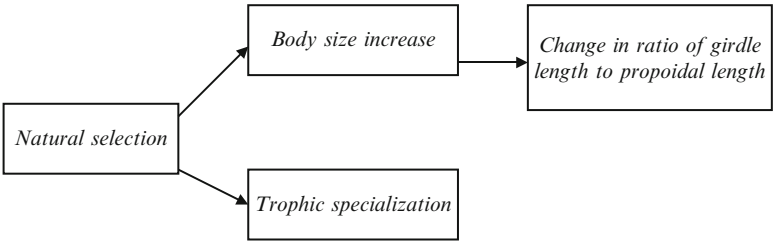


Fig. 12.1 Causal claims implicit in the work of O'Keefe and Carrano (2005)

C1 Natural selection is the cause of certain trends in plesiosaur morphology, especially body size increase and trophic specialization.
C2 Body size increase in the plesiosaurs caused other changes in the morphology of the locomotor system, in particular, changes in the ratio of girdle length to propoidal length.

Notice that C1 seems to imply that historical trends can stand in causal relationships with *historical processes*, such as natural selection, whereas C2 seems to imply that trends can stand in causal relationships with *other trends*.

Consider first claim C1. Although O'Keefe and Carrano do not explicitly claim that natural selection is what drove either body size increase or trophic specialization in the plesiosaurs, they do claim that body size increase was an "active" trend. In support of this claim, they invoke an empirical test first proposed by McShea (1994) and known as the stable minimum test. McShea argued that the stable minimum test can be used to determine whether an evolutionary trend is passive or driven (see also Sect. 12.4 of this chapter for further discussion of passive trends). To a first approximation, we can say that a trend is *driven* when there is a directional bias in the state space – for example, when body size increases are more probable than decreases. When a trend is *passive*, the clade does a "random walk" away from a fixed boundary in the state space. (For a lovely illustration of this concept, see Gould's (1996) description of the "drunkard's walk.") When a trend is passive, the size of the smallest members of the clade should remain constant. But if the size of the smallest members of the clade increases over time, the increasing minimum suggests that something is driving the evolution of the clade toward larger size. Traditionally, most scientists have just assumed that the "driver" would be natural selection (see, e.g., Hone and Benton 2005). Although O'Keefe and Carrano use slightly different terminology – they talk of "active" rather than "driven" trends – many, though not all, scientists use those terms interchangeably (Wang 2001 is one exception). Applying McShea's stable minimum test to the plesiosaurs, O'Keefe and Carrano conclude that body size increase in the plesiosaurs was an active trend. The background assumption here is that the cause of an active trend would be natural selection. Natural selection is also a pretty obvious candidate for being the cause of

trophic specialization in plesiosaurs. Although this idea would still need to be tested, the morphological trend in HN ratios plausibly results from the adaptation of different groups of plesiosaurs to different food sources and to different ecological roles.

Although the mainstream view is that natural selection is a cause of evolutionary change, there is an interesting minority view according to which natural selection is itself just a special sort of historical trend – a biased directional trend in trait frequencies (or, perhaps better, in gene frequencies) in a population (Matthen and Ariew 2002; Walsh et al. 2002). Without weighing in on the ongoing controversy concerning the status of natural selection, it is at least worth pointing out that if the statisticalist view of selection is correct, then it would follow that whenever scientists explain macroevolutionary trends in terms of natural selection, they are essentially explaining trends in terms of other trends.

What about claim C2? The conclusion of O'Keefe and Carrano's paper is telling:

> Two broad trends are demonstrable in the evolution of the Plesiosauria – an active trend of body size increase and a trend toward divergent trophic specialization. Measures of the locomotor system show a complex set of correlations with these trends. Concerning body size, identical changes in the geometry of the locomotor system are evident in all plesiosaur subclades. This suggests that the physical constraints of thrust production placed demands on the locomotor system that *resulted* in allometric changes, specifically the relative shortening of propoidals and lengthening of girdle elements. (2005, p. 672, emphasis added)

They seem to be making a causal claim here: Body size increase, together with certain facts about the biomechanics of swimming, provides a causal explanation of the trend in locomotor morphology. Tellingly, they refrain from arguing that trophic specialization caused any trends in locomotor morphology:

> The trend toward trophic specialization is also correlated with stereotyped geometries in the locomotor system. These patterns are statistically significant ... but we lack the data to constrain speculation about why the observed correlations occur. (2005, p. 672)

Their cautious restraint here seems to suggest that they do take themselves to have the data to support the causal claim expressed in C2.

12.3 The Metaphysics of Historical Trends: Some Preliminary Distinctions

In order to get clear about what it means to say that trends can stand in causal relations, we first need get clear about what trends *are*. Here I offer some initial observations about the metaphysics of trends. The metaphysics of trends remains a relatively under-explored topic, and these observations barely scratch the surface. But they will be helpful in what follows.

First, recall that I defined a historical trend as a persistent directional change in some variable. Thus defined, a trend is one kind of pattern. All trends are patterns, but not all patterns are trends. A good example of a pattern that should not be considered a trend is the 32 million-year periodicity of mass extinction events that

paleontologists David Raup and Jack Sepkoski thought they had discovered (Raup and Sepkoski 1984). That was an alleged pattern in the fossil record, but the pattern was cyclical and did not involve any persistent directional change.

Second, it might be useful to distinguish at the outset between *population-level trends* and *trends in the properties of individuals*. For example, as a child grows up, his/her height increases in a directional fashion. Height, in this case, is a property of one individual. By contrast, an increase in the average height of American 5th graders over a given interval of time would be an example of a population-level trend. Here I will focus exclusively on population-level trends, such as body size increase and trophic specialization in plesiosaurs.

Third, the reality of a trend is relative to one's decision about which spatial and temporal scale to focus on. To give a simple example, over a given time interval, unemployment might fall in one particular region, even while it increases at the national scale. Although it is a bit awkward, we can say that "rising unemployment" is a real trend at the national scale but not at the local scale. Similarly, if we fix attention on a particular region, then rising unemployment might be a real trend over some relatively short time interval – say, 6 months. But if we zoom out and look at a longer 5-year interval, rising unemployment may no longer show up as a real trend. This fact about scaling effects is crucial for understanding debates within paleontology concerning macroevolutionary trends, such as Cope's rule of size increase. Cope's rule does show up as a real trend at some spatiotemporal scales, and in some clades, but not in others (McShea 1998; Turner 2011, p. 110).

Fourth, some trends are constitutive parts of others. For example, if a professor begins his/her career as a tough grader but eases up over time, the trend in average grades awarded by that person might be part of a larger grade inflationary trend on campus. When thinking about trends, it will be helpful to attend to the distinction between *causation* and *constitution*. The professor may be a contributing cause of grade inflation (let us save that issue for later), but it would not be quite correct to say that the upward trend in his/her grading is a cause of grade inflation. Rather, grade inflation on campus is constituted, in part, by the trend in this one faculty member's grading practices.

The approach I take, beginning in the next section, is to start with Woodward's (2003) articulation of an interventionist theory of causation, and I attempt to see whether that theory can shed some light on O'Keefe and Carrano's work on plesiosaur evolution. One clear limitation of this approach is that it ignores all the other going philosophical theories of causation. One can envision a more grandiose philosophical project that would involve surveying those theories in order to see how well each one can handle causal claims about trends, but that is much more than I can take on here.

It may also be helpful to bypass some deeper questions about the metaphysics of trends. For example, are trends and patterns abstract objects? (See Dennett 1991 for a fascinating discussion.) If so, what is at stake in debates about whether trends such as Cope's rule, or even global warming, are real? Can abstract objects stand in causal relations? Using Woodward's work to ground the subsequent discussion will happily make it possible to set some of these very difficult questions to one side.

12.4 An Interventionist Proposal

Woodward's (2003) interventionist theory of causation might seem to offer a solution to this puzzle about how trends can stand in causal relations. One reason for optimism is that Woodward treats causation as a relation that obtains between different variables, not as a relation between concrete objects or events:

> It is most perspicuous to think of causal relationships as relating variables or, to speak more precisely, as describing how changes in the value of one or more variables will change the value of other variables. This is also the way that many scientists think about causal relationships. (2003, p. 38)

Since a trend is merely a persisting directional change in the value of some variable, Woodward's approach seems very amenable to the idea that trends can be causes and effects. Woodward's thought, to a first approximation, is that "X causes Y" means that if you could perform an experiment in which you manipulate the value of X while holding fixed the values of all the other relevant variables, then the value of Y would change accordingly. And we do seem to be able to manipulate variables in a persistent, directional way. Think, for example, of our manipulation of atmospheric CO_2 concentrations.

Woodward himself does not say anything about historical trends. When he discusses the hypothetical manipulation of variables, he does not consider cases that would involve persistent directional changes in the values of those variables. Nevertheless, his interventionist approach seems able to accommodate such cases.

One possible hitch is that Woodward seems to focus exclusively on individual-level properties. Thus, he writes that:

> Values of variables are always possessed by or instantiated in particular individuals or units, as when a particular table has a mass of 10 kg. (2003, p. 39)

This seems to exclude cases involving directional changes in population-level properties. For example, the average body size of plesiosaurs is a variable whose values are not instantiated in any particular individual or unit. Even with this restriction, Woodward's approach might be able to explain how trends in individual properties (see Sect. 12.2 above) could stand in causal relations. It is not entirely clear, however, that this restriction is essential to his view. There seems to be no principled reason why we could not intervene on variables associated with aggregate or population-level properties. What is needed is some additional principle to cover cases in which the values of variables change in a persistent, directional way.

Explaining how trends in population-level properties can stand in causal relations poses a bit more of a challenge. The distinction between constitution and causation (introduced in Sect. 12.2) might seem to cause some trouble here. Consider Dennett's (1991) example of the mean geographical center of population of the USA. The mean geographical center of population is an example of a population-level property. The US Census Bureau defines the mean center of population in the following way:

> The mean center of population is the point at which an imaginary, flat, weightless, and rigid map of the United States would balance if weights of identical value were placed on it so that each weight represented the location of one person.[1]

Now suppose we perform an experiment in which one individual relocates from Connecticut to California, while somehow holding fixed the locations of everyone else. As a result, the mean center of population will shift a bit to the west. In this example, we are wiggling the value of an individual-level variable – that is, the location of a single citizen – while holding other relevant variables fixed, and the value of the population-level variable wiggles along. Thus, Woodward's view would seem to imply that individuals can *cause* changes in the mean geographical center of population by moving across the country. At least, it would imply that if (contrary to Woodward's restriction mentioned above) we took that theory to apply to population-level variables.

One potential problem with this example of an intervention on the mean geographical center of population is that it seems to conflate *causation* with *constitution*. The mean center of population is just an aggregate measure of the addresses of all individual US citizens. Those individual addresses are best described as constituting the mean center of population, rather than causing it. In general, it sounds odd to say that an average (or some other aggregate measure) is an effect of the properties of the individuals in a population. Perhaps one reason for the oddness is that we usually think of causes as temporally preceding their effects, whereas the constitution relation is not a temporal one. The change in the mean geographical center of population is simultaneous with the individual's relocation. At any rate, if we take this distinction between constitution and causation seriously, we seem driven toward the view that changes in the values of population-level variables are not, in general, caused by changes in the values of individual-level variables. Instead, they are constituted by changes at the individual level. This may help explain why Woodward chooses to restrict his focus to individual-level variables. As long as we focus exclusively on individual-level variables, the distinction between constitution and causation does not become an issue.

In the example from Sect. 12.1, the main challenge was to try to understand how a trend in one population-level variable could cause a trend in another population-level variable. In spite of the aforementioned worry about the constitution/causation distinction, Woodward's interventionist account might seem to make sense of causal relations between different population-level variables, in much the same way as it makes sense of causal relations between different individual-level variables. Perhaps what we need is a restriction saying that causal relations can only obtain between variables at the same level; population-level properties are generally constituted, not caused by, individual-level properties. This, however, leaves entirely open the possibility that population-level variables might stand in causal relations with one another.

---

[1] From the website of the US Census Bureau, last accessed on April 10, 2012 (http://www.census.gov/population/www/censusdata/files/popctr.pdf).

Reisman and Forber (2005) take an interventionist line in response to the statisticalist views of Walsh, Ariew, Lewens, and Matthen. Reisman and Forber point out that in certain experimental settings, scientists can manipulate drift and selection and check to see how those manipulations make a difference to the subsequent evolutionary patterns. For example, experimentalists can and do control the "strength" of drift by manipulating the starting size of a population (of fruit flies, bacteria, or whatever). This line of argument may suggest that population size may cause population-level trends, in an interventionist sense of "cause." What this argument does not do, however, is help make sense of the idea that one trend can be a cause of another. The sort of causal claim exemplified by C2 remains in need of clarification.

The following interventionist proposal seems promising:

> *Trend-Trend Causation*: "A trend in the value of *X* causes a trend in the value of *Y* (where the values of *X* and *Y* are both population-level properties, or both individual-level properties)" *means that* "If you could manipulate the value of *X* in a persistent, directional way, while holding fixed all the other variables that might make a difference to *Y*, then the value of *Y* would also move in a correspondingly persistent, directional way."

This proposal slightly modifies Woodward's original theory in order to accommodate both individual property and population-level historical trends. It certainly seems consistent with the spirit of the original theory.

The above proposal makes good sense of the initial case that I described in Sect. 12.1. In that opening example, O'Keefe and Carrano seemed to be saying that the trend in plesiosaur body size was causing certain other morphological trends relevant to plesiosaur locomotion. If body size increases (in a persistent, directional fashion) while certain biomechanical constraints on swimming remain fixed, then something else has to give. In this case, there was a resulting trend in the ratio between the lengths of the bones in the shoulder and pelvic girdles, on the one hand, and the lengths of the propoidal bones, on the other. The interventionist proposal above clarifies what we mean by this talk of the trend in body size causing the trend in the morphology of the locomotor system. We are, in essence, talking about a hypothetical experimental intervention. If we could somehow manipulate the body size of a plesiosaur while holding fixed all the biomechanical constraints on swimming, then in order for the animal to remain viable at all, certain other aspects of its locomotor system would have to change: The bones in the girdle would get relatively longer, while the propoidal bones would get relatively shorter. It is tempting to say, with a bit of personification, that natural selection did in fact carry out this experiment during the Mesozoic.

Thus, the interventionist proposal seems to make good sense of the idea that population-level trends can stand in causal relations with other population-level trends, and thus illuminates the type of scientific practice described in Sect. 12.1. We now seem to have a clearer idea of what is going on when scientists causally

explain one historical trend in terms of another. Indeed, the fact that interventionism can make sense of cases that its defenders have not yet considered would seem to count strongly in favor of an interventionist approach to causation. In the next section, however, I argue that there are some serious problems with the appealing interventionist proposal.

12.5 Passive Trends

The analysis of trend-trend causation developed in Sect. 12.4 represents a plausible extension of Woodward's interventionist theory and one that sheds light on a least some of the causal claims that figure in the historical sciences. For example, the claim that increasing atmospheric CO_2 concentration is a cause of global warming is an example of a claim about trend-trend causation. However, the proposal developed above only goes so far. I will now consider a rather different case, also drawn from paleontology, that is difficult to understand in interventionist terms.

Suppose that we are tracking the mean body size of some clade. And suppose that increases and decreases in body size are equally probable. With each time interval, it is as if a fair coin is flipped to determine whether size increases or decreases. Next, suppose that there is a fixed boundary in the state space, or a fixed minimum size for the clade. That fixed boundary could result from natural selection working against small-bodied organisms, but it could also be a result of biomechanical constraints. If the clade starts out at or near this fixed lower boundary, then the mere process of diversification will lead to an increase in the mean body size, even if there is no directional bias in favor of larger size, and even if natural selection "doesn't care" about body size. The clade will do a random walk away from the fixed boundary. This idea was first proposed as an explanation of Cope's rule by Stanley (1973) and has been invoked by other scientists since then (e.g., Gould 1988, 1997).

Passive trends do involve persistent directional change in the mean, but that change is (by definition) not caused or driven by a persistent directional change in some other variable. If a passive trend were caused by some other trend, in somewhat the same way that the morphological changes in plesiosaur limbs, hips, and shoulder joints were caused by body size increase, then it would not be passive. So the proposed extension of interventionism that I explored in Sect. 12.4 (trend-trend causation) will not help here. Is there some other way of thinking about the causes of passive trends in interventionist terms? Or would it be best to say that passive trends have no (interventionist) causes at all?

We can approach these issues in a systematic way by considering all the possible variables that one could intervene on in ways that might make a difference to the character of a passive trend. For illustrative purposes, it will help to stick with the example of body size evolution. In principle, one could manipulate (i) the starting body size of the clade, (ii) the fixed boundary in the state space, or the minimum body size for the clade, or (iii) the strength of the directional bias in the state space. I will consider each of these three possibilities in turn.

12.5.1 Manipulating the Starting Value of the Target Variable

One initially plausible suggestion is that the starting mean body size of the clade is a cause of the passive trend. If we could explain why mammals started out so small, we would thereby have (partly) explained the subsequent size increase. Moreover, the starting body size of the clade is certainly a variable on which we can perform hypothetical interventions. It is easy to imagine a counterfactual scenario in which the first mammals are the size of mastodons. Nevertheless, what we are looking for is some *other* variable that would stand in a causal relationship to mean body size. When we imagine interventions on the starting mean body size of the clade, all we are doing is imagining how later values of that variable might depend on the starting value of the same variable.

It is an interesting question what an interventionist should say about cases where the latter values of a variable depend on the earlier values. For example, if you heat up a pot of water and let it sit at room temperature for 5 min, the temperature of the pot at the end of that interval will depend on the starting temperature. Should we say that the starting temperature is a cause (in the interventionist sense) of the latter temperature? Interventions on the starting temperature will certainly make a difference to the temperature at the end of the 5-min interval. One problem here is that the starting temperature and the ending temperature are not different variables; they are just different values of the same variable. For that reason, the starting temperature would not count as a cause on Woodward's view.

Recall the definition of "trend" as any persistent, directional change in some variable of interest. The initial value of that variable will always be something that one can (in principle, at least) manipulate. But what we want to explain, in the case of a trend, is the persistent, directional change.

12.5.2 Manipulating the Fixed Boundary in the State Space

One can also imagine hypothetical interventions on the fixed boundary in the state space. Imagine, for instance, that the minimum body size for mammals increases over time, so that at some later time, the smallest possible mammal is the size of a large dog. This could happen if minimum body size were determined by some environmental factors (say, the strength of the earth's gravitational field, or the density of the atmosphere) that change over time. Most scientists do in fact think that the oxygen content of the atmosphere imposes a size maximum on certain kinds of organisms, especially insects, and that this maximum has changed in the past. The shifting lower boundary could well drive a change in the mean body size of the clade. In this case, however, the trend would no longer fit the canonical definition of "passive." It would be an example of what I have elsewhere called a "shifting boundary" trend (Turner 2009). In the literature, passive trends are typically associated with fixed boundaries in the state space.

According to the interventionist picture, two variables stand in a causal relationship to one another when changes in the value of one are sensitive to changes in the value of another (other things being held fixed). The problem is that in the case of a passive trend, mean body size increases, while the size minimum remains fixed in the state space. We could drive up the mean body size by moving the minimum size upwards – that would be a shifting boundary trend – but what about a case where the mean body size trends upward with no change at all in the fixed boundary? In such a case, the fixed boundary would not seem to count as a cause of the upward trend. The fact that shifting the boundary could be an (interventionist) cause of a trend in body size does not mean that the fixed boundary is a cause where the trend is passive. In the case of a passive trend, the boundary in the state space is one of the "background" variables that are being held fixed.

12.5.3 Manipulating the Directional Bias in the State Space

One important difference between passive and driven trends is that the latter are generated by a directional bias in the state space. For example, if increases in body size were more probable than decreases, the resulting trend toward larger body size would be driven. The strength of the directional bias is also a variable that one could hypothetically manipulate. (See, e.g., Turner 2009 for some discussion of "shifting bias" trends.) Intuitively, where a trend is driven, we might want to say that the directional bias in the state space is the cause. It seems natural, in other words, to say that body size increases because there is a directional bias toward larger size. This is akin to saying (in the context of coin tossing) that we have obtained a long sequence of heads because the coin is weighted toward heads, that the bias toward heads is the cause of the pattern. If in the case of driven trends, we are inclined to say that the directional bias causes the trend, then why not say that the *absence* of any bias is what causes a passive trend? We can cause a passive trend by setting things up so that the probabilities of size increase and decrease are equal.

One problem with the above suggestion is that it is impossible to generate a passive trend merely by manipulating the strength of the directional bias in the state space. A passive trend also requires the fixed boundary in the state space, and it requires that the system start out at or near that fixed boundary. Suppose, for example, that we start out with a clade (a set of evolving lineages) whose mean size is well above the minimum boundary. If we suppose that size increases and decreases are equally probable, there is no reason to expect any trend at all toward larger size. Of course, such a trend is always possible. But saying that a trend toward larger size is caused by the absence of a directional bias would be akin to saying (in a coin tossing context) that a long sequence of heads is caused by the fact that the coin is fair.

Recall that interventionism treats causation as a relation between variables. In the case of passive trends, we have seen that there are three variables that one could, in principle, intervene on: (i) the starting value of the variable of interest (e.g., the

starting mean size of the evolving clade), (ii) the fixed boundary in the state space, and (iii) the strength of the directional bias in the state space. Passive trends require that all three of these factors be set up in just the right way. Grantham (1999) aptly refers to these factors as "structuring causes," and there may well be some loose sense in which this setup is the "cause" of a passive trend. The interventionist approach has some difficulty with this case, however. Taking each of these three variables in isolation, it is difficult to make any sense of the idea that any one of them is an interventionist cause of the passive trend. Obviously, by intervening on any one of these variables – say, by shifting the boundary, introducing a bias, or moving the starting value away from the boundary – we can make a difference to the resulting pattern or trend. However, the crucial thing to see is that in a passive diffusion model, all of these variables are *held fixed*: The size minimum does not change, the initial mean size of the clade does not change as the system evolves, and the directional bias remains set to zero. Yet mean body size trends upward. Because these variables all remain fixed while mean body size increases, we cannot point to changes in these variables as causes of the change in mean body size.

12.6 Conclusion

I began with a straightforward observation about the practices of the historical natural sciences. Historical scientists often investigate the causes and effects of trends. Indeed, they often develop and test causal explanations of historical trends. One task for philosophers of science is to see whether our best philosophical theories – in this case, theories of causation – can make this bit of scientific practice intelligible. Here I have argued that with a small modification, Woodward's interventionist theory can indeed make sense of many causal claims about trends. This interventionist approach does a good job explaining what it might mean to say that one trend causes another; thus, it does a good job illuminating the paleontological case study with which I began. It does, however, run into problems in the case of passive trends. There the trends in question seem not to have any (interventionist) causes. Insofar as scientists do explain a trend by claiming that it is passive, they are not offering a causal explanation of it, at least not in the interventionist sense explored here.

Acknowledgements Earlier versions of this chapter were presented at the Society for Philosophy of Science in Practice (SPSP) and at Connecticut College in 2009. I am grateful to members of both audiences for helpful feedback.

References

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science, 69*(3), 474–496.
Cleland, C. E. (2011). Prediction and explanation in historical natural science. *British Journal for Philosophy of Science, 62*(3), 1–32.
Dennett, D. C. (1991). Real patterns. *The Journal of Philosophy, 88*(1), 27–51.
Gould, S. J. (1988). Trends as changes in variance: A new slant on progress and directionality in evolution. *Journal of Paleontology, 62*(2), 319–329.
Gould, S. J. (1996). *Full house: The spread of excellence from Plato to Darwin*. New York: WW Norton.
Gould, S. J. (1997). Cope's rule as psychological artifact. *Nature, 385*(6613), 199–200.
Grantham, T. A. (1999). Explanatory pluralism in paleobiology. *Philosophy of Science, 66*(3), 223–236.
Hone, D. W. E., & Benton, M. J. (2005). The evolution of large size: How does Cope's Rule work? *Trends in Ecology and Evolution, 20*(1), 4–6.
Kemp, T. S. (1999). *Fossils and evolution*. Oxford: Oxford University Press.
Matthen, M., & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *The Journal of Philosophy, 99*(2), 55–83.
McShea, D. W. (1994). Mechanisms of large-scale evolutionary trends. *Evolution, 48*(6), 1747–1763.
McShea, D. W. (1998). Possible largest-scale trends in organismal evolution: Eight 'live hypotheses'. *Annual Review of Ecology and Systematics, 29*, 293–318.
O'Keefe, F. R. (2001). A cladistic analysis and taxonomic revision of the plesiosauria. *Acta Zoologica Fennica, 213*, 1–63.
O'Keefe, F. R., & Carrano, M. T. (2005). Correlated trends in the evolution of the plesiosaur locomotor system. *Paleobiology, 31*(4), 656–675.
Raup, D. M., & Sepkoski, J. J. (1984). Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences, 81*(3), 801–805.
Reisman, K., & Forber, P. (2005). Manipulation and the causes of evolution. *Philosophy of Science, 72*(5), 1113–1123.
Sepkoski, D. (2012). *Rereading the fossil record*. Chicago: University of Chicago Press.
Sober, E. (1994). Progress and direction in evolution. In J. H. Campbell & J. W. Schopf (Eds.), *Creative evolution!?* Boston: Jones and Bartlett Publishers.
Stanley, S. M. (1973). An explanation for Cope's rule. *Evolution, 27*(1), 1–26.
Turner, D. (2009). How much can we know about the causes of evolutionary trends? *Biology and Philosophy, 24*, 341–357.
Turner, D. (2011). *Paleontology: A philosophical introduction*. Cambridge: Cambridge University Press.
Walsh, D. T., Lewens, T., & Ariew, A. (2002). The trials of life: Natural selection and random drift. *Philosophy of Science, 69*(3), 452–473.
Wang, S. C. (2001). Quantifying passive and driven large-scale evolutionary trends. *Evolution, 55*(5), 849–858.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Part IV
Bridging the Two Disciplines

Chapter 13
Aspects of Human Historiographic Explanation: A View from the Philosophy of Science

Stuart Glennan

Abstract While some philosophers of history have argued that explanations in human history are of a fundamentally different kind than explanations in the natural sciences, I shall argue that this is not the case. Human beings are part of nature, human history is part of natural history, and human historical explanation is a species of natural historical explanation. In this chapter, I shall use a case study from the history of the American Civil War to show the variety of close parallels between natural and human historical explanation. In both instances, I shall argue that these explanations involve narrative descriptions of causal mechanisms. I shall show how adopting a mechanistic approach to explanation can provide resources to address some important aspects of human historiographic explanation, including problems concerning event individuation, historical meaning, agency, the role of laws, and the nature of contingency.

Keywords Historical explanation • Mechanism • Laws • Agency • Naturalism

13.1 Introduction

While some philosophers have suggested that explanations of events in human history are of a fundamentally different kind than explanations of natural events, I shall argue that this is not the case. Human beings are part of nature, human history is part of natural history, and human historical explanation is a species of natural historical explanation. This view is sometimes called naturalism, and its converse is called anti-naturalism or exceptionalism.

S. Glennan (✉)
Butler University, Department of Philosophy and Religion, 4600 Sunset Ave, Indianapolis, IN 46208, USA
e-mail: sglennan@butler.edu

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_13,

My case for naturalism starts from a rather general view about the science and nature. Briefly put, that view is that natural phenomena, including human and social phenomena, are all (or at least nearly all) produced by the operation of structures called mechanisms, and that scientists explain phenomena by describing the mechanisms that are responsible for these phenomena. This "new mechanicism" or "new mechanical philosophy" has received considerable discussion in the last decade (Glennan 1996; Machamer et al. 2000; Bechtel and Abrahamsen 2005). For purposes of this chapter, the differences among the new mechanists are inessential. My main supposition is widely shared and relatively noncontroversial, namely, that (most) scientific explanations are causal and mechanistic rather than based upon laws of nature. In a recent paper (Glennan 2010b) I have considered the implications of this view for the explanation of historical events, claiming that these events, both natural and human, are typically the products of *ephemeral mechanisms* – mechanisms whose organization is fleeting and one-off in character – and that what historians call narrative explanations are in fact descriptions of these mechanisms.

This chapter will extend the case for naturalism by considering how the mechanistic approach can be applied to the explanation of a particular event in human history, that is, the battle of Antietam, which occurred during the American Civil War. The exploration of this case will show how the mechanistic approach to explanation works in historical cases and demonstrate that various supposedly exceptional features of human historiographic explanation have close analogs in the explanation of nonhuman phenomena.

Because of the many uses of the term "history," it will be helpful to clarify at the outset some terms connected with history, historiography, and explanation. In the first place, it is essential to distinguish history, the actual events and processes that have occurred in the past, from historiography, which is the activity of discovering, describing, and explaining those events. The relation between historiography and history then is analogous to the relation between science and nature. The next question, though, is "the history of what?" In popular parlance the term "history" is often synonymous with human history and indeed recorded human history (as opposed to prehistory). Nonetheless, all things in this world, from humans to other species, geological formations, continents, and galaxies, have their histories; but the people who study nonhuman history have typically not been called historians or historiographers but scientists. Given all of this I shall use the term "history" broadly to refer to events in the past and will contrast human historiographic explanation from natural historiographic explanation. Human history is a part of natural history, but I shall reserve the term "natural history" to refer to the history of nonhuman things. To be strictly proper we might use the term "natural historiography" and "natural historiographer" to refer to those sciences and scientists that are concerned with nonhuman natural history – but unless context demands, I will live with the more standard "natural science" and "natural scientist."

If history is the collection of past events, then historiographic explanation is the explanation of past events. And while the fact that historiography focuses on *past* events is epistemologically significant, from the point of view of explanation, the important point is that historiographic explanation is the explanation of *events* –

that is, singular occurrences situated in particular places and times.[1] These historiographic explanations can be human and sociocultural – explaining why Henry VIII split from the Roman church – or natural, for example, explaining why Pangaea split into the modern continents some 175 million years ago. Human or natural, this form of explanation is causal. To ask why Henry split from the Roman Church or why Pangaea split into the modern continents is to ask what caused these things to happen.[2]

Causal explanation of single events is not the only kind of explanation that historiographers (human or natural) engage in, but it is arguably the primary one. It is also the sort of explanation which distinguishes historiographic explanations from scientific explanations of patterns or regularities. Accordingly, it is this kind of explanation that will be the focus of this essay.

Any answer to questions about the relationship between historiographic and scientific explanation will presuppose certain views about scientific explanation and about the nature of science more generally. My view, and it is a common one (e.g., Cleland 2008; Danto 1985; Kuhn 1991), is that the debate over the relation between historiography and the sciences over much of the last century has been misdirected, because it starts from an image of natural science that is fundamentally mistaken. That image suggests, among other things, that scientific theories are collections of laws, that scientific hypotheses are falsifiable, that observation can be separated from theory, and that social and cultural presuppositions can at least ideally be eliminated from science. In the 50 years since the publication of Kuhn's *Structure of Scientific Revolutions*, this image of science has been extensively revised and has reached a point in which many of the features that supposedly distinguished the natural sciences from the social sciences (including historiography) have vanished.

While it is difficult to summarize all of the features of this revised view of the nature of science, two important developments are (1) that philosophers of science have come increasingly to understand science as a search for mechanisms as opposed to laws of nature and (2) that scientists typically explain natural phenomena by providing idealized models of those mechanisms that cause these phenomena as opposed to complete theoretical descriptions that invoke laws of nature. This shift is important because much of the supposed distinction between explanations of natural phenomena and of human action depends upon the claim that natural phenomena, but not human actions, are law-governed.

---

[1]Historical explanation can also explain facts and states of affairs, which are things that are somewhat different than events, but all of which crucially are "local" – that is, holding at particular places and times.

[2]Tucker (2008) has suggested that philosophical approaches to causation in human history can be divided into two kinds: unificationist approaches suggest that causes in human and natural history are of the same kind, while exceptionalist approaches suggest that human causes are of a different kind or perhaps that human action cannot be understood causally at all. Given that historiographic explanations are causal, the debate about the supposed distinctiveness of historiographic explanation is closely bound to this question about causation in human history, which in turn is connected to more general questions about the nature of causation.

13.2 The Battle of Antietam: A Brief Narrative

I will use the battle of Antietam as my central case for examining aspects of historiographic explanation, so it would be helpful to begin with a brief summary of the circumstances of that battle.[3] The battle took place near the village of Sharpsburg, Maryland, some 60 miles north of Washington DC on September 17, 1862. The battle, which pitted about 75,000 men of Union General George B. McClellan's Army of the Potomac against 55,000 men of Confederate General Robert E. Lee's Army of Northern Virginia, stands as the bloodiest single day of United States history, with around 23,000 dead and wounded.

Earlier in 1862, the Union had appeared poised to defeat the Confederate Army. McClellan had invaded Virginia and had brought a powerful army close to the Confederate capital of Richmond. Largely because of Lee's leadership and McClellan's excessive caution, that campaign ended with the retreat of the Union Army. After McClellan's retreat, Lee went on the offensive. At the end of August, Lee's army defeated another Union Army, the Army of Virginia, under the command of General John Pope in the Second Battle of Bull Run. This opened the way for an invasion into the state of Maryland, a border slave state that had sided with the Union but which had important pockets of Confederate sympathizers. McClellan's army (which had absorbed the remnants of Pope's army) pursued the Confederate Army, catching it near Antietam Creek.

Although McClellan's forces outnumbered Lee's, a lack of coordination and initiative on the part of McClellan and his commanders prevented them from bringing their full forces to bear. In the end, the battle was a stalemate, with neither side claiming the field. Nonetheless, from a strategic point of view, the battle of Antietam is considered a decisive Union victory. Lee's losses were such that he had to retreat from Maryland, ending the threat to the northern states. The effects of this cascaded in a number of important directions. It had a significant impact on midterm congressional elections, allowing Lincoln's Republican Party to maintain its majority in Congress. It made the British and French governments, which had been on the verge of recognizing the Confederacy as a sovereign state, decide not to intervene and call for negotiations. Most importantly, it gave Lincoln a victory that he felt he needed in order to announce his Emancipation Proclamation – the order by which he freed all slaves within the rebellious states. This act changed the war. What had started as a war to suppress a rebellion became a war to free slaves.

One advantage to the battle of Antietam for our case study is that the basic facts about what happened are well known. The events of the American Civil War are relatively recent, participants in and proximal observers of these events were highly literate and documented these events extensively, and the American Civil War has been a subject of sustained historical investigation. These facts allow

[3] Information in this essay about the battle and its context in the American Civil War is drawn from McPherson (2002).

us to focus on the question of how historians explain those facts. In making this choice, I do not want to suggest that knowing the facts is easy. In some historical investigations, there are profound difficulties with establishing the most basic facts, and even in cases where the basic facts are well established, explanation may rely upon discovering hitherto unknown but causally relevant facts. Nonetheless, in what follows, I shall take the facts for granted and focus on the question of how historians assemble facts into explanatory relations.

13.3 Mechanisms and Historical Explanation

The model of explanation I am defending suggests that human historical processes are mechanistic and that narrative historiographic explanations are ultimately descriptions of these mechanisms. In order to specify the content of this claim, we must say something about what mechanisms and mechanistic explanations are. Advocates of the new mechanicism have sometimes disagreed about just what constitutes a mechanism, but there seems, notwithstanding the details, to be something of a consensus about the basic features that all mechanisms share. Illari and Williamson provide a sort of minimal definition that captures this consensus:

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (Illari and Williamson 2012)

A paradigmatic example of a mechanism like a clock consists of a collection of entities (gears, watch hands, crystal, battery, etc.) whose activities (turning, vibrating, etc.) are organized in such a way that they produce some phenomenon, for example, the turning of the hands at a constant speed. While such paradigmatic examples count as mechanisms on this definition, so too do many other things. The sorts of things that can count as entities and activities in mechanisms extend far beyond what appears in classical machines; entities can include anything from molecules to globular clusters, and activities can be anything from the chemical interactions of neurotransmitters, the flowing of rivers, the erupting of volcanoes, and the play of children.

While the new mechanists have argued for the primacy of mechanisms and mechanistic explanation over laws and nomological explanation, laws, or at least non-accidental generalizations, do play an important role in the characterization of mechanisms.[4] Generalizations have two important roles in relation to mechanisms. On the one hand, activities of and interactions between parts of mechanisms can be described by generalizations. If, for instance, two gears interact within some mechanical device, there is a non-accidental generalization that will describe how a change in the position of one gear will produce a change in a position of the

[4]The relationship between mechanisms, laws, and other sorts of generalizations is widely discussed in the mechanisms literature (Glennan 2002, 2011; Andersen 2011; Leuridan 2010).

other gear. These are so-called change-relating generalizations. On the other hand, generalizations can be used to describe the behavior of mechanisms as a whole. For instance, Mendel's laws, which describe relationships between the distribution of genes in parents and offspring, describe an aspect of the behavior of reproductive mechanisms. Such laws (if we are to call these laws) are *mechanically explicable*. Mechanically explicable laws or generalizations, while descriptively essential, are not at the metaphysical heart of the matter, since these laws obtain only in virtue of the existence of mechanisms.

A final important feature of mechanisms is their hierarchical organization. The entities and activities of mechanisms are typically themselves complex, where lower-level mechanisms may explain the properties of these entities and activities. This means, among other things, that the generalizations describing activities and interactions of the entities that are parts of a mechanism will themselves be mechanically explicable. So for instance, if the clock contains a battery that generates a current within the system, the generalizations about how the battery behaves will themselves be explicable by examination of the parts of the battery and the activities and interactions in which they engage.

The term mechanism is sometimes used to refer to systems or structures, while at other times it used to refer to processes (Glennan 2002). Systems are complex "things" – organized collections of entities that act in regular and repeatable ways. Clocks, synapses and stomachs, and legislatures are all mechanical systems. Processes on the other hand are most easily thought of as sequences of activities, interactions, and events. Many processes can be thought of as resulting from the operation of mechanical systems – for instance, stomachs are one of the systems involved in the process of digestion. However, not all processes derive from the operation of a system. Here is a process: I swing a golf club, striking a ball lying on a tuft of grass; the ball travels through the air 150 yards, slicing to the right, landing on the ground whereupon it rolls down a hill into a bunker. There are entities (me, the golf club, the ball, the grass, the bunker, etc.) and activities (swinging, slicing, rolling, etc.) but there is no system here. For one thing, the particular combination of the ball's lie and place on the course is, more or less, unique. For another, my swing is (sadly) not repeatable, so that two swings will not produce the same results. A process like this, that is not the product of the operation of a stable system, is an ephemeral mechanism. More specifically, an ephemeral mechanisms is one in which the way entities and activities are organized is the result of chance or exogenous factors and in which that organization is short lived, non-stable, and not an instance of a multiply-realizable type (Glennan 2010a).

Historical mechanisms are typically best understood as processes rather than systems, and these mechanical processes are to a large degree ephemeral. Indeed, one way of understanding the distinction between historians and social scientists is that historians are concerned with the particularities of processes that lead to particular outcomes at particular places and times, whereas sociologists, political scientists, or economists are concerned with systems that give rise to stable and repeatable processes (cf. Gaddis 2002, Chap. 4).

There is a close connection between mechanistic and narrative explanations. Narrative is the principle mode of explanation of singular historical events (cf. Danto 1985; MacDonald and MacDonald 2008), and it is often thought that the use of narrative explanation is one of the marks that distinguish historiography from the natural sciences. A mechanistic explanation characterizes entities and activities, describing how their organization in space and time gives rise to some phenomenon. This is in essence a narrative.[5]

McPherson's (2002) account of the events on the day of the battle is a typical narrative. Let us consider how this narrative fits within the paradigm of mechanistic explanation. McPherson's account seeks to explain many things about the battle of Antietam, including the circumstances that led to the battle, the decision and indecision of commanders in the battle, the effects of technology, training and terrain on tactical outcomes, the downstream effects of the battle upon the emancipation question and on the national elections, and so on. To illustrate the ways in which this narrative describes a mechanism, let us focus on a singular explanatory question: What explained the tactical outcome of the battle, conceived primarily as the final positions of the armies at the end of the day's fighting, along with the casualties that each army suffered? McPherson's narrative describes the various entities involved the battle, which are for the most part the various military units and their commanders. These are the entities. McPherson also describes the activities and interactions in which these entities are engaged: deliberating, giving and receiving orders, marching, shooting, suffering casualties, retreating, etc. The narrative pays particular attention to the organization of these entities' activities in space and time; for it is upon this that the battle turns. For instance, the casualty rates in a part of the battlefield depend upon the position, orientation, and size of opposing forces. If McPherson and other historians are correct, much of the explanation of the failure of the Union to achieve a more complete victory had to do with their failure to appropriately time their attacks and concentrate their forces. This example illustrates how an event like a battle has all of the key features of a mechanism – entities, activities and interactions, and organization – and how a historical explanation describes these things.

[5]One way of understanding the place of narrative explanation within natural science is to distinguish historical natural science (or natural historiography) from experimental natural science (Cleland 2008). According to this approach, natural historiography is concerned with the representation of past events of natural history and their causes, and so, like human historiography, explains via narrative descriptions of these processes. Experimental science, on the other hand, is concerned with repeatable and law-governed phenomena and, accordingly, uses different forms of explanation. Interestingly, however, the mechanistic approach suggests that even the phenomena studied by experimental science are in fact susceptible to narrative explanation. Regular and repeatable phenomena are simply the products of the operation of widespread and reliable mechanisms. Descriptions of these processes form generalized narratives (Glennan 2010a; Wise 2011).

13.4 Selected Problems of Historiographic Explanation

Using this basic framework of historiographic narrative as mechanistic explanation, I turn now to a selective discussion of some important problems in the theory of historiographic explanation. This discussion will show that the mechanistic approach provides some important resources for thinking about these problems. It will also allow us to see some often unappreciated parallels between explanatory practices in human historiography on the natural sciences that collectively bolster the case for naturalism.

13.4.1 Problems of Object and Event Individuation

Causal relations are most commonly understood as being relations between events. To offer a causal explanation of an event then involves identification of other events that cause the explanandum event. If the battle of Antietam is an event, it will be explained by events in its past and may help to explain events in its future. This conception of causal explanation, however, raises many questions about the nature of events and their descriptions. The central challenge for an advocate of a naturalist and realist approach to causal explanation is to square a broadly ontic conception of explanation – one in which explanatory relations obtain between events that exist independent of mind and theory – with the evident fact that the description of events, and explanatory practices more generally, is deeply dependent upon a variety of pragmatic factors.

Let us consider the battle of Antietam as an event. What makes it the event it is and distinguishes it from other events? Historians do not, to my knowledge, find this question too problematic. The battle is an aggregation of smaller events – marching, shooting, killing, fleeing, etc. – taking place within a well-defined region a couple of miles around Sharpsburg for around 12 h beginning at dawn on September 17, 1962. But how clear is this? Why for instance do we delimit the battle at 12 h, as opposed to including a preliminary skirmish that occurred the evening before or occasional shots fired the day after? A related question concerns how much the identity of an event depends upon its properties and constituents. Had one brigade arrived later on the field than it did, or for that matter one cook arrived later to breakfast, would it have been a different battle? Such questions do not have clear answers, and reflection on them can lead one to the sort of skepticism exemplified by Louis Mink that it is not the case that "there is a determinate historical actuality, the complex referent for all our narratives of 'what actually happened,' the untold story to which narrative histories approximate" (quoted in Ankersmit 2008, p. 202).

If Mink's skepticism is actually warranted, then we should have similar concerns about natural history. Here is one example: Speciation events may have different sorts of causes, but it is generally believed that many speciation events occur as a result of the operation of the mechanism of allopatric speciation. In such

cases, a population of individuals belonging to a species become geographically isolated from other members of that species to a point where interbreeding becomes impossible. Over time, genetic drift and differential selective environment lead to genotypic and phenotypic divergence between populations to the point where a new species is formed.

How exactly does one describe and individuate a speciation event? As with other historical events, there are clearly times before and after the event, but it is difficult to identify when exactly the event begins and ends. Using the biological species concept, populations are members of distinct species when they no longer have the potential to interbreed. But what counts as having the potential to interbreed is a vague and theoretically difficult question. Again, as in the case of Antietam, it is difficult to say how different things would have to be for an event to count as the same speciation event. In allopatric speciation a population often becomes isolated through the creation of a geographical barrier. For instance, flooding might create a boundary between two parts of a population. But which individual organisms end up on which side of the barrier can be a highly contingent affair. Had the organisms that formed the population and its gene pool been slightly different than those that actually did, would it really have been the same speciation event?

So Mink's skepticism, if it is warranted, is as much a problem for natural history as it is for human history. The difficulty is to find a way to answer these questions about how one describes historical objects and events that both recognizes the pragmatic dimensions of such descriptions while saving our intuition that the events in question have a reality independent of those descriptions. This problem has been much discussed by advocates of the new mechanicism, and something of an answer is already implicit in the characterization of mechanisms discussed above. Here again is Illari and Williamson's characterization:

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.

It is key that there is no definition of a mechanism as such, but only of a mechanism *for a phenomenon*. The point (cf. Glennan 1996) is that decompositions of mechanisms into parts can only be carried out in light of a description of what a mechanism is doing. To take a simple biological example, consider all of the various phenomena produced by human bodies – pumping blood, sweating, eating, excreting, moving, playing tennis, writing books, etc. The entities and activities that produce these phenomena can be quite different, and the boundaries will overlap. There are various systems – for example, pulmonary, digestive, muscular-skeletal, nervous, cognitive – which are productive of different behaviors and which divide up a human body and its activities in different ways. We can make a similar point about a system that might be a matter of historical investigation like the United States Congress. The Congress can be decomposed into entities and activities of different and overlapping kinds – by states, by committee affiliation, and by party affiliation to name a few. Different activities undertaken by Congress (different phenomena) will be explained by different causal mechanisms that appeal to these different entities and to the activities in which they engage.

In a causal-mechanical explanation, identification of a mechanism's phenomenon is identification of the explanandum. And as is widely understood, the identification of explananda is dependent upon the context in which explanation is being sought. But this context dependence need not suggest either that choices of explananda are arbitrary or that the resulting articulation of entities, activities, and events do not refer to real things.

Consider more closely some explanatory questions surrounding the battle of Antietam. To seek an explanation is to ask a why question, and there are many different such questions that have been of interest to historians. A basic question is this: Why did the battle of Antietam occur? This question is most commonly posed in the context of the strategic situation during the summer of 1862. The way in which the occurrence of the battle is characterized will be quite coarse grained, because the explanation is essentially contrastive and the implied contrast class involves very different sets of events. So the question of why the battle of Antietam occurred might be cashed out in this way: What caused Lee to invade Maryland and what caused McClellan to chase him and to seek out a battle? The answer to this question will individuate the battle quite coarsely – as a battle taking place between Lee's and McClellan's armies in Maryland in the late summer of 1862. From this strategic perspective, fine-grained descriptions of the time, place, and entities involved are irrelevant. The battle (and its explanation) would still be the same if it had taken place a few days earlier or later, if a few regiments more or less had taken part, or indeed if the battle had taken place some miles away from Antietam Creek. The description that we are really operating under is something like "the battle that occurred in which McClellan attempted to halt Lee's invasion of Maryland." Other questions will individuate the battle more finely. For instance, one might ask why the battle was fought at Antietam Creek as opposed to a few miles away or why it occurred on the 17th of September instead of the 16th.

Once the explanandum is identified, there will be non-arbitrary reasons for articulating the parts of the mechanism responsible for producing the event to be explained. The articulation of the mechanism responsible for the coarse-grained explanation will involve description of various agents whose perceptions and decisions were responsible for the Confederate decision to invade Maryland and the Union response – people like President Lincoln, General McClellan, Confederate President Jefferson Davis, and General Robert E. Lee. At the strategic level, the two armies can be treated as unitary entities. If the explanatory question turns to explaining something like why the battle took place on September 17, the articulations of entities and their activities and their interactions will have a higher resolution. Reference must be made to individual corps, divisions and regiments within the army, their specific locations within the vicinity of the battle, and the various activities and interactions of the numerous commanders and staffs.

While in some sense there is no privileged set of explanatory questions surrounding a historical event, historians have good reasons for choosing a particular question and grain given their larger explanatory interests. The reason, for instance, to focus on the coarse-grained strategic description of the battle of Antietam, is that it was on this coarse-grained outcome that so much of the subsequent history of the American

Civil War appears to have depended. In explaining the historical significance of the battle, it likely does not matter that the casualties were 6,500 instead of 5,000 or that the battle occurred at Sharpsburg rather than Frederick. What made a difference strategically was, among other things, that the invasion of Maryland was halted, that the battle's outcome changed the electorate's attitude toward Lincoln and the Republicans, that it enabled Lincoln to emancipate the slaves, and that it persuaded the British not to intervene in the war.

13.4.2 The Problem of Historical Meaning

Arthur Danto famously argued that historiography has a different character than science because of its use of narrative forms and particularly of a particular sort of description he called a narrative sentence. The distinguishing mark of narrative sentences is that they fix the referent to entities and events that occur in the past by means of events that occur further in the future. An example of such a sentence is the claim "The commander of the Army of Northern Virginia in the Maryland Campaign was born in Virginia." Such a claim describes the birth of Robert E. Lee, but it does so in a way that not even an omniscient "ideal chronicler" could describe at the time that it occurred. The birth of Lee was not the birth of the commander of the Army of Northern Virginia until many years later.

Danto (1985, p. 182) thought that the prevalence of such references in historical sentences showed that historiography was not science, but if one includes in science those fields like astronomy or evolutionary biology which study the origins of particular things, the problem is far from unique. Events of nonhuman history obtain their historical meaning retrospectively just as do those of human history. Consider again the idea of allopatric speciation. Suppose a group of animals crosses a river, leaving some of their brethren behind. Subsequently, the river floods and thereby creates a boundary. Over time natural selection operates on the different populations to such an extent that the descendants of the population become a new species. The ideal chronicler could not have identified this population as the founding population of a new lineage at the time that it split off because at that time there simply was not a new lineage.

The concept of historical meaning, while not uniquely applicable to human history, can be quite helpful to understanding the grounds for non-arbitrarily identifying explanatory questions and explanandum events. In the discussion above I emphasized that the most common way to characterize the event known as the battle of Antietam was coarse grained because it was the event at this grain that was of strategic significance. "Strategic significance" is but another way of talking about historical meaning. The primary reason why historians care about the battle of Antietam is that it appears (retrospectively) to be a turning point in American history. (McPherson's history of Antietam is in fact part of a series from Oxford University Press called "Pivotal Moments in American History.") The battle is both an event to be explained and an event that is crucial to the explanation of

future events. For the historian, the essential characteristics of the battle are not the particular place, time, or participants, but the set of characteristics that allowed it to play this causal role.

13.4.3 The Problem of Agency

Probably the most familiar argument for taking a nonnaturalistic and exceptionalist position toward human historiographic explanation has to do with human agency. Human agents cause things to happen in the world, and if the actions of these agents cannot be woven into the naturalistic fabric of causes and effects, then indeed explaining events in human history would be a very different kind of matter. Philosophers have sometimes suggested that human agents and actions have a number of special properties that make it impossible to integrate them into a naturalistic model of causation and explanation. The actions of human agents are said to be free and undetermined; they are based upon reasons, which cannot be causes; they are not governed by laws.

It is not possible to delve too deeply into these matters in this essay. Suffice it to say that some of these claims about human agency are clearly incompatible with the naturalist thesis. What I would like to argue, however, is that a naturalistic approach to human historiographic explanation need not deny the importance of human agency. In fact, successful mechanistic explanations of human events must take into account certain special properties of human agency; but none of these special problems are genuinely incompatible with a causal and naturalistic approach.

What are these special properties? All of them are connected in some way or other with intentionality. Human agents have beliefs and desires, and the explanation of human action is (in practical terms at least) impossible without them.[6] This is in large part because an agent's actions are not responses directly to what is happening in the world, but to the agent's beliefs about what is happening in the world. Thus, causal explanations of events produced by human actions must appeal to these beliefs.

This feature of human action is especially salient in military history because the actions of both generals and common soldiers are based upon their beliefs about their enemies, and these beliefs are often mistaken in ways that make big differences to outcomes. The army with better reconnaissance or with officers who are smarter (or luckier) in their guesses about the dispositions of their enemies will often be victorious. The battle of Antietam provides numerous examples. During the Maryland campaign, Lee's invading Confederate army had approximately 55,000 men, while the Union Army had more than 75,000 men. To compound this

[6]There is a vast literature in philosophy of psychology, of which Fodor (1989) and Dretske (1988) are representative, that has been directed at developing a naturalistic account of these intentional properties. My analysis assumes that some such account is on the right track.

numerical disadvantage, Lee had divided his forces and, by great good fortune, McClellan had learned of Lee's plans. Notwithstanding these facts, McClellan substantially overestimated the forces arrayed against him and accordingly was slow to press his advantage. McPherson describes McClellan's behavior in the days before the battle:

> On the 16th McClellan had 55,000 troops on hand with another 14,000 within six miles. Lee's force had not yet increased to much more than 25,000. Having informed Washington three days earlier that he would crush Lee's army while it was separated, McClellan had missed his first opportunity to do so on the 14th. He missed his second chance on the 16th as he spent much of the day planning an attack on September 17 – by which time all of the Army of Northern Virginia would be united except for A. P. Hill's division. Without Hill, Lee had 36,000 men, which McClellan tripled in his mind. (McPherson 2002, Chap. 4)

When McClellan finally attacked on September 17, he still held a two-to-one numerical advantage, but he believed he was outnumbered and so held one third of his forces in reserve. Most historians believe that McClellan's failure to commit these reserves, along with similar caution on the part of some subordinate commanders, prevented the Union from achieving what could have been a decisive victory. This was Lincoln's conclusion as well, as he relieved McClellan of command after McClellan failed to pursue the retreating Confederate army.

What this example demonstrates is that what caused the particular outcome at Antietam was as much McClellan's beliefs as the soldiers' weapons. Thus, a narrative explanation of the outcome will inevitably describe those beliefs and what caused them to be formed. Because such large consequences can follow from individual judgments, political and military history make the dependence of human action on beliefs especially clear, but the phenomenon is ubiquitous.

The appeal to beliefs (including false beliefs) is an essential feature of narrative explanations of human historiography, but many explanations in natural science and natural historiography have similar features. In the first case, representation and misrepresentation are essential to explaining many aspects of nonhuman animal behavior. In some cases this behavior will involve states that have many of the same features as human intentional states. For instance, it is difficult to formulate an explanation of many animal behaviors without referring to a predator's beliefs about their prey. Even animals that do not have anything like the mental capacities of humans or other cognitively advanced predators will utilize representations. Bees and ants, for instance, have internal information bearing states that allow them to return to their nests.

The applicability of semantic concepts to the explanation of biological systems in fact goes far beyond their use in the study of animal behavior. Much has been made in the last decade of the concept of information in biology. Genes are often thought of as coding molecular information and developmental information. Adaptations can be seen as representing information about the environment. Critics of gene-centered views of evolutionary biology do not deny that genes carry information, but instead argue that information (and with it, misinformation) is widely distributed across "developmental systems" (Oyama 1985). All of this is just to say that there is no obvious conflict between naturalistic and causal explanation, on the one hand,

and semantic and intentional explanation, on the other. The entities and activities studied by natural scientists, as much as by human historians, may have various semantic properties. And while it is clear that we are far from understanding exactly how to think about such properties, it is equally clear that they are part of the natural world.

13.4.4 The Problem of Laws

As has been widely noted, one of the chief reasons to take an exceptionalist attitude toward historiographic explanation has had to do with the suspicion that laws do not figure in historiographic explanation in the way that they do in scientific explanation; to the extent that philosophers of science have in recent decades discredited nomological approaches to scientific explanation, this argument against naturalism has lost much of its force. Nonetheless, it is implausible to think that either scientific or historiographic explanations do not rely in important ways upon generalizations that express non-accidental regularities.

As indicated in our preliminary discussion of mechanisms, generalizations play a twofold role in the description of mechanisms – they can describe the behavior of a mechanism as a whole, or they can describe the character of the entities, activities, and interactions that produce that behavior. Let us focus first on this latter role by considering some generalizations that play a role in the explaining events surrounding the battle of Antietam.

An ephemeral mechanism is one whose arrangement of parts is fragile, short lived, and one-off, but in which the activities of and interactions between those parts – given their relatively stable properties and dispositions – will be robust and regular. To take a simple example, the circumstances that might lead a single gun crew to fire a round at a particular moment and place in a battle will be ephemeral, but the interaction between a match and loaded cannon is quite robust and regular, as is the interaction between a cannonball and its target. An effective narrative explanation will show how these various pieces came together, and how, given this organization, the stable dispositions of the parts interact to produce the outcome.

One set of generalizations that is important in the explanation of human historical events is the generalizations describing human dispositions. These can be generalizations about the behavior of human beings generally, about the behavior of specific groups of human beings (e.g., mid-nineteenth-century West Point-educated officers) or about the behavior of specific people. Let us consider some generalizations about the behavior of specific people that are relevant to the explanation of events at Antietam. The two commanding generals, McClellan and Grant, had rather different dispositions as commanders, and it is possible and informative to form generalizations about them. McClellan, as alluded to above, was very cautious and was inclined to overestimate the strength of forces arrayed against him. Lee, on the other hand, was a risk taker, inclined to leave certain areas unprotected so that he could go on the attack and keep his opponent off balance. Generalizations like this

are crucial to explaining the generals' actions and with them the outcome of the battle. Consider this narrative of events in the center of the battlefield during the afternoon of September 17:

> The broken Southern brigades fell back in disorder almost half a mile. Lee's center was wide open except for some artillery and a handful of dazed infantrymen that Confederate officers including Longstreet desperately scraped together back along the Hagerstown Pike. "There was no body of Confederate infantry in this part of the field that could have resisted a serious advance," wrote a Southern officer. "Lee's army was ruined," added Longstreet's artillery commander melodramatically, "and the end of the Confederacy was in sight." Now was the time for McClellan to send in his reserves. Longstreet himself later said that if 10,000 fresh Union troops had been put in at that juncture, the Confederates would have been swept from the field.
>
> McClellan had those 10,000 available in Franklin's corps, and several thousand more in Porter's. The normally cautious Franklin pleaded to be unleashed. But Sumner, who was still shocked by what had happened to Sedgwick's division, counseled against it. Fearing that Lee must be massing his own supposedly abundant reserves for a counterattack, McClellan accepted Sumner's advice. ... So the opportunity passed. (McPherson 2002, Chap. 4)

A crucial point in explaining the outcome of the battle is explaining McClellan's failure to commit troops to exploit the Confederate retreat. What caused McClellan to make this choice was the evidence and advice presented to him, in combination with his own dispositions and judgment. His disposition to caution was robust and stable. On multiple occasions in the battle of Antietam, as well as in earlier battles during the peninsular campaign, McClellan had failed to exploit opportunities because of a tendency to overestimate the forces arrayed against him. Thus, McClellan's decision at this point is predictable and explanatory.

It would be odd to call this generalization a law, but it is invariant in the sense of Woodward (2003) and it is mechanically explicable. The generalization is simply a description of McClellan's dispositions. McClellan does not act as he does because of the generalization; rather the generalization holds true because of the particular psychological structures which constitute McClellan's personality, and these in turn have a history of particular causes.

There are many other generalizations besides generalizations about the psychological dispositions of agents that may play a role in historical explanation. In military engagements, for instance, there are numerous explanatory generalizations about the ways in which opposing forces might interact – for example, of the susceptibility of certain kinds of infantry formations to artillery file, of the favorable or unfavorable effects of terrain, or of the amount of casualties that typically will lead to the destruction of unit cohesion. These generalizations can be explanatory and also predictive. It is indeed their belief in the truth of these generalizations that explains why commanders made the choices they do. Confederate commanders chose to concentrate their forces around a bridge over Antietam Creek because of their beliefs about the defensive advantages of such a position. Such generalizations describe real regularities, but these regularities arise because of the similarities across particular mechanisms.

The fact that generalizations are mechanistically explicable helps to explain their *ceteris paribus* and exception-ridden character. Mechanistically explicable

generalizations only hold true in the right context. Given violations of certain background or boundary conditions, the mechanism will break and the regularity the generalization describes will fail. For example, artillery batteries are mechanisms for firing cannon balls, and there are non-accidental generalizations describing this behavior, including such properties as range and rate of fire. But these generalizations hold true only in virtue of a wide variety of background conditions. The rate of fire depends, *inter alia*, on the location and availability of munitions and on the health, level of fatigue, and psychological state of gun crews.

There is nothing special that distinguishes how generalizations figure in human historiographic explanations from how generalizations figure in natural historiographic explanations, except for the fact that some of the generalizations that figure in human historiographic explanation will be generalizations about intentionally driven human behavior. For the sake of comparison, consider some generalizations involved in explaining the outbreak of a forest fire. The particular circumstances that explain the ignition of a forest fire will be ephemeral – a chance lightning strike or a wind gust that ignites the embers of a passing backpacker's campfire. But there are many generalizations that will figure into explaining how likely a fire is to occur in an area, how far a fire spreads, how hot it burns, and when it ends. Those generalizations will reference things such as the climatological conditions like wind speed and humidity, the kind of growth in the forest, the rainfall in that year, and so on. And while these generalizations can help both to predict and explain the progress of a forest fire, they are not laws in the realist sense, but simply descriptions of the various mechanisms involved. Ultimately, what causes a forest fire are the local interactions of the various parts, single sparks, individual trees, and very local weather conditions.

13.4.5 The Problem of Contingency

Finally, let us consider the role of necessity and contingency in mechanistic explanations of human history. While the idea of contingency is often associated with indeterminism, especially indeterministic interpretations of human freedom, I will follow Ben-Menahem, who suggests that "contingency and necessity be understood in terms of stability, that is, sensitivity or insensitivity to initial conditions and intervening factors" (2008, p. 121). Contingent events are, on this view, just as causally determined as necessary events. Events are contingent when small changes in causal antecedents lead to significant changes in outcomes. A simple physical example will illustrate the difference. If I drop a marble anywhere inside a hemispherical bowl, it will, regardless of where it landed inside the bowl, eventually settle at the bowl's center; if I turn the same bowl over and drop the marble on top of it, the marble's final resting place will vary widely depending where the marble landed exactly, as well as on its spin and velocity. The former outcome is necessary, while the latter is contingent. Contingency in this sense is a familiar feature of nonhuman natural history. The historical conditions which give rise to

planets, geological formations, or species may be highly contingent. The degree to which the shape of "the tree of life" is contingent is in fact a widely discussed problem (Gould 1990; Beatty 1995). The notion of historical contingency is closely related to the ephemerality of historical mechanisms. Mechanisms are ephemeral to the extent that the organization of entities and activities is contingent.

The events leading up to the battle of Antietam provide a spectacular example of contingency. As McPherson narrates, on September 12, 1862:

> the Army of the Potomac marched into Frederick greeted by delirious citizens waving flags, kissing McClellan, and hugging his horse. The 27th Indiana stopped that morning in a farm field outside of town. Corporal Barton W. Mitchell flopped down in the shade of a tree along a fenceline to enjoy a welcome rest. As he relaxed, however, Mitchell noticed a bulky envelope lying in the grass. Curious, he picked it up and discovered inside a sheet of paper wrapped around three cigars. As a comrade went off to hunt for a match so they could smoke their lucky find, Mitchell noticed that the paper contained writing under the heading "Headquarters, Army of Northern Virginia, Special Orders, No. 191." (McPherson 2002, Chap. 4)

Special Order 191 contained Lee's orders dividing his forces into four parts. The orders were passed on to McClellan, and based upon them, McClellan issued orders to move his army to catch Lee's portion of the divided force. These troop movements led, 5 days later, to the engagement at Antietam.

Corporal Barton's fortunate discovery illustrates the interplay between necessity and contingency that is characteristic of ephemeral mechanisms. While the events leading to Barton's picking up the orders were highly contingent, the consequences of that event were not. The parts of the Union command and control mechanism functioned as expected and the order was passed on until it reached McClellan. McClellan's staff, in keeping with their professional training, made efforts to authenticate the document and correctly judged the orders to be genuine. McClellan, reflecting his professional training, recognized that this information could be decisive in bringing about his goal of defeating the Confederate army. Reflecting his famed excess of caution, however, he was methodical in his preparations to move his army, and he did not get his troops moving until 18 h after he saw Lee's order. Lee, by another stroke of good fortune, received intelligence on the 14th that his plans had been compromised and, being Lee, moved very quickly to concentrate his forces, mitigating considerably the advantage McClellan had gained.

It is cases like these that lead historians to emphasize the unique character of historical narratives. But necessity is a matter of degree, and different historical processes will have greater and lesser degrees of contingency. Compare Ronald Reagan's electoral victory over Walter Mondale in the US presidential election of 1984 with President George W. Bush's victory over Al Gore in 2000. Reagan's victory was a landslide, carrying 49 states and winning the popular vote by nearly 20 %. In contrast, George Bush won a disputed election, losing the popular vote, and only gaining the slightest edge in the electoral vote. The outcome of the latter race was highly contingent, depending ultimately (it appears) upon a 5–4 vote of the US Supreme court.

While there are, objectively speaking, more or less contingent historical processes, contingency also depends upon the grain at which explanandum events are described. The coarser the grain, the less contingent outcomes appear. That the Union and Confederate armies met on September 17 near the village of Sharpsburg, Maryland, was highly contingent upon operational details, including, most notably, the discovery of Order 191. That the Union and Confederate armies met that fall somewhere in Maryland was considerably less contingent. McClellan was under orders to pursue and engage the Confederate Army, and it was likely, given the imperatives under which the various commanders operated, that a major battle would necessarily have occurred sometime that fall – "necessarily" in Ben-Menahem's sense of stability.

Reflection on this case shows that there are very close parallels between issues in human and natural history. Just as with a case like Antietam, some events may be more contingent than others, but judgments of contingency will be connected to explanatory grain. Also, in both areas of inquiry, there are good reasons for explanatory pluralism – looking both for detailed and more contingent narratives about particular individuals and events and for coarse-grained but stable historical patterns (Sterelny 1996). It is tempting to argue that human affairs are more contingent than other parts of natural history, but this is likely an artifact of perspective. We naturally identify imaginatively with individual human beings and life-changing events for individual human beings are, as we know too well, highly contingent. If, though, we were to take a personal interest in a single grain of sand on the beach, I expect we would find a degree of contingency not unlike that which characterizes our individual lives. But we do not worry about the individual grain of sand, choosing instead to focus on the predictable long-term changes to the beach.

13.5 Conclusion

Philosophical thinking about human historiography has frequently involved articulating reasons why there is something special about human history, something which demands special methods and modes of explanation. I cannot think of a better way to counter this view than to show that some important and putatively special features of human historical explanation in fact have close parallels in the explanation of natural events and phenomena. In saying that human historiography is not special, I am not suggesting that there are or should be no differences between the explanatory practices of historiography and the natural sciences; I am only saying that the variations between these practices are no greater than one finds within the practices of the natural sciences themselves. Reflection on parallels between human and natural history and historiography not only helps make the case for naturalism; it also reminds us of the many historical questions we find within the traditional domains of natural science.

References

Andersen, H. K. (2011). Mechanisms, laws, and regularities. *Philosophy of Science, 78*(2), 325–331.

Ankersmit, F. R. (2008). Narrative and interpretation. In A. Tucker (Ed.), *A companion to the philosophy of history and historiography* (pp. 199–208). New York: Wiley-Blackwell.

Beatty, J. (1995). The evolutionary contingency thesis. In G. Wolters & J. G. Lennox (Eds.), *Concepts, theories, and rationality in the biological sciences. The second Pittsburgh-Konstanz colloquium in the philosophy of science*. Pittsburgh: University of Pittsburgh Press.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in the History and Philosophy of Biology and the Biomedical Sciences, 36*(2), 421–441.

Ben-Menahem, Y. (2008). Historical necessity and contingency. In A. Tucker (Ed.), *A companion to the philosophy of history and historiography* (pp. 120–130). New York: Wiley-Blackwell.

Cleland, C. E. (2008). Philosophical issues in natural history and its historiography. In A. Tucker (Ed.), *A companion to the philosophy of history and historiography* (pp. 44–62). New York: Wiley-Blackwell.

Danto, A. C. (1985). *Narration and knowledge: Including the integral text of analytical philosophy of history*. New York: Columbia University Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge: MIT Press.

Fodor, J. (1989). Making mind matter more. *Philosophical Topics, 17*(1), 59–79.

Gaddis, J. L. (2002). *The landscape of history: How historians map the past*. New York: Oxford University Press.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*(1), 49–71.

Glennan, S. S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(3S), 342–353.

Glennan, S. S. (2010a). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*(2), 251–266.

Glennan, S. S. (2010b). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research, 81*(2), 362–381.

Glennan, S. S. (2011). Singular and general causal relations: A mechanist perspective. In P. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 789–817). Oxford: Oxford University Press.

Gould, S. J. (1990). *Wonderful life: The burgess shale and the nature of history*. New York: Norton.

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*(1), 119–135.

Kuhn, T. S. (1991). The natural and the social sciences. In D. R. Hiley, F. J. Bohman, & R. Shusterman (Eds.), *The interpretive turn: Philosophy, science and culture* (pp. 17–24). Ithica: Cornell University Press.

Leuridan, B. (2010). Can mechanisms really replace laws of nature? *Philosophy of Science, 77*(3), 317–340.

MacDonald, G. & MacDonald, C. (2008). Explanation in historiography. In *A companion to the philosophy of history and historiography* (pp. 131–141). New York: Wiley-Blackwell.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

McPherson, J. M. (2002). *Crossroads of freedom: Antietam*. New York: Oxford University Press.

Oyama, S. (1985). *The ontogeny of information: Developmental systems and evolution*. Cambridge: Cambridge University Press.

Sterelny, K. (1996). Explanatory pluralism in evolutionary biology. *Biology and Philosophy, 11*(2), 193–214.

Tucker, A. (2008). Causation in historiography. In *A companion to the philosophy of history and historiography* (pp. 98–108). New York: Wiley-Blackwell.

Wise, M. N. (2011). Science as (historical) narrative. *Erkenntnis, 75*(3), 349–376.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Chapter 14
History and the Sciences

Philip Kitcher and Daniel Immerwahr

Abstract The apparent power of the covering-law model of scientific explanation inspired efforts to make historical explanation fit within it. After the demise of that model, many philosophers of history have proposed more liberal approaches to historical explanation, and some reflective historians have questioned the thesis that offering explanations is the business of good history. We attempt to sort through a number of conflicting ideas about historical explanation and about the historian's commitment (or duty?) to offer the truth about the past. We suggest that histories are diverse, that historians sometimes provide explanations, that the types of explanations they offer are highly various, and that delivering the truth is often important. The picture that emerges illuminates the sciences, by reminding philosophers of the range of questions to which scientific research is directed. It also brings out affinities, not only between history and the natural sciences but also between history and anthropology and history and literature. None of these enterprises should be seen in light of a simple model of successful inquiry. None should be viewed as committed to a single monolithic aim.

Keywords Explanation • Objectivity • Aims of history • Aims of science

---

We are delighted to dedicate this chapter to Arthur Danto.

P. Kitcher (✉)
Department of Philosophy, Columbia University, New York 10027, NY, USA
e-mail: psk16@columbia.edu

D. Immerwahr
History Department, Northwestern University, 1881 Sheridan Road, Evanston 60208, IL, USA
e-mail: daniel.immerwahr@northwestern.edu

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_14,

14.1 Introduction

The history of philosophical reflection on history is dominated by attempts to determine the relation between history and the sciences. To a large extent, the times at which the discussion is turned in a new direction are marked either by a sense that the practice of history is at odds with accepted stereotypes or by a novel account of the character of the natural sciences. In the twentieth century, two such crucial moments are C.G. Hempel's publication of his article "The Function of General Laws in History"[1] and Arthur Danto's publication of *Analytical Philosophy of History*.[2] Hempel's article aimed to resolve a long-standing controversy about the relationship between history and the *Naturwissenschaften* (a debate that had raged particularly fiercely in the late nineteenth and early twentieth centuries, although there had been important earlier eruptions[3]), by deploying a philosophical reconstruction of the natural sciences that he and his colleagues were in the process of developing.[4] Danto's penetrating study of issues in the philosophy of history appeared when the account of the sciences offered by Hempel and others had begun to be challenged, and Danto was quite conscious of the fact that some of the emerging perspectives on the sciences might enable the philosopher of history to accommodate many of the complaints that historians and historiographers had leveled against Hempel's approach to historical explanation.[5]

Since 1964, even since 1985, philosophical views about science – or better, philosophical views about the sciences – have changed again. At the beginning of the twenty-first century, it may no longer be possible to talk of a consensus view about the character of the sciences. Yet because so much of the contemporary

[1] Originally published in *The Journal of Philosophy* in 1942. We'll refer to the reprinting in Hempel (1965).

[2] Danto (1964) from Action, Art, History: Engagements with Arthur C. Danto, by Daniel Herwitz, Michael Kelly. Copyright © 2007 Columbia University Press. Reprinted with permission of the publisher.

[3] Wilhelm Dilthey and Benedetto Croce are the most prominent advocates of the difference between history (as a *Geisteswissenschaft*) and the natural sciences, and Hempel's article is naturally read as responding to them – despite the fact that he doesn't mention them and takes, as his official target, a related view advanced by Maurice Mandelbaum. We think that the controversy about the scientific status of history goes back at least to the Enlightenment.

[4] Essentially, the view that has become known as "logical empiricism," articulated by Hempel, Rudolf Carnap, Hans Reichenbach, Ernest Nagel, and Karl Popper; although there were important differences among these philosophers, all shared the following views: (1) Scientific theories are deductive systems. (2) Scientific laws are universal generalizations. (3) Laws and theories are tested by deriving from them statements that can be tested by empirical observation. (4) Scientific explanation consists in producing arguments that use general laws to derive a description of the phenomenon to be explained. It is worth noting explicitly that, despite the prominence of Hempel's work in discussions of historical explanation, much more detailed (and more nuanced) proposals about historical explanation and the role of general laws were offered by Nagel and Morton White.

[5] Danto (1985, p. xi); Danto refers to the impact of the work of N.R. Hanson and, particularly, of Thomas Kuhn. We agree that Hanson, and especially Kuhn, forced a rethinking of many points that logical empiricism had taken for granted.

discussion in the philosophy of history seems to take for granted views about the natural sciences that virtually no philosopher of science writing today would accept, it seems worth returning to the old question of the relation between history and the sciences with a more up-to-date philosophical perspective. We'll approach the issue against the backdrop of a view of the sciences that one of us has developed elsewhere[6]; if the account of the sciences proves unconvincing at particular points, we hope that philosophers of science who hold different views will be inspired to address the status of history using their own preferred accounts.

14.2 Some Theses to be Scrutinized

To assert the kinship of history and the sciences might be to claim any of several different things. The most prominent candidates, however, are to maintain a kinship of method, a common aim, or similar achievements – or, of course, any combination of these.[7] Superficially, of course, there are major differences between the methods employed by historians and those used by stereotypical natural scientists; historians aren't noted for their propensity to perform experiments – rather, they trudge off to archives, assemble documents (and other remains of the past), and scrutinize them. To the extent that philosophers have sought community of method throughout the natural sciences, however, they haven't hoped to discover it at this level of description; it's a commonplace that there are areas of scientific inquiry in which investigators make observations rather than performing experiments (particularly in astronomy and in the study of animal behavior); and, as we might expect, "historical sciences" like paleontology and historical geology reveal researchers who use the rock strata very much in the way that historians use their archival sources.[8]

One principal theme in the suggestion that the natural sciences share a common method has been the idea of a theory of confirmation that applies irrespective of subject matter. However natural scientists obtain their data, it's supposed that there are general standards for assessing the degree to which the data support the hypotheses they entertain. Proposals about these standards are controversial, and some philosophers have been skeptical of the notion that context-independent standards can be precisely formulated.[9] Insofar as one focuses on the particular

[6]See Kitcher (2001b); some parts of the picture are articulated in more detail in an earlier book, (Kitcher 1993a), but, where there are differences, the views of the later book are to be preferred.

[7]Of course, claims about aims typically constrain theses about achievements and about methods, so one shouldn't assume that all these elements can vary independently.

[8]It's no accident that we talk of the "fossil record" and the "rock record." And, of course, one of Darwin's most extensive defenses of his views draws an extended analogy between the sequence of organisms whose remains have been preserved and an incomplete, tattered, and defaced library (*Origin of Species*, Chap. IX [closing paragraph]).

[9]Kuhn is a major source of this kind of skepticism, not only in *The Structure of Scientific Revolutions* (1962) but also in "Objectivity, Value Judgment, and Theory Choice" (in Kuhn 1977).

claims that historians often defend in their writings – "When Parliament began to 'tell stories to the *People*' in the Grand Remonstrance of 1641, the members had no intention of deposing their king."[10] "Joan of Arc merely checked the English advance by reviving Dauphinist morale, and the Regent managed to halt the counter-offensive. It was not the Maid who ended English rule in France"[11] – we think that there's no reason to hold that the standards for assessing evidence are any different in history than they are in the natural sciences or, indeed, in the social sciences, in literary attribution, musicology, the reconstruction of artworks, criminal detection, plumbing, salesmanship, or whatever. If a satisfactory formal theory of confirmation can be given for the natural sciences, in all their forms, we see no reason to think it wouldn't work equally well for all kinds of inquiry, including the marshalling of evidence for specific historical theses; if no such theory is available, it will remain possible to identify the kinship between historical inferences and arguments and those that are advanced in many areas of scientific inquiry.

We shall not elaborate on these points because we think that the interesting issues about the relation of history and the sciences concern *aims* and *achievements*, not standards of evidence. Those issues arise in two ways, depending on the guiding conception of the aims of the sciences. If one assumes, as many writers tacitly or explicitly do, that the natural sciences aim at truth, the controversy can be formulated in terms of whether truth is an aim of history.[12] Alternatively, if the starting point is the familiar suggestion that the aims of the natural sciences are explanation, prediction, and control, the dispute emerges rather differently. With a small number of exceptions, most writers about history will agree that historians rarely aim to predict or control, so that the kinship between history and the sciences will be debated in terms of the aspiration to provide explanations in history that are akin to those offered by natural scientists. This, of course, is the classic way in which Hempel raised the question, and from 1942 to the present, many scholars have taken the issue of the relation between history and the sciences to be a question about the character of historical explanation.[13]

There are thus several theses that we think it worthwhile to present explicitly.

For a thorough survey of the leading contender for an account of scientific confirmation, see Earman (1992).

[10] Morgan (1988, p. 55).

[11] Seward (1999, p. 213).

[12] Here, one may compare Berkhofer (1995) and Appleby et al. (1994).

[13] Partly because of Danto's important book and partly because of historians' concerns about styles of history ("narrative" versus "analytic"), this has shaded into a discussion of the character of historical narratives (See, e.g., Roberts (2001), which collects many of the classic contributions). As we'll try to make clear, we think that these disputes have been stymied because of failure to probe the broader questions about the aims of the sciences and of history.

- *Veritism about the sciences*. The natural sciences aim at, and sometimes achieve, truths about various aspects of nature.
- *Bernardism about the sciences*.[14] The natural sciences aim to provide explanation, prediction, and control.
- *Veritism about history*. History aims at, and sometimes achieves, truths about various aspects of the past.
- *Impracticality of history*. History rarely, if ever, aims at prediction and control.
- *Explanationism about history*. History aims to provide explanations.
- *Strong explanationism about history*. The principal aim of history is to provide explanations.

We think most of these theses deserve considerable scrutiny.[15]

14.3 Problems with Strong Explanationism

One exception is the uncontroversial *Impracticality of history*; despite the prevalence of slogans advertising the importance of learning from history, we take it that their plausibility rests on the thought that a historical account of some past events can provide the basis for a hypothesis in some area of social science, a hypothesis that can then be applied to a new context.[16] If the *Impracticality* thesis is in place, then it's not hard to understand how those who believe in the kinship

[14]This thesis is named for Claude Bernard, whose study of experimental physiology and medicine is one of the classic sources of the view that the sciences aim at explanation, prediction, and control.

[15]As Isaac Levi pointed out to us, this list would be rejected by many philosophers in the pragmatist tradition, who would set up the issues very differently. We take the point and regard our list as emblematic of a version of logical empiricism that is antithetical to pragmatist themes and modes of formulation – a version more evident in Carnap and Hempel than in Nagel. We also believe that the view of the sciences and of history's relations to the sciences that we elaborate below is far more akin to the pragmatist approach to scientific inquiry.

[16]Perhaps this is too quick, in that there are instances in which history might be credited as the ultimate source of claims, advanced by economists or political scientists who make predictions – as, for example, when they suggest that economies planned by powerful authoritarian governments have a high probability of leading to disastrous consequences for the citizenry. Moreover, as with the lines between history and other disciplines, the distinction between history and economics (or that between history and political science) may be blurred. Even if these caveats about the *Impracticality* thesis are correct, they will not affect the main conclusion we draw from it, for it would be hard to dispute the idea that any social scientific predictions drawn from history are obtained through the historical explanations that have been given, and this would leave intact the primacy of explanation as a goal.

between history and the sciences are led towards – if not all the way to – *Strong explanationism about history*.

Start with *Bernardism about the sciences*. If history shares an aim with the sciences, then it must be that history aims at explanation and/or prediction and/or control. By *Impracticality of history*, explanation is the only candidate. Hence, we arrive at *Explanationism*. But suppose *Strong explanationism about history* were false. Then, the provision of historical explanations wouldn't be the most central or prominent aim of history, so that there would have to be some *nonscientific* ends of equal or greater importance. Hence, an account of history that assimilated it to the sciences would be inadequate. Conclusion: To assert a kinship between history and the sciences, you need something like *Strong explanationism about history*.

As Hempel saw, *Strong Explanationism* isn't enough. To demonstrate the kinship, you also have to scotch doubts about differences between the explanations provided by scientists and those offered by historians. So a familiar dialectic begins. Hempel presented a general model of explanation, according to which explanations are arguments whose conclusion describes the phenomenon to be explained and among whose premises is at least one general law. Faced with the obvious objection that historians rarely[17] state (or are in a position to state) general laws, Hempel suggested that they don't offer *complete* explanations but only *explanation sketches*. As Danto (1985, 203ff) saw, there's some tension in the proposal that scientists achieve their aims, but that the principal aims of history are almost always unattainable – at least if one wants to emphasize the kinship between history and the sciences. Hence, *the* task of the analytical philosophy of history came to be that of showing how historians achieved a distinctive form of explanation, narrative explanation, and that this was quite respectable. Exposing "the structure of narratives" became a cottage industry – and the wheels still hum.

We suggest that Hempel's particular account of science infects the whole discussion at quite an early stage. This isn't simply a matter of adherence to the covering-law model of explanation but also the unquestioned deployment of the categories of *Bernardism* and the paradigms on which Hempel and his co-workers concentrated their attention. We'll start with the theses *Explanationism about history* and *Strong explanationism about history*.

It seems clear to us that there are some historical works that do attempt to provide explanations. Historians have offered rival explanations for the fall of the Roman Empire (in 410, 476, or 1514!), the outbreak of the Civil War in England, the growth of the abolitionist movement in North America, and the origins and course of the First World War.[18] In some instances, the historian focuses on a particular event –

[17]There are important exceptions. Historians sometimes draw on generalizations about the transmission of infectious agents or about the effects of various kinds of missiles; for examples, see McNeill (1976) and Keegan (1978).

[18]We offer a handful of representative texts: Gibbon *The Decline and Fall of the Roman Empire;* Grant (1976), Stone (1972), Davies (1966), MacDonald (1987), Keegan (1998), and Ferguson (1999).

a change of government, a battle, and the official acceptance of some doctrine – and tries to show why that event occurred; in other instances, the project involves linking a sequence, or a complex web, of events and of explaining why all these events occurred.[19] Philosophical reflections on history have been largely directed towards works of these types – at cost, we think, of overlooking other kinds of historical venture.

In *Montaillou*, Emmanuel Le Roy Ladurie does something very different. He uses the records of the Inquisition, which tracked down the Cathars (proponents of the Albigensian heresy) in the early decades of the fourteenth century in the villages of the French slope of the Pyrenees (Montaillou is a mountain village that sheltered an unusually large number of Cathars and Cathar sympathizers). Le Roy Ladurie is not principally interested in explaining why (or how) the tireless inquisitor Jacques Fournier succeeded in routing out the heretics or anything similar. His aim is to take us into the world and the lives of the Pyrenean community, constructing the kind of ethnographic account that an anthropologist might give for some distant group. Summing up his conclusions about the conflicts between local clans in Montaillou, he introduces an obvious but useful image:

> The study of Montaillou shows on a minute scale what took place in the structure of society as a whole. Montaillou is only a drop in the ocean. Thanks to the microscope provided by the Fournier Register, we can see the protozoa swimming about in it. (Le Roy Ladurie 1979, p. 276)

One of the most prominent "protozoa" under Le Roy Ladurie's "microscope" is a likeable shepherd, Pierre Maury, to whose actions and attitudes he devotes over 60 pages. Here is a typical passage:

> . . . Pierre Maury had his leisure moments. When necessary he got his friends to look after his sheep for him while he went down to the neighboring town, to take, or to collect, money (iii.166). Or he might absent himself for purely personal reasons, without any problems of time-keeping or supervision, to go and visit friends, mistresses (unless they came up directly to see him in his *cabane*) or fellow-sponsors, friends acquired at baptisms recently or long ago. (Le Roy Ladurie 1979, p. 174)[20]

Nothing much is *explained* here, but this passage, in combination with plenty more like it, provides us with a picture of how Pierre Maury lived. We learn what it was like to be a Pyrenean shepherd, with heretical leanings, at a particular time (Or, attending to Le Roy Ladurie's claim about the relation between Montaillou and the broader society, we learn what it was like to be a French peasant at a particular time).

---

[19] Thus, one of the differences among historians who try to explain the origins of the First World War consists in their specification of the congeries of events that are to count as the beginning of that war; this difference in *explananda* doesn't occur when the task is to explain something like Constantine's declaration that Christianity was to be the official religion of the Empire.

[20] A *cabane* is a mountain hut, typically occupied by several shepherds and constituting a social unit; the reference to sponsoring indicates Pierre's involvement in Cathar religious practices; the parenthetical numerical reference is to the published version of Fournier's inquisitorial register.

Similar points could be made with respect to other influential microhistories. Natalie Zemon Davis' *The Return of Martin Guerre* is, as she says, concerned not just to bring forth the actions of her subjects but also "the world they would have seen and the reactions they might have had."[21] Unlike Le Roy Ladurie's *Montaillou*, Davis' book tells a story – indeed a compelling story.[22] A Basque peasant, Martin Guerre, goes off adventuring, leaving his home and his wife, Bertrande de Rols. Years later, another peasant, Armand du Tilh, comes to the town, claiming to be Martin. Because he looks like Martin and has learned many things about Martin, he is hesitantly accepted as Martin – even by Bertrande, who surely knew the difference. Armand played his role successfully for a few years, before his trial, at which he was being prosecuted (unsuccessfully) for fraud, was interrupted by the entrance of the real Martin Guerre.

Davis is clearly interested in explaining a number of things, including why Armand decided to impersonate another man and why Bertrande accepted the impersonation. Giving these explanations depends on doing something else, something we take to be Davis' overarching aim, namely, to enable a contemporary reader to understand what it was like to live in a particular sixteenth-century French culture. Davis doesn't simply list the reasons that might have moved Bertrande: she tries to make us view the world through Bertrande's eyes. Here is part of her account:

> What Bertrande had with the new Martin was her dream come true, a man she could live with in peace and friendship (to cite sixteenth century values) and in passion. It was an invented marriage, not arranged like that of her own of eighteen years earlier or contracted in a customary way like that of her mother and Pierre Guerre. It started off with a lie but, as Bertrande described it later, they passed their time "like true married people, eating, drinking, and sleeping together." ... In the marriage bed of the beautiful Bertrande things now went well.[23] Within three years, two daughters were born to them. ...
>
> The evidence for the relationship between the new Martin and Bertrande comes not from this peaceful period of three years, but from the time when the invented marriage was called into doubt. Yet it everywhere attests to his having fallen in love with the wife for whom he had rehearsed and her having become deeply attached to the husband who had taken her by surprise. When he is released from prison in the midst of later quarrels, she gives him a white shirt, washes his feet, and receives him back into her bed. When others try to kill him, she puts her body between him and the blows. Before the court he addresses her "gently"; he puts his life in her hands by saying that if she swears that he is not her husband he will submit to a thousand deaths.[24]

The power of this passage is to make Bertrande, and her apparently odd behavior, comprehensible to us. Davis does this not simply by laying out Bertrande's reasons for accepting Armande but by making her emotions immediate to us – by presenting the couple as tender lovers. After we read this account, Bertrande no longer seems alien, because Davis has given us a way to assimilate her experiences to our own.

---

[21] Davis (1983, p. 5).

[22] As witnessed by the fact that it became a moderately successful film.

[23] The real Martin had apparently been impotent (Davis 1983, p. 19); things went better after Martin's return – see ibid., p. 124.

[24] Davis (1983, pp. 44–46). See also ibid., p. 55, p. 61, pp. 79–80, and p. 92.

And that, of course, is the principal point. The *entrée* to the world of sixteenth-century French peasants isn't a means to answering the burning question "Why did Bertrande de Rols accept the false Martin?" but rather the end at which Davis is aiming. Skillful historian than she is, Davis has combined her introduction to a past culture with a particularly poignant story, so that readers want answers to questions that can only be addressed by entering the culture. Those in the grip of *Strong Explanationism* might insist that the aim of *The Return of Martin Guerre* is to answer explanation-seeking questions – like "Why did Bertrande accept Armand as Martin?" – but this is to overlook the enormous difference between such questions and the usual paradigms, questions like the following: "Why did Constantine declare that Christianity was to be the official religion of the Empire?" "Why did Napoleon lose at Waterloo?" By Davis's own lights, Bertrande wasn't a historically important person who did historically consequential things[25]; she is interesting because she is a gateway through which we can enter a strange and intriguing world.

We've already emphasized the similarity between the historical works we've been reviewing and projects in ethnography, and it's easy at this point to take a wrong turn by invoking an influential *theory* of what such ethnographies do. Many historians and anthropologists have been inspired by Clifford Geertz' famous essay "Deep Play: Notes on the Balinese Cockfight" and by his deployment of the Rylean notion of "thick description." It's become fashionable to suggest that historians – or really up-to-date historians – aren't in the business of giving explanations or causal analyses but rather give accounts of the "meaning" of cultural institutions and practices.[26] We intend neither to lurch from *Strong Explanationism* to its contrary nor to acquiesce in a tendentious theoretical description of the kinship between illuminating ethnography (of which we take Geertz' account of the Balinese cockfight to be an outstanding example) and the microhistories of Le Roy Ladurie and Davis. The relations between Geertz' account and some notion of "meaning" for cultural items (as well as the relations between Geertz' account and causal analysis) require more extensive treatment than we can offer here. For our purposes, it's enough to identify important historical works that serve as *prima facie* counterexamples to *Strong Explanationism* and to be able to specify the kinds of questions that they address (Of the latter, more shortly).

---

[25] Another famous microhistory, Carlo Ginzburg's *The Cheese and the Worms* (1980), underscores the point we make here. Ginzburg takes us into the world of peasants in the Friuli region of Italy, by focusing on a miller, Menocchio, who was tortured and executed for his heretical beliefs. Ginzburg isn't trying to persuade us that a person previously deemed unimportant has great significance. The aim is to show us how the world appeared to people who normally get left out of histories. The same purpose could have been achieved by concentrating on a different peasant, perhaps Marcato, who came from the same town and was also executed. What distinguishes Menocchio is that we happen to know something about him. But, as the last sentence of *The Cheese and the Worms* tells us (Ginzburg 1980, p. 128), "About this Marcato, or Marco – and so many others like him who lived and died without a trace – we know nothing."

[26] See Berkhofer (1995, pp. 31–33).

The microhistories surely can be described as providing materials for explanation in the sense(s) typically employed in discussions of the aims of history, but to describe them in this way would be to miss their point. The authors give so much detail not so that we can answer a plethora of why-questions (formulated about individuals about whom we have no antecedent interest) but so that something can be evoked in the reader, so that there can be a psychological change through which Pierre Maury, Bertrande, and Menocchio cease to be remote deviant peasants and become fellow humans, who, for all their apparent strangeness, are more like ourselves than we had thought. We don't want to assimilate this evocation to explanation – let alone to make it the central feature of "historical explanation" – but we do want to recognize its importance as a mode of historical knowledge.[27] *Strong Explanationism* seems to have left philosophers the unfortunate choice between denying important historical aims and accomplishments and adopting an implausible view of explanation as the achievement of empathy.

It's important to recognize that the features we've discerned in the microhistories are also present in a much wider spectrum of historical writings. Military history has traditionally been an obvious field in which authors attempt to offer explanations – indeed, critics of *Explanationism* are quite reasonably challenged to account for the large number of works devoted to the origins and resolution of battles and wars. Even in military history, however, we can find historians whose concerns are similar to those of Le Roy Ladurie and Davis. John Keegan's celebrated book *The Face of Battle* has much to tell us about why Agincourt, Waterloo, and the Somme went the ways they did: Keegan provides rich accounts of the outcomes of these three battles. Nevertheless, that is not all – and, we believe, not primarily – what he intended to do. At the end of the first paragraph, Keegan tells us

> . . . I have never been in a battle. And I grow increasingly convinced that I have very little idea of what a battle can be like. (Keegan 1978, p. 13)

The investigation he undertook, on which his book was based, was motivated by his sense of his own ignorance of the very points he felt he should be conveying to his students, all of whom were cadets at Britain's elite Royal Military Academy at Sandhurst; Keegan felt difficulty in answering "what, for a young man training to be a professional soldier, is the central question: what is it like to be in a battle?"[28]

---

[27] There are other distinctive psychological changes that histories might endeavor to induce. Sometimes historians attempt to provoke a moral reaction by explaining how some contemporary institution has been deliberately designed to exclude a particular class of individuals or to detract from their welfare; a prime example is Mike Davis *City of Quartz* (1990), which shows how various aspects of Los Angeles were set up to make life hard for the indigent. Of course, there's a long tradition of histories "to a moral purpose," as well as an extensive critique of their propriety – perhaps most famously encapsulated in Ranke's dictum; we won't attempt to resolve the thorny issues here.

[28] Keegan (1978, p. 16); it's entertaining to think that, had he read Thomas Nagel's famous essay, Keegan might have entitled his book *What Is it Like to be in Battle?* Other military historians have approached the same question, particularly in the case of the First World War; see, for example, MacDonald (1978, 1980, 1983) and Ellis (1976).

One way to approach that question is to describe the circumstances of the individual participant in major historical battles in ways that enable readers to relate the soldiers' predicaments to their own experiences – to describe in some detail the gear that would have been worn, the equipment carried, the ways in which various types of encounters would have gone, the sounds that would have been heard, the limitations on visibility, the effects of incidents that occurred, and so forth. Here is Keegan's description of the crucial failure of a French charge at Waterloo.

> The men at the front could see their officers, see the enemy, form some rational estimate of the danger they were in and of what they ought to do about it. The men in the middle and the rear could see nothing of the battle but the debris of earlier attacks which had failed – discarded weapons and the bodies of the dead and wounded lying on the ground, perhaps under their very feet. From the front came back to them sudden crashes of musketry, eddies of smoke, unidentifiable shouts and, most important, tremors of movement, edging them rearward and forcing them, crowd-like, in upon each other. (Keegan 1978, p. 174)

This passage contributes to two quite different historical projects. Keegan is interested in explaining a particular incident, late in the day at Waterloo, when the Imperial Guard, charging the British position, was met with sudden and unexpected fire and, as was recorded by soldiers on both sides, those in the center and rear of the columns (soldiers who were in less danger than those in front of them) turned and retreated. He also wants to convey to readers who have never had military experience what it was like to advance in column at the end of a long battle, and he does so by connecting the predicament of the soldiers who fled to experiences most of us have had. We may not know exactly, or even approximately, what it was like to be a French soldier in that charge, but, because we have been in crowds that suddenly pitched us in unanticipated directions, Keegan's description provides us with a much better appreciation than we would otherwise have had.

It would be easy to multiply examples. Keith Thomas' (1970) magisterial study of the ways in which religion, magic, and the emerging science catered to a broad variety of human needs during the sixteenth and seventeenth centuries could be viewed as explaining the trend indicated in his title, the "Decline of Magic." But Thomas is concerned to display for us the rich variety of magical practices and the diversity of ways in which the church attempted to assimilate them. Paul Cohen's (1999) illuminating approach to the Boxer Rebellion, *History in Three Keys*, explicitly commits itself to a difference among three styles of approaching the same events, the first offering an explanation of what occurred, the second attempting to reconstruct the world of the Boxers, and the third examining the various ways in which the Rebellion has been interpreted to illuminate later political programs. It's small wonder that many historians have resisted analytical philosophy of history, feeling that the complex texts they most admire are somehow reduced or eviscerated by philosophical analyses. Moreover, as we'll argue in a moment, they've been right to object to the model of scientific explanation that has almost invariably been wheeled out when philosophers try to identify the links between history and the sciences. But the rot goes deeper.

The trouble lies with *Strong Explanationism*. In terms of the views of explanation typically presupposed in discussions of history, the types of studies just reviewed count as a decisive refutation of *Strong Explanationism.* As we'll now argue, there

are excellent reasons for abandoning those approaches to explanation, both for history and for the sciences. When we do so, the centrality of explanation to history will appear in a new light.

14.4 Liberalism about Historical Explanation

Claims about explanation in history can be read in several ways, of which we'll distinguish three:

(1) *The Strict Interpretation*. Explanations are arguments in which general laws figure in the premises.
(2) *The Orthodox Interpretation*. Explanations are answers to why-questions.
(3) *The Liberal Interpretation*. Explanations are answers to questions of many different types (how-questions, what-questions, when-questions, and so forth, as well as why-questions).

The most vigorous program for assimilating history to the natural sciences, the Hempelian program, attempts to defend *Strong Explanationism* under the *Strict Interpretation*. Sensitive readers of historical texts notice that those texts rarely succeed in giving the right kinds of arguments and appear to contain a lot of interesting material of a different kind. So they reject the *Strict Interpretation* in favor of the *Orthodox Interpretation*, contending that historians have a special way of answering why-questions, which proceed through the construction of narratives, and that the important philosophical project is to understand the structure – or logic – of narratives. We agree that historians sometimes construct narratives and that these narratives answer certain kinds of questions (whether they are best construed as why-questions are a topic we'll take up below). But we think *Strong Explanationism* is doomed even on the *Orthodox Interpretation*. Friends of the *Orthodox Interpretation* sometimes recognize that the achievement of some type of empathetic understanding of historical actors is a goal of many historical works – but they distort the point by relentlessly insisting that this has something to do with answering why-questions. Our claim is that recreating past experience, enabling a modern reader to have access to a past world, and answering questions of the form "What was it like to be ...?" are valuable quite independently of answering any why-question. Paul Cohen's separation of his first two approaches to the Boxer Rebellion is exemplary in this respect.

Adopting the *Liberal Interpretation* would appear to attenuate the connection between history and the natural sciences, and, indeed, that would be so if one continued to insist on the *Strict Interpretation* or the *Orthodox Interpretation* in reading *Bernardism about the sciences*. We propose, however, to adopt a *Liberal Interpretation* consistently, for explanation in the sciences as well as for

explanation in history. As we'll emphasize below, the natural sciences aim at lots of different things, and the questions they answer are heterogeneous. So the possibility of kinship between the aims and achievements of history and the aims and achievements of various areas of natural science is not foreclosed by our rejection of *Strong Explanationism* when construed by way either of the *Strict* or the *Orthodox* interpretation.

Although we believe that the rejection of the covering-law model of explanation, with the shift to the *Orthodox Interpretation* and the advocacy of narrative as a mode of historical explanation, doesn't go far enough, we agree that it was correct to abandon Hempel's account. Indeed, the covering-law model has been under severe attack *as a model of scientific explanation* for three decades or more, and it will be helpful for later discussions to examine why its fortunes have waned.[29] Hempel's lucid analysis of explanation encountered four major difficulties, two of which are pertinent to its failure with respect to history.[30] The two on which we'll focus are the insufficiency of Hempel's conditions on explanation and the incompleteness of his discussions of how explanations relate to context.

Setting aside details that are irrelevant for our purposes, Hempel is committed to the view that any deductively valid argument among whose premises is a general law explains its conclusion. It's not hard to think of many different counterexamples, but two general types are especially forceful. *Asymmetries* of explanation arise when there is a pair of arguments differing only in the fact that each is obtained by switching a premise and a conclusion in the other, where each set of premises contains a general law and of which one, but not the other, strikes us as explanatory. So, to cite a standard example, one can explain why a flagpole casts a shadow of a certain length by appeal to the height, the elevation of the sun, and the law that light travels in straight lines; but, although one can derive the height of the flagpole from the length of the shadow, the elevation of the sun, and the principle of rectilinear propagation of light, that derivation is not explanatory.[31] *Irrelevancies* in explanation result from the possibility of stating general laws that don't identify a factor that is explanatorily crucial. Here, a standard example concerns a man who takes birth control pills; even though it's a matter of scientific law that ingesting those pills prevents pregnancy, we can't explain the man's failure to become pregnant by appealing to the law and the fact of his peculiar diet.[32]

In the context of historical explanation, the same general problem was identified by J.H. Hexter who saw that Hempel's model permitted trivial derivations of no

[29]For more detail on this issue, see the opening sections of Kitcher (1989) and the later parts of Salmon (1989).

[30]The two problems we won't consider are the intractability of the problem of specifying the notion of scientific law and the counterintuitive consequences of Hempel's model of probabilistic explanation. Both troubles are presented very clearly in Salmon (1989).

[31]The example was originally devised by Sylvain Bromberger in the early 1960s.

[32]This example was introduced by Salmon in "Statistical Explanation and Statistical Relevance"; Salmon notes that an earlier example of the same type was formulated by Henry Kyburg.

explanatory value. Hexter noted that one can't explain the presence of the Giants in the 1951 World Series by deducing the conclusion "The Giants played in the 1951 World Series" from premises asserting that the Giants won more games than any other National League Team and that whenever a National League Team wins more games than any other National League Team, it goes to the World Series.[33] Now, one might quibble about whether the generalization he cites should count as a general law (after all, it refers to a particular social entity, the National League), but it wouldn't be hard to develop Hexter's example to avoid any such objection.[34] Indeed, once one appreciates the problems posed by explanatory asymmetry and irrelevance, it becomes clear that nonexplanatory derivations that fit the Hempelian model are legion. Hexter (1971, p. 31) saw this very clearly – he concludes that his Hempelian argument about the Giants' victory doesn't "tell the questioner what he wants to know" – and, insightfully, he goes further, claiming that the philosophical discussion of history has been distorted: "... the notion that the sole appropriate response of the historian to his commitment to communicate what he knows is something designated "explanation" is wildly arbitrary."[35]

Given the deficiencies of the covering-law model in stating sufficient conditions on scientific – or historical – explanation, philosophers have sought to isolate what is problematic about the nonexplanatory derivations. Although no current model of explanation enjoys the widespread acceptance that Hempel's account once had, the most popular suggestion has been that explanations have to identify causally relevant factors.[36] Invoking the notion of causation was anathema to Hempel and his colleagues, for whom much of the point of an analysis of explanation consisted in demonstrating that one didn't need to appeal to any (suspect) causal concept.[37] For our purposes, however, questions about whether causal concepts need analysis (and, if so, how the analysis is to be given) are secondary; once the specification of causes is seen as crucial to scientific explanation, there seems to be a much more straightforward connection between history and the natural sciences. Causal explanation is common to human affairs, to evolutionary and developmental biology, to geological studies that trace the emergence of mountain ranges and other

[33] See Hexter (1971, p. 30). Hexter formulates the point a bit more carefully than we've done here.

[34] Here's a recipe for doing so: One can give a Hempelian argument for a conclusion asserting that the Giants won any specific victory (say their 37th) by using physical laws to derive the trajectory of the winning hit; now, add the true conditional statement that if the Giants won that game, they would win the pennant.

[35] Hexter (1971, p. 29); see also 71, where Hexter adopts what we've called the *Liberal Interpretation.*

[36] See, for example, Salmon (1984) and Humphreys (1990).

[37] Logical empiricism was mindful of Humean strictures about causation. Some of those who have identified the need for a causal constraint – Salmon, for example – have accepted the thought that invocation of an *unexplicated* notion of causation is illegitimate. Interestingly, the approach that Salmon has adopted, which sees causation in terms of the transmission of conserved quantities, seems very hard to apply in the context of historical explanation.

large topographical features, and to cosmological investigations of the formation of atoms, nebulae, stars, and planets.[38]

At this stage, it's useful to take up the second major difficulty with the Hempelian approach to explanation, the lack of any detailed account of how explanations are responsive to contextual variables. Hexter's example of the Giants' success in 1951 is pertinent here – we might imagine *some* contexts in which the Hempelian argument serves to explain to someone why the Giants went to the World Series (consider someone who is just very ignorant about this kind of competition, for whom it's a genuine option that you might go on to the final phase if you won more than a particular percentage of the games, or defeated a particular opponent, or scored most runs); but the contexts that readily come to mind are ones in which the Hempelian account doesn't provide what a person asking the explanation-seeking question wants to know. Further, it isn't obvious that insisting that genuine explanations specify causes helps to resolve the trouble, for one might argue, with some plausibility, that Hexter's derivation actually satisfies the causal constraint. The trouble, quite evidently, is that there are causes and causes; some are remote and some are very close to their effects; some strike us as unimportant or uninteresting; others are salient. The essential context dependency of specifying causes was brought out very clearly by N. R. Hanson, and we'll amend a famous example of his.[39]

Why did the Princess of Wales die? We don't know the details, but there was surely a moment shortly before the fatal crash at which the wheels of the car were set on a trajectory that was inevitably going to lead the vehicle into a high-velocity collision with unyielding concrete. So there's some mechanical story that specifies an event that caused the crash and another mechanical-physiological story that specifies the damage produced in Lady Diana's body. Imagine that you are given these accounts in any amount of detail. Have you been offered an answer to the question?

We think not – at least not if the context in which the request for explanation was posed was relatively normal. We can envisage accounts at many different levels of analysis, some that appeal to blood-alcohol levels and unfastened seat belts, others that focus on the paparazzi and their intrusions into Diana's life, yet others that

[38]It should be noted that causal-historical explanation is prominent in some areas of the sciences (like the ones we've listed); that a different type of causal explanation (causal-mechanical explanation) is widespread in others, as, for example, in biochemistry and solid-state physics; and that there are some parts of theoretical science in which it's something of a strain to think in causal terms (the theory of the chemical bond, sex ratio theory). See Salmon (1998).

[39]See Hanson (1958, p. 54). We should note that the context dependency of explanation has been thoroughly analyzed by Bas van Fraassen (1980, Chap. 5) who offers a pragmatic theory of explanation. One of us has criticized van Fraassen's theory on the grounds that it trivializes the notion of explanation (Kitcher and Salmon 1987), but the objection would now be modified; as we'll argue below, what counts as the right sort of causal relation to invoke in answering an explanation-seeking why-question is contextually determined, and van Fraassen was insightful in pointing this out.

concentrate on her unhappy marriage and the attitude of the Windsor family. No one of these will answer to every normal context of requesting explanation, although for each, there's a range of mundane contexts in which it would be appropriate. It's not enough, then, to replace Hempel's covering-law model with the suggestion that explanations specify causes. The right sorts of causes must be picked out, and they must be given their due – and what "right" and "due" mean depends on the context in which the why-question is posed.

Hexter's discussion of the 1951 World Series comes close to making this point. He presents a graph, showing the number of games by which the Giants trailed the Dodgers from August 13, when they were thirteen games behind, to the dead heat at the end of the season, the three-game play-off, with the third game run deficit inning by inning (with, of course, the dramatic Bobby Thomson home run represented by a final upward spike).[40] Hexter suggests that we can use the graph to understand why some proposed explanations succeed, proposing, in effect, that it's the points of sudden change that mark the places at which causes are especially to be sought. In our judgment, this isn't quite right: we can envisage circumstances under which it would be precisely the points at which the Giants maintained ground (or didn't lose too much) that corresponded to the important causal foci – imagine that August and September 1951 were marked by outbreaks of intestinal flu that laid many baseball players low and that the Giants held their own even when barely able to field a team. So we draw a somewhat different conclusion: historical explanations seek particular kinds of causal information, and there's no context-independent way to specify the types of causal information that are salient.

But we don't believe that matters are any different when one turns to the natural sciences, particularly to those sciences whose modes of explanation are closest to history. Consider the process that begins with the fertilization of an egg and culminates in a mature organism. The causal history behind the presence of a particular trait can have the same complexity as that behind the death of Princess Diana – perhaps there was a particular allelic combination that gave rise to a protein that might have been modified in the presence of a cytoplasmic constituent that wasn't available, and the subsequent receipt of molecules from the ambient environment triggered an increase in the rate of cell division in a specific developmental field, and so on and so forth in a cascade of effects. Just as we could devise in the case of the car crash (or in the case of the Giants' success) any number of causal stories that focus on factors inapposite in any normal context, so too with the embryological example; by analogy with the causal-mechanical account of the fatal collision, we can select some late developmental stage and show that available intracellular energy doesn't suffice to break the bonds of appropriately chosen constituent molecules. Moreover, as we imagined a variety of narratives that emphasized different factors – the alcohol, the paparazzi, the

[40] Hexter (1971, p. 35), Fig. 1. We are grateful to David Sidorsky for pointing out to us that Hexter's account does not mention the controversy about whether the Giants were stealing the Dodgers' signs.

Windsors' disapproval – so too we might concentrate on the organism's genotype, on the details of maternal inheritance that led to a missing cytoplasmic constituent, the signals from the environment, or the increased rate of cell division.

Our examination of *Strong Explanationism* thus leads us to two conclusions, one that militates against the assimilation of history to the natural sciences and one that favors the assimilationist program. The negative point is that, on both the *Strict Interpretation* and the *Orthodox Interpretation*, *Strong Explanationism* must be rejected; as Hexter saw, historians are not simply in the business of giving explanations, conceived as answers to why-questions (or how-did-it-come-about-questions).[41] The positive point is that Hempel's account of explanation for the natural sciences must give way to a much looser causal-contextual view, a view that allows for affinities between scientific explanations and historical explanations. It looks, then, as though part of what historians aim at and achieve might prove similar to what (some) natural scientists aim at and achieve and that the extent of the similarity might vary quite widely depending on which historians and which scientists we pick out.

We think that this conclusion is along the right lines, but that it needs refinement. We'll try to improve it by taking up some of the other theses we promised we'd scrutinize.

14.5 History and Truth

A different way of specifying the aims and achievements of the sciences is to invoke the idea of the pursuit of truth and advocate *Veritism about the sciences*. Just as there are different ways of interpreting *Strong Explanationism*, depending on the concept of explanation chosen, so too with *Veritism* and the notion of truth. We'll approach the issues by adopting a relatively modest version of the correspondence theory of truth.[42] We hold that there's a relation of reference between the singular terms of our language and mind-independent entities and between the predicates of our language and sets of mind-independent entities and that a sentence is true by virtue of corresponding to the way the world is just in case the entities referred to by its singular terms stand in the right relationship to the sets referred to by its constituent predicates – where the right relationships are those characterized by Tarski.[43] There are influential arguments to the effect that, on this interpretation,

[41] Hexter (1971, p. 30) rightly appreciates the greater naturalness of "How did it come about that . . . ?" rather than "Why . . . ?" in historical studies.

[42] The modesty comes in two ways. First, we don't suppose that there are special entities – *facts* – to which true sentences correspond. Second, we don't assume that the core notion of reference can be specified in a physicalist vocabulary (as, e.g., Field (1972) proposes). For further exploration of the position, see Kitcher (2002).

[43] See "On the Concept of Truth for Formalized Languages" (Tarski 1956) or any presentation of the semantics for first-order logic in a logic text. Effectively, our proposal adds to Tarski's well-known

Veritism about the sciences can't be sustained – or that it can only be upheld for certain kinds of scientific claims (those that are concerned only with observable entities).[44] Since one of us has argued at some length that *Veritism* can be defended against these challenges, we'll simply take *Veritism about the sciences* (on our modest correspondentist interpretation) for granted in what follows.[45]

Some historians and philosophers of history have resisted *Veritism about history*, at least when that thesis is articulated via a correspondence approach to truth. We want to start by identifying some *Veritist* themes that are quite innocuous.

In his celebrated *The Age of Constantine the Great*, the nineteenth-century historian Jacob Burckhardt tells his readers about the birth of the future emperor:

> ... [the Alemanni] were defeated at Windisch by the General Constantius Chlorus under Aurelian (274), and indeed on the same day that his son Constantine was born. (Burckhardt 1949, pp. 71–72)

Even though this is the second clause of a two-part sentence, it is quite complex. One way to expose its logical structure would be as follows: there is an event e, such that e is in 274 and e is a battle and e is a defeat of the Alemanni by Constantinus Chlorus and Constantinus Chlorus is at the time of e a general under Aurelian and e is at Windisch and on the same day as e, there is an event f which is a birth and a son is born to Constantinus Chlorus in f and that son is Constantine. It's easy to recognize that there are plenty of ways in which Burckhardt's claim might turn out to be false. Indeed, we'd agree with the judgment that *certainty* about any conjunction like this involving happenings in the distant past is too much to hope for. *Veritism*, however, isn't about certainty but about truth. We judge that Burckhardt aimed to tell the truth, in the modest correspondence sense, that he assembled evidence to this end and that, given the evidence, there's good reason to think he attained it. To consider the terms that figure in our reconstruction of the sentence, there are singular terms ("Alemanni," "Windisch," "Constantine," and so forth) and predicates ("is a battle," "is on the same day as," "is a birth," "is a general under"). According to the modest correspondence theory, the singular terms refer to entities that are independent of the psychological life of Burckhardt or his contemporary reader, to a tribe, a place, and a person; similarly, the predicates have in their extensions events, ordered pairs of events, events, and ordered pairs of people, respectively. There is nothing obscure or metaphysically dubious in this account of the truth of Burckhardt's sentence. Nor is it mysterious how a chain of informants might provide evidence for each of the constituent claims. There are, of course, interesting issues about how historians should satisfy themselves

account only the idea that the reference relation connects linguistic items with mind-independent entities.

[44] For a general skepticism about *Veritism* as we've interpreted it, see Rorty (1982), Putnam (1981), and Goodman (1978); more local versions are advanced by van Fraassen (1980) and Laudan (1984). See also Fine (1986).

[45] For the defense, see Kitcher (1993a, Chap. 5), Kitcher (2001b, Chap. 2), and especially Kitcher (2001a).

that their conclusions are backed by a reliable sequence of informants – it might even turn out that the resolution of those issues might raise suspicions about some part of Burckhardt's judgment – but the general possibility of finding out (say) that Constantinus Chlorus defeated the Alemanni at Windisch shouldn't be dismissed. Hence, as long as we focus on sentences like the one we've quoted, *Veritism about history* seems unproblematic.

Where then does trouble come in? Although Burckhardt's sentence is logically complex, it has a certain type of *conceptual transparency*. What we mean by this is that the language, particularly the predicates, it contains doesn't seem to embody either a classificatory scheme that might easily be rejected or a categorization that depends on subjective judgment. It's possible that our descendants might reject such categories as *battle*, or *being on the same day as*, but the possibilities seem too remote to buttress a charge that the historian can't aim at or achieve truth because the classificatory scheme presupposed in the representation of historical events is always laden with the values and prejudices of the writer's time and circumstances. We can bring out the contrast by considering the account that one of Burckhardt's predecessors gives of the character of the Empress Theodora. Gibbon's description of her is full of references to acts of "prostitution" (a category that covers both her alleged affairs and her public performances on the stage) and her "licentiousness" (which, if we ignore the real possibility of Gibbonian irony, might be viewed as expressing a moralistic disapproval of female sexual desire). Here is a relatively short sentence:

> Her chastity, from the moment of her union with Justinian, is founded on the silence of her implacable enemies; and although the daughter of Acacius might be satiated with love, yet some applause is due to the firmness of a mind which could sacrifice pleasure and habit to the stronger sense either of duty or interest.[46]

We imagine opponents of *Veritism* protesting that Gibbon's sentence isn't true, that it embodies categories that he was entitled to use but that we are entitled to reject, and that similar infection permeates all historical writing (The infection is just much harder to recognize in a sentence like the one previously quoted from Burckhardt).

We agree that no contemporary historian should be tempted to use Gibbon's sentence (despite its elegance) in a description of Theodora – it would be right to say that some of his words are not ours.[47] That, however, shouldn't be confused with the issue at hand, the truth of Gibbon's claim. Here, it may help to consider parallel examples in the sciences. Gibbon's rough contemporaries Joseph Priestley and Georges Cuvier used terminology we'd reject: Priestley identified the properties of a gas he called "dephlogisticated air," noting that it supports combustion and

[46]For Gibbon's description of Theodora, see Chap. XL, part 1, of *The Decline and Fall of the Roman Empire*. The quoted sentence is from p.56 of volume 5 of the Oxford English Classics edition (1827).

[47]Famously, Oscar Wilde replied to the prosecutor who asked if his works constituted blasphemy, "That is not one of my words."

respiration better than ordinary air, and Cuvier presented his admiring audiences with new fossil species (assuming a fixed, monotypical, notion of species). We reject the language they employed, abandoning Priestley's term "dephlogisticated air" and attaching a different concept to Cuvier's "species," but this doesn't prevent us from recognizing that Priestley says true things about oxygen (the gas he sometimes refers to using "dephlogisticated air") and that Cuvier correctly separates different fossil species.[48] At least part of Gibbon's sentence can be retrieved in a similar way. When he talks of Theodora's "chastity" after her marriage to Justinian, we recognize that he means to refer to her sexual fidelity. Thus, the first part of his sentence might be recast as the claim that after marrying Justinian, Theodora didn't have sexual relations with anyone else, together with the suggestion that the lack of rumors about her behavior (in a context in which she had many detractors) serves as evidence for this. Once this has been done, Gibbon's claim seems no more problematic than Burckhardt's.

Let's now look at an alternative way in which conceptual transparency might fail. Consider a passage we've already quoted from Desmond Seward's *The Hundred Years War*:

> Joan of Arc merely checked the English advance by reviving Dauphinist morale, and the Regent managed to halt the counter-offensive. It was not the Maid who ended English rule in France.

One might worry that this claim presupposes a subjective interpretation of how causal categories are to be applied, that Seward has focused only on the relative short-term consequences of Joan's actions and failed to appreciate her influence on events that took place after her death. In articulating his view, he notes that Joan's initial successes (the relief of Orléans, the march through English-Burgundian territory to Rheims, and the coronation of the Dauphin) were followed by a period in which Bedford, the Regent, had a number of victories, a period that ended with Joan's capture, trial, and execution. Seward thus emphasizes the fact that something more was needed to drive the English out of France, something beyond the revived morale of the French – he points to the ineptness of Cardinal Beaufort's military policy (especially after Bedford's death), the Franco-Burgundian alliance, and the emergence of improvements in artillery technology (particularly associated with Maître Jean Bureau).[49] Enthusiasts for Joan (and for traditional celebrations of her) might suggest that her influence was decisive – without her, there would have been no possibility of driving the English out. The sophisticated historian, reviewing this clash of judgments, may declare that there's no fact of the matter. History is just indeterminate as to whether Joan ended English rule in France.

We agree that there are several different ways of elaborating such causal notions as "ending English rule," but we believe that, once the meanings have been fixed, it's possible to talk about the objective truth (or falsehood) of historical statements.

---

[48] For detailed defense of the claim about Priestley, see Kitcher (1978).

[49] Seward (1999, pp. 221–262).

To a first approximation, the traditionalist insists that, without Joan's intervention, the English would have continued to dominate Northern France (and Guyenne): if we imagine a world very like the actual one, in which Joan doesn't intervene (she doesn't hear the voices, or she is turned away from the Dauphin's camp), then the English presence remains. Likewise, a rough way to gloss Seward's claim is that many continuations of the course of events at Joan's death would have led to the preservation of English rule: in worlds like the actual one in which Beaufort is less powerful (or more clear-headed) or in which Burgundy stays allied to England or Maître Bureau doesn't achieve his technological advances, the English don't get driven out. Historians are sometimes suspicious of counterfactual claims, but we concur with those authors who believe that counterfactual explorations are embedded in historical practice.[50] We don't believe that it's easy to give a full theory of historical counterfactuals that will reveal how they are objectively true and false,[51] but we can defuse the argument for maintaining that counterfactual judgments must be subjective that appeals to such clashes as that between Seward and the traditionalist. For in this, and kindred, cases, the disambiguation of the causal claims exposes the fact that *both* might be correct, we see no difficulty in supposing that Joan-less worlds would have seen continued English domination and that worlds with Joan but without (say) Bureau would have unfolded to the same end.

We offer a further consideration against the worry that counterfactual claims are mere flights of the historian's fancy. In some instances, the counterfactual judgment mirrors the decision-making of a historical protagonist: Joan was moved to go to the Dauphin because she thought that her intervention was needed to save Orléans; Maître Bureau worked with his brother on improving artillery because he thought it would make a difference to the French success. Setting aside the heavenly voices of the one and the commercial interests of the other, we can endorse the idea that *both* had a clear understanding of the possible futures. When historical agents consider their options, they may sometimes be myopic or deceived, but, where we retrospectively find no basis for impugning their judgments, we'd expect their most central decisions to involve suppositions and counterfactuals that are objectively correct.[52]

We've been arguing for a particular elaboration of *Veritism about history*, and it will be worth presenting it explicitly.

> *Veritism about historical statements*. History aims at, and sometimes achieves, true statements about some aspects of the past, even when the statements in question may be couched in categories that later historians might reject or when those statements contain causal concepts.

[50] See the Introduction to Ferguson (2001, especially 87) and also Hawthorn (1991).

[51] The most prominent philosophical account is that of David Lewis (1974), which deploys a notion of similarity across possible worlds. An obvious worry is that similarity depends on a choice of respects and degrees and that such judgments are irremediably subjective.

[52] For a similar assessment, see Niall Ferguson's Introduction to his *Virtual History*.

Even if (as we hope) we've been successful in defending this thesis, it may seem beside the point. For although aiming at true statements might be *part* of the historian's enterprise – that's what accounts for the long hours in the archives – there's plenty of room to doubt that it's the whole or even a major part. If that were all there were to doing history properly, then history would be easy: all one would have to do would be to find some hitherto unworked piece of archival material – the journal of a nineteenth-century Shropshire pig-farmer, say – establish the reliability of the source, and then proceed to regale the learned (and maybe unlearned) world with true statements about the past ("On April 18 1836, there were intermittent showers on Wenlock Edge ..."). Opponents of *Veritism* probably have little patience with our efforts to support *Veritism about historical statements*, because they consider the historian's task as one of producing *histories*, and although histories contain statements (and although historians want those constituent statements to be true), the collection of statements doesn't exhaust the history. Indeed, opponents will continue by suggesting that the interesting thesis of *Veritism about history* is the claim that the aim of history is to produce *true histories* – *Veritism about historical statements* is a necessary condition for that interesting thesis, but it falls far short of being sufficient.

This objection contains several important insights, which deserve careful articulation. We'll begin, however, with a cautionary point. Truth is primarily an attribute of sentences or statements; there may be derivative notions of truth that apply to thoughts or to visual representations. Any notion of truth that is supposed to apply to a complex of statements – a historical work, a narrative, or a history – must be carefully explained in terms of the core notion of truth, that is, truth as a property of individual representations (paradigmatically statements). Casual invocation of truth for complex texts (histories, narratives) and direct denials that such texts are true are both misguided.[53] We need first to understand how a concept of truth might be supposed to apply here.

At this point, it will be useful to explore a parallel source of confusion in the philosophy of science. One influential line of argument against realist approaches to the aims and achievements of the sciences, the "pessimistic induction on the history of science," begins from the judgment that past science is full of theories that once appeared extremely successful and which we now reject. The conclusion we're invited to draw is that none of our current theories, however successful they may appear, is true. Indeed, to suggest that we ever achieve true theories may be a serious deception; and, if true theories inevitably lie beyond our reach, true theorizing can't be our aim.[54]

But what does it mean to say that a theory is true? According to a once popular notion of scientific theory, there's an easy answer: a scientific theory is a collection

[53]The philosopher of history who is clearest on this point is Ankersmit (1983).

[54]The most fully developed version of this argument appears in Laudan (1981); this essay is essentially reprinted as Chap. 5 of his *Science and Values* (1984).

of statements, consisting of a set of principles and their deductive consequences; the theory is true just in case the conjunction of the principles (the axioms of the theory) is true. Now, although philosophers have reconstructed a few parts of science in this way – most notably in theoretical physics – it has become increasingly evident that there are vast areas of the natural sciences that the axiomatic conception of scientific theories fits badly, if at all.[55] Even where it is applicable, however, the axiomatic conception and its coordinate notion of truth lead to an interesting reappraisal of the "pessimistic induction." To show that a successful theory is false, all that is needed is to find some false constituent statement – one fault infects the whole. Realists can thus reply that all that has been shown is that the theories we've so far developed aren't *completely* true. More exactly, they can defend *Veritism about scientific statements* – the sciences aim at and sometimes achieve true statements.

And here, of course, the impatient anti-Veritist protests. To say that the sciences aim at true statements is far too weak. Truth is cheap. Without large government grants (or private funds), you can discover vast numbers of truths about nature: Look around! There are indefinitely many languages you could use to announce indefinitely many truths about the immediate vicinity. If *Veritism about the sciences* simply retreats from the claim about seeking true theories, maintaining *Veritism about scientific statements* instead, then it has trivialized the scientific enterprise.

The negative point is correct. The bare substitution of *Veritism about scientific statements* is inadequate. But it's a mistake to think that the old idiom of "true theories" was satisfactory or to suppose that theories are the be-all and end-all of good science. Instead, we propose to adopt

> *Veritism about significant scientific statements*. The sciences aim at, and sometimes achieve, significant true statements about aspects of nature.

Similarly, we maintain

> *Veritism about significant historical statements*. History aims at, and sometimes achieves, significant true statements about aspects of the past.

Neither of our theses is worth much, of course, until we've said something about the notion(s) of significance involved.

---

[55] Some challenges have developed the "semantic conception of theories," according to which theories are families of models; for an accessible presentation, see Giere (1988, Chaps. 3 and 4). Others have emphasized the apparently non-axiomatic structure of evolutionary biology, molecular biology, the geological sciences, and so forth; see Kitcher (1993a, Chaps. 2 and 3). Another source of trouble emerges from the powerful account of "normal science" offered in the early chapters of Kuhn's *The Structure of Scientific Revolutions*.

For the case of the sciences, we summarize an approach one of us has developed elsewhere.[56] Significant statements are answers to significant questions. To say that a question is significant is not to say that it's posed to us by nature[57], but that *people in a particular context* find it to be worth addressing. There are general sorts of considerations that make a question worth addressing. Sometimes we need to know how to predict an outcome in order to achieve our ends; sometimes we need to know something in order to intervene successfully in nature; sometimes we are simply curious about some aspect of the natural world. Here is the point at which *Bernardism* and *Veritism* connect. When the connection is made, however, it's important not to suppose that all instances of disinterested curiosity – all cases in which the significance of a question is epistemic rather than practical – involve why-questions. Both in the sciences and in history, there are many different kinds of questions to which we'd like answers.

Consider the following sample: "What are the constituents of eukaryote cells?" "Will the universe continue to expand indefinitely?" "Is there intelligent life elsewhere in the universe?" "When did human language evolve?" "How many species of australopithecines were there, and how are they related?" "To what extent can one form a range of silicon compounds that rivals the diversity of carbon compounds?" "What is the natural host organism for the Ebola virus?" "Can nonhuman animals count?" We suggest that these questions are significant, that they are significant independently of any practical use we might make of answers to them,[58] that we aren't interested in them because answers would constitute a law or a theory,[59] and that none of them is naturally reformulated as a why-question. Scientific significance is much more heterogeneous and messy than traditional philosophical accounts have recognized. We can defend *Bernardism* only if we're prepared to view explanations as answers to significant questions, which fall into a wide variety of types – in short, only if we're prepared to adopt the *Liberal Interpretation*.

Historians, too, are concerned with a wide range of questions. We've already noted that the point of some historical works is to answer questions of the form "What was it like to be . . . ?" Yet, even in the case of historical texts that might seem to be directed towards causal explanation of some outcome, it would be wrong to insist that a single why-question is the focal point. It's tempting to think that a history of the Hundred Years War is effectively an explanation of why the English

[56] Kitcher (2001b, Chap. 6).

[57] Here, it seems to me that Rorty's skepticism about nature's agenda is insightful; see the introduction to Rorty (1982).

[58] One might worry that the issue of the natural host for the Ebola virus is a practical question. Indeed, knowing the answer might enable us to prevent future outbreaks of Ebola. Nevertheless, even if we had a surefire vaccine for this disease – and were thus unconcerned about passage of the virus to human populations – we'd still be interested in knowing where the virus originally came from.

[59] In some instances, of course, we might achieve an answer by developing a general theory; but even in such cases, we'd be interested in the answer whether it came as a consequence of theory or not.

were driven out of France, but the primary concern is surely to inform the reader about what happened in a particular region at a particular time. Many histories are far more interested in the route than in the terminus. A striking example is Robert Hughes' brilliant evocation of "the system," that is, the settling of Australia by convicts (and the law enforcement officers who disciplined them).[60] It would be a travesty to confine Hughes' account to a single why-question, or even a small set of why-questions, for example, "Why was the system abandoned?" His rich treatment answers a wide range of questions: "Who were these convicts?" "What were their lives like?" "What opportunities were available for them?" "How harshly were they disciplined in Australia?" "How did they graduate from the system?" We're given what many histories provide, a picture of a place and a social group during a particular period, that is, a "portrait of an age."

At this point, we can return to the impatient critic and to the insight that there's more to a history than a collection of statements. What the critic sees is that a serious historical work *structures* the constituent statements and that the criterion for good structuring is not one of correspondence to reality. A familiar way of developing the point is to refer to the structure as a "narrative" and then to debate whether narratives are answers to why-questions or whether they should be understood by deploying categories from literary criticism (or theory).[61] Precisely because we understand histories as answering a range of significant questions, we adopt the neutral term "structure." What makes for a good structure, we suggest, is the provision of answers to significant questions. Thus, an unstructured list of true statements fails as history because it doesn't answer any significant question (beyond whatever significant questions would have been answered by the constituent statements). In the work of a gifted historian, however, the *combination* of the statements provides answers to a much broader set of significant questions. Thus, for example, the individual details of Hughes' *The Fatal Shore* compose into a portrait of early Australian life.

There are forms of historical structure that are very close to, even identical with, structures that inform scientific works. One kind of significant scientific question concerns the ways in which aspects of nature have come to be as they are – thus, there are important works of science that develop accounts of how the universe began, how it evolved, how the earth's surface came to be as it is, how the continents reached their present positions, how life evolved, how hominids originated and radiated, and how different human groups became differentially successful.[62] Equally, for groups of human beings, both large and small, there are historical texts that tell structurally similar stories, histories of particular nations, or

[60] Hughes (1986).

[61] For this debate, see Roberts (2001). The approach to historical narratives in literary terms was pioneered by Hayden White in *Metahistory* (1973). We'll briefly discuss the approaches of Ankersmit and White in the final section below.

[62] For the first and last, see Weinberg (1977) and Diamond (1998), respectively. There are vast numbers of books on the history of life and on hominid evolution.

institutions, or local communities. Second, as we've already noted, some scientific investigations are concerned to identify the causes of complicated phenomena – to explain, for example, the distribution of the biota in a particular region of the globe – and their proposals are structurally similar to those offered by some historians, interested in such things as the causes of the First World War or the schism between Catholicism and the Orthodox Church.

Yet historians, particularly the most creative of them, raise new kinds of significant question. They expose features of our own lives which we overlook or take for granted, by showing how people lived when those features were absent. They illustrate possibilities we had not considered by taking us into past societies or situations. The historian's selection of true statements about the past may bring into new relations things with which we are familiar or expand the world of our mundane experience. Tapping into our curiosity about the character of our own lives and the possibilities of human experience, they may make us interested in people or periods that we had not previously seen as significant, Bertrande de Rols or the late eighteenth century in Botany Bay.

So, while our pair of *Veritist* theses bring out the commonality between history and the sciences, we think it right to emphasize the ways in which historians can generate new significant questions by drawing on our curiosity about the possible forms of human life. Once again, we reach a mixed conclusion. History shares with the sciences the aim of reaching significant truth. It differs from the natural sciences in having special opportunities for generating new significant questions.[63] Historical works that address the types of significant question addressed in the natural sciences will foster the impression that there's no significant difference. Some historical works that raise radically different kinds of significant question will contravene that impression. And many texts will present a mixed picture. We now want to close with some brief reflections on the links between history and anthropology and between history and literature.

14.6 Style in History

Here's an obvious counter to the assimilation of history to the sciences: style matters in history, but not in science. Is that correct?

In general, we suggest, rhetoric should be judged by its ability to promote the function the text is supposed to serve. It's a mistake to think that rhetorical considerations don't matter in science: on the contrary, scientific presentations adopt a very particular style, one designed for the cognitive ends that are to be attained.[64] Yet, it seems that the original point can be restated; there are great historical works

[63] Plainly, the sciences often generate new questions that have practical significance for us. The difference we're trying to characterize here is that, because of our background curiosity about the possible forms of human life, history has a particular way of generating new issues for us.

[64] See Kitcher (1991, 1993b).

that are rightly prized for their literary merits, while in correspondingly major scientific texts, any suspicion of a literary flourish is sacrificed to the rigid demands of the conventions of scientific rhetoric.[65] Doesn't this suggest that the cognitive functions are different?

Everything depends on the kinds of questions that historians and scientists are trying to answer. As we've suggested, there are species of history that are very close in aim to the historical sciences – histories that relate what happened in a given time period or that are focused on answering a single why-question (or a cluster of related why-questions). In these instances, the cognitive ends to be served are the perspicuous presentation of a sequence of events or of an ensemble of causes. The style of the historian should be adapted to those ends, enabling the reader to appreciate the elements of the sequence or the comparative importance of the causes. There's no significant difference whether the topic is the medieval papacy or the evolution of the vertebrates. Here, we think that stylistic considerations bear in the same ways on the historical work and on the scientific presentation: what is good for one is good for the other.

Pure cases are probably quite rare. Most histories, we believe, are interested in answering a broader range of questions, and among the questions they'll attempt to address are issues about what a particular past situation was like. In doing that, of course, they'll need to make the past vivid, to present the telling detail in ways that prompt an imaginative response on the reader's part. We alluded earlier to the similarity between this sort of historical writing and the construction of a good ethnography. Because the aims of this type of history are like those of some anthropologists, it's entirely appropriate that texts that offer a "portrait of an age" should be held to the standards of ethnographic writing – standards that typically diverge from those in force in the natural sciences. What succeeds in delineating the precise relations among a complex of causes may not work at all for conveying the lives of past people.

Insofar as both history and anthropology aim to introduce possible ways of living of which readers hadn't antecedently been aware, doing so by highlighting individual situations and characters, they share goals with works of literature. It should therefore be unsurprising that some historical writing has a literary flavor. This is not simply a matter of outmoded, preprofessional, history – the familiar point that Gibbon is worth reading just for the glories of his style; the passage quoted above from Natalie Zemon Davis wouldn't disgrace a work of fiction, and we could make similar claims for many of the historians we have cited.

These observations of important differences between some historical writings and works in natural science – grounded in the kinds of cognitive aims we've tried

[65]The general point can be appreciated by comparing the major works of great scientists with the books in which they summarize their views for a general audience. But there are important exceptions: the famous laconic last sentence of the Watson-Crick paper announcing the structure of DNA, some of Stephen Jay Gould's professional articles in paleontology, and, reverting to an earlier time, Darwin's *Origin*.

to survey in previous sections – lead us to two closing questions. The first concerns the place of literary analysis in the understanding of the practice of history. Since we agree that some historical enterprises may have similar aims to those of literary works, we believe that the tools used to elucidate the latter may prove valuable with respect to the former (and conversely). We can thus welcome studies that identify the literary tropes in historical writing. We are sympathetic to the pioneering ventures of Hayden White, provided that such efforts distinguish the various kinds of cognitive functions served in historical writing and do not operate with the presupposition that all historical texts are structured by narratives.[66] We should not lose sight of the ways in which histories aim to be objective (providing true answers to questions about the past) and either ignore the practice of causal explanation or else distort it by supposing that there's a special type of historical understanding that can only be specified by using the vocabulary of literary criticism (or theory). Just as our view of the practice of history makes room for the expression of modifications of some of Hempel's ideas, so too we'd make room for a transformed version of White's literary analyses, focused on precisely those parts of historical practice where the connections in aim with literature are closest.

Our second question raises the need for truth in history. We've written so far as if the truth of the constituent statements of a historical work were a *sine qua non*: we could be given lots of truths about the past without having a good history, but, it's seemed, we can't have good history in the absence of lots of truths about the past. It seems possible, however, that a historical work might get the big picture right and have most – even all – of the details wrong. Perhaps there are histories that are groundbreaking in their bringing to bear kinds of descriptions that others have overlooked, introducing categories that are crucial to understanding the causes of events or that bring into focus the lives of past people, and yet, for all that, the deployment of these categories is inaccurate. So, for example, it may well be true that there was the kind of shift in understandings of madness that Foucault claims, even though he has picked out the wrong historical episodes and the wrong agents for documenting it.[67] More radically, a historian might deliberately choose to introduce into a historical discussion conjectures for which there's no evidence or even statements known to be false.[68] Doing so might make the past more vivid than it would otherwise have been or might open up possibilities for the reader that had previously not been appreciated. If these are proper goals of historical writing, shouldn't we allow that constituent truth isn't a necessary condition of good history? Hence, we can envisage areas of historical study for which *Veritism* isn't an appropriate aim.

---

[66] See White (1973). White is quite explicit in claiming that all historical texts are structured by narratives, so the position we are recommending requires some adjustment of his views.

[67] Foucault (1965). Although we allude here to a common criticism of Foucault, to the effect that he's wrong about the facts of the history of attitudes to insanity, we don't want to take a stand in this controversy.

[68] See Schama (1991).

We've emphasized throughout that there are many varieties of history, some that are close to the (historical) natural sciences, some that border on anthropology, and some that border on works of literature. Far from restricting historical writing, we'd encourage the development of many different genres, including those that cross the boundaries between history and fiction. Moreover, we'd recall that the natural sciences often fictionalize the phenomena, introducing ideal entities for purposes of shedding light on the behavior of messier and more complicated things. There's no reason to deny historians the same license. Indeed, insofar as we relax *Veritism* for one group, in the interests of significance, we should relax it for the other as well.

But, of course, scientists are usually quite clear where they are pretending, noting explicitly that the pivot isn't frictionless and that the breeding population doesn't satisfy the conditions of the Hardy-Weinberg law. Perhaps historians should honor the same demand, making it clear to their readers just where they have embellished the account.[69] Otherwise, we may be deceived into thinking that we have history "as it actually was" and thus take the work to answer questions to which it wasn't properly directed. In his attempt to draw firm boundaries between history and fiction, the eminent historian Eric Hobsbawm claims that "If history is an imaginative art, it is one which does not invent but arranges *objets trouvés*."[70] Whether or not they receive the label "history," we allow for collages that touch up the objects a bit, provided that the artist acknowledges the handiwork.

How then do we sum up the relation between history and the sciences? Are they akin or are they different? We suggest that the questions invite oversimplified answers – and thus foster unprofitable controversy. A harmless, but not very informative, response would be to point out that the enterprises we group among the sciences are diverse, that the practice of history is also diverse, and that some things we count as history are similar to some of the things we categorize as sciences in their aims, achievements, and methods. A better answer is to provide a picture of both kinds of diversity and to identify the points of similarity and difference among specific historical studies, specific parts of natural science, specific work in anthropology, and specific types of literature. We've been trying to clear the ground for providing that better answer.

References

Ankersmit, F. (1983). *Narrative logic. A semantical analysis of the historian's language*. Boston: Martinus Nijhoff.

Appleby, J., Hunt, L., & Jacob, M. (1994). *Telling the truth about history*. New York: Norton.

Berkhofer, R. F. (1995). *Beyond the great story*. Cambridge, MA: Harvard University Press.

Burckhardt, J. (1949). *The age of Constantine the Great*. Berkeley: University of California Press.

Cohen, P. (1999). *History in three keys*. New York: Columbia University Press.

[69] Schama himself is quite scrupulous in this regard.

[70] Hobsbawm (1997, p. 272); the passage is from the essay "Identity History Is not enough."

Danto, A. (1964). *Analytical philosophy of history*. Cambridge: Cambridge University Press.
Danto, A. (1985). *Narration and knowledge*. New York: Columbia University Press.
Davies, D. B. (1966). *The history of slavery in western civilization*. New York: Oxford University Press.
Davis, N. Z. (1983). *The return of Martin Guerre*. Cambridge, MA: Harvard University Press.
Davis, M. (1990). *City of quartz. Excavating the future in Los Angeles*. London: Verso.
Diamond, J. (1998). *Guns, germs, and steel*. New York: Norton.
Earman, J. (1992). *Bayes or bust?* Cambridge, MA: MIT Press.
Ellis, J. (1976). *Eye-Deep in Hell*, New York: Pantheon.
Ferguson, N. (Ed.). (1997). *Virtual history*. London: Picador.
Ferguson, N. (1999). *The pity of war*. New York: Basic Books.
Field, H. (1972). Tarski's theory of truth. *Journal of Philosophy, 69*(13), 347–375.
Fine, A. (1986). *The shaky game*. Chicago: University of Chicago Press.
Foucault, M. (1965). *Madness and civilization. A history of insanity in the age of reason*. New York: Pantheon.
Giere, R. (1988). *Explaining science*. Chicago: University of Chicago Press.
Ginzburg, C. (1980). *The cheese and the worms*. Baltimore: John Hopkins University Press.
Goodman, N. (1978). *Ways of worldmaking*. Indianapolis: Hackett.
Grant, M. (1976). *The fall of the Roman Empire. A reappraisal*. Radnor: Annenberg School Press.
Hanson, R. N. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.
Hawthorn, G. (1991). *Plausible worlds: Possibility and understanding in history and the social sciences*. Cambridge: Cambridge University Press.
Hempel, C. G. (1965 [1942]). The function of general laws in history. In C. G. Hempel (Ed.), *Aspects of scientific explanation*. New York: The Free Press.
Hexter, J. H. (1971). The rhetoric of history. In J. H. Hexter (Ed.), *Doing history*. Bloomington: Indiana University Press.
Hobsbawm, E. (1997). *On history*. New York: Norton.
Hughes, R. (1986). *The fatal shore*. New York: Knopf.
Humphreys, P. (1990). *The chances of explanation*. Princeton: Princeton University Press.
Keegan, J. (1978). *The face of battle*. London: Penguin.
Keegan, J. (1998). *The First World War*. London: Hutchinson.
Kitcher, P. (1978). Theories, theorists, and theoretical change. *Philosophical Review, 1987*(4), 519–547.
Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation. Minnesota studies in the philosophy of science volume XIII*. Minneapolis: University of Minnesota Press.
Kitcher, P. (1991). Persuasion. In M. Pera & W. Shea (Eds.), *Persuading science: The art of scientific rhetoric*. Sagamore Beach: Science History Publications.
Kitcher, P. (1993a). *The advancement of science*. New York: Oxford University Press.
Kitcher, P. (1993b). The cognitive function of scientific rhetoric. In H. Krips, J. E. McGuire, & T. Melia (Eds.), *Science and rhetoric* (pp. 47–66). Pittsburgh: University of Pittsburgh Press.
Kitcher, P. (2001a). Real realism: The Galilean strategy. *Philosophical Review, 110*(2), 151–194.
Kitcher, P. (2001b). *Science, truth and democracy*. New York: Oxford University Press.
Kitcher, P. (2002). On the explanatory power of correspondence truth. *Philosophy and Phenomenological Research, 64*(2), 346–364.
Kitcher, P., & Salmon, W. (1987). Van Fraassen on explanation. *Journal of Philosophy, 84*(6), 315–330.
Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Kuhn, T. (1977). *The essential tension*. Chicago: University of Chicago Press.
Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science, 48*(1), 19–49.
Laudan, L. (1984). *Science and values*. Berkeley: University of California Press.
Le Roy Ladurie, E. (1979). *Montaillou*. New York: Vintage.
Lewis, D. (1974). *Counterfactuals*. Oxford: Blackwell.
MacDonald, L. (1978). *They called it Passchendaele*. London: Joseph.

MacDonald, L. (1980). *The roses of no man's land*. London: Joseph.
MacDonald, L. (1983). *Somme*. London: Joseph.
MacDonald, L. (1987) *1914*. London: Joseph.
McNeill, W. (1976). *Plagues and peoples*. New York: Doubleday.
Morgan, E. S. (1988). *Inventing the people*. New York: Norton.
Putnam, H. (1981). *Reason, truth and history*. Cambridge: Cambridge University Press.
Roberts, G. (Ed.). (2001). *The history and narrative reader*. New York: Routledge.
Rorty, R. (1982). *Consequences of pragmatism*. Minneapolis: University of Minnesota Press.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Salmon, W. (1989). Four decades of scientific explanation. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation. Minnesota studies in the philosophy of science volume XIII*. Minneapolis: University of Minnesota Press.
Salmon, W. (1998). Scientific explanation: Causation *and* unification. In W. Salmon (Ed.), *Causality and explanation*. New York: Oxford University Press.
Schama, S. (1991). *Dead certainties. Unwarranted speculations*. New York: Knopf.
Seward, D. (1999). *The Hundred Years War*. New York: Penguin.
Stone, L. (1972). *The causes of the English Revolution*. New York: Harper and Row.
Tarski, A. (1956). On the concept of truth for formalized languages. In A. Tarski (Ed.), *Logic, semantics, metamathematics*. Oxford: Oxford University Press.
Thomas, K. (1970). *Religion and the decline of magic*. New York: Columbia University Press.
van Fraassen, B. (1980). *The scientific image*. Oxford: Oxford University Press.
Weinberg, S. (1977). *The first three minutes. A modern view of the origin of the universe*. New York: Basic Books.
White, H. (1973). *Metahistory*. Baltimore: John Hopkins.

Chapter 15
Explanation and Intervention in Coupled Human and Natural Systems

Daniel Steel

Abstract "Coupled human and natural systems" (CHANS) has emerged within the last two decades as a designation for interdisciplinary research focused on complex interactions between human activities and ecosystems. I examine CHANS from a manipulation approach to explanation advocated by Jim Woodward, according to which causal generalizations are distinguished by being invariant under interventions. Several philosophers object that causal generalizations about complex social and biological systems, such as CHANS, often fail to be invariant. This chapter develops the concept of a *robust intervention* to answer this objection, where an intervention is robust to the extent that its ability to promote the intended result is insensitive to errors in the causal model. However, this necessitates rethinking the concept of intervention used by Woodward. Whereas Woodward's concept requires that interventions be *exogenous*, robust interventions are often non-exogenous insofar as involving a sequence of actions wherein later choices are conditional on the results of prior ones. I explain how robust interventions are related to adaptive policies, often discussed in relation to CHANS.

Keywords Explanation • Intervention • Invariance • Coupled-human systems • Natural systems

15.1 Introduction

The label "coupled human and natural systems" (CHANS) has emerged within the last two decades as a designation for interdisciplinary research focused on complex interactions between human activities and natural systems such as the

D. Steel (✉)
Department of Philosophy, Michigan State University, 503 S Kedzie Hall,
East Lansing, MI 48824-1032, USA
e-mail: steel@msu.edu

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_15,

climate, forests, oceans, and rivers (Liu et al. 2007a, b). This chapter focuses on a case study of CHANS research, namely, explanations of forest degradation in the Wolong Panda Reserve in China (hereafter, Wolong). I consider this case in connection with Jim Woodward's (2003) theory of causal explanation, according to which explanatory generalizations are distinguished by being invariant under interventions. Woodward's approach appears initially promising in relation to this case given the emphasis placed in his theory on the connection between explanation and intervention. Unsurprisingly, explanations of forest degradation in Wolong are explicitly intended to assist the design of more effective habitat conservation policies. However, it is unclear whether explanatory models developed by these researchers satisfy Woodward's requirements for an invariant generalization. Indeed, the CHANS literature consistently emphasizes the potential for "surprises," that is, outcomes that differ significantly from what would have been expected on the basis of a seemingly well-confirmed explanatory model. This point is closely related to objections to Woodward's theory raised by Sandra Mitchell (2009, Chap. 4) and Julian Reiss (2009, pp. 25–26) in relation to biology and social science, respectively. Both Mitchell and Reiss argue that causal relationships may be fragile rather than invariant under interventions, and hence that invariance is problematic as a general criterion of causal explanation.

I suggest that adequately addressing this difficulty within the context of Woodward's theory requires modifying his concept of intervention. More specifically, I distinguish between *experimental interventions* whose purpose is to learn about causal relations and *practical interventions* that aim to promote a desired outcome, such as protecting panda habitat. Although the practical need to intervene in our surroundings is a central motivation for Woodward's approach (see 2003, Chap. 2), his definition of intervention is an abstract characterization of a randomized controlled experiment (2003, p. 98). Moreover, the types of interventions most relevant to the case study I examine – and many other complex systems – differ in significant respects from the ideal experimental interventions considered by Woodward. In particular, they are (or should be) designed to accommodate the possibility of "surprises," in other words, the possibility that our explanatory model might fail to be invariant in unexpected ways. But for reasons I will explain, such interventions cannot qualify as interventions at all given Woodward's definition of that concept, and hence his approach is incapable of exploring the relationship between such interventions and explanation. Consequently, I explore what a manipulation approach to explanation along the lines of Woodward's approach might look like if it utilized a more expansive notion of intervention. Finally, on the basis of this modification, I suggest that the concept of a "ceteris paribus model" is helpful for understanding how, from a manipulationist perspective, a causal model could be deemed explanatory despite failing to be invariant.

15.2 CHANS

Specialization is an obvious and inevitable characteristic of modern science; as scientific knowledge has grown, the breadth of topics in which a single researcher can maintain expertise has persistently shrunk. The interest in CHANS as a separate research focus in its own right arises from this specialization together with the inconvenient fact that the world does not neatly divide itself up according to disciplinary boundaries found in science. Most fundamentally, while the division between social and natural science is longstanding and deeply entrenched, human-nature interactions are at the heart of pressing environmental issues from climate change to toxic chemicals to biodiversity. Awareness of this mismatch between the borders of scientific disciplines and environmental problems has resulted in explicit efforts within the scientific community to foster research on CHANS. For example, since 2001, *Dynamics of Coupled Natural and Human Systems* has been a funding category for the National Science Foundation in the United States, and there is an *International Network of Research on Coupled Human and Natural Systems* (home page: chans-net.org), which promotes events such as symposia on CHANS at the 2011 meeting of the American Association for the Advancement of Science.

However, scientific study of CHANS not only requires some reorganization of scientific social structures, but it also involves grappling with a number of features of CHANS that pose significant challenges for attempts to discover causal explanations that can serve as the basis for informed policy making. General discussions of CHANS (see Liu et al. 2007a, b) highlight several of these, which I summarize below.

- *Nonlinearity and Thresholds*: The impact of the cause upon the effect is not constant across distinct levels of the cause. For instance, the detrimental effect of lakeshore housing upon fish habitat may sharply increase once housing surpasses a critical density (Liu et al. 2007a, p. 1514).
- *Legacy Effects and Time Lags/Indirect Effects*: Causes and effects may be very distant in time, space, and in terms of the intermediate steps in the causal chain (Liu et al. 2007a, p. 1515, b, pp. 640–641). Climate change is one obvious example.
- *Heterogeneity*: Distinct groups and locations within the area under study respond differently to the same causes (Liu et al. 2007a, pp. 1515–1516, b, p. 642). For example, a conservation policy might have distinct impacts depending on cultural or economic characteristics of the people involved.
- *Reciprocal Effects and Feedbacks*: Coupled human and natural systems often exert mutual influences upon one another that play out over extended periods of time (Liu et al. 2007a, pp. 1513–1514, b, pp. 639–640). For

(continued)

(continued)

example, a pristine ecosystem attracts tourists but tourism can degrade the qualities of the ecosystem that attract tourists.

– *Surprises/Vulnerability*: Policy interventions in CHANS often have unintended, and undesirable, consequences (Liu et al. 2007a, pp. 1514–1515, b, p. 641). For instance, a species introduced to an ecosystem for a particular purpose (say, to serve as prey for an existing predator species) may produce different results than expected (say, eating the predator's young).

The characteristics listed above are by no means unique to CHANS, but are features often associated with complex systems generally (see Mitchell 2009; Taylor 2005). The last of them, surprises/vulnerability, is best conceived as a consequence of other aspects of complexity, not as something that explains why the system is complex. Thus, unexpected results of policies might stem from such things as thresholds or heterogeneity in the CHANS. In the next section, I will discuss the relationship between the above characteristics of CHANS and Woodward's manipulation approach to causal explanation. For now, I will describe an example of CHANS research – focused on the Wolong Nature Reserve in China – that illustrates the features noted above.

Created for the purpose of protecting the habitat of the endangered giant pandas (*Ailuropoda melanoleuca*), the Wolong Nature Reserve (henceforth, Wolong) was initially established in 1963 and designated as a national nature reserve in 1975, at which time it was expanded to its current size of approximately 200,000 ha and was protected from commercial logging. Wolong is a natural site for CHANS research for several reasons. First, Wolong is a prime example of human-nature interactions. Besides pandas and other wildlife, Wolong contains several thousand local human residents, and the protection of the habitat of wild pandas depends in large measure on managing the interactions between humans and forests. In addition, the reserve is about a 4-h bus ride into the Qionglai Mountains northwest of Chengdu, the capital of Sichuan Province and metropolis of over seven million people, making it easily reachable by tourists from around the globe. The influx of tourists has had a significant impact on the local population of Wolong, which in turn affects the forest habitat that the reserve was created to protect. Secondly, the history of Wolong illustrates the potential for well-intentioned policies to yield surprising and undesirable results that stem from a failure to adequately appreciate the inherent complexity of CHANS.

Perhaps the classic CHANS-genre article on Wolong is titled "Ecological Degradation in Protected Areas: The Case of Wolong Nature Reserve for Giant Pandas" (Liu et al. 2001). This article describes how remote sensing data shows that the degradation of the pandas' forest habitat accelerated after the creation of the reserve in 1975 and in fact proceeded at a higher rate than in nearby areas outside the reserve. This result was especially surprising given that a ban on commercial

logging was one of the primary effects of designating Wolong a national nature reserve. The increased habitat degradation appears to stem from factors relating to the local population in connection with the impact of tourism. The local population expanded from 2,560 residing in 421 households in 1975 to 4,260 residing in 921 households in 1997 (Liu et al. 2001, p. 100). Since local residents harvest forest wood for fuel and building materials, this increase in the population and number of households accelerated habitat degradation. The increase in local population can be traced to several factors. About 80 % of the local residents in Wolong are members of ethnic minorities who are not subject to China's "one child" policy and hence whose birth rates are higher than in some neighboring areas (ibid). In addition, the tourist boom following the establishment of the reserve counteracted the urban migration of working-age people that is typical of rural China, and the portion of the population between 20 and 59 years of age increased by 60 % from 1982 to 1996 (ibid). The tourist business also stimulated forest degradation related to economic activities, such as clearing land to grow food to sell to hotels and restaurants within the reserve (ibid). This was not the last "unpleasant surprise" outcome of a well-intentioned intervention in Wolong. For instance, a forest conservation program initiated in 2001 paid local residents per household to monitor illegal wood collection, which resulted in an increase in the founding of new households, thereby creating more demand for wood for fuel and building material (Liu et al. 2007a, pp. 1414–1415).

These examples of unpleasant surprises illustrate several of the aspects of CHANS highlighted above. Foremost among these is the role of inadequately anticipated feedbacks: a panda habitat attracts tourism which in turn has detrimental effects on the panda habitat; payments for monitoring illegal wood collection inadvertently provide an incentive to form new households which increases wood harvesting. The role of lagged and indirect effects is also evident in this case. A newly designated nature reserve does not become a popular tourist destination overnight: it must be advertised through travel media, such as guidebooks, and infrastructure to accommodate visitors, such as roads and hotels that must be built. Moreover, the detrimental effects of tourism on panda habitat are largely indirect and "behind the scenes," working through the economic impacts of tourism on the local population rather than resulting directly from the presence of the tourists themselves. The Wolong case also illustrates heterogeneity, since the ethnic and cultural makeup of the local populations matters and differs across panda habitats in China. Finally, the Wolong case illustrates a common feature of complex systems not listed above, namely, the "bushiness" of effects stemming from a cause. The creation of the Wolong Nature Reserve had a number of effects, including stopping commercial logging in the reserve and promoting tourism, and these effects in turn also had multiple effects. Furthermore, separate lines of causal influence are often at cross-purposes, some working for and the others against the desired result. This not only makes it difficult to predict the overall impact but also makes unintended negative consequences likely when some contrary causal paths are overlooked.

Useful scientific study of CHANS, then, requires some means of integrating knowledge concerning various aspects of complex human-nature interactions.

Agent-based modeling (ABM) is one approach currently being used for this purpose and applied specifically to the Wolong case (An et al. 2005; Chen et al. 2012). ABM takes a "ground up" approach, building a model in which the behaviors of a large number of interacting agents can generate surprising "emergent" properties as a whole. For example, the An et al. model focuses on factors implicated in harvesting wood among the local population in Wolong, which is one of the key factors in panda habitat degradation as mentioned above. In this model, households are laid out on a grid of pixels, where each pixel contains information about such things as elevation, slope, and land cover (An et al. 2005, pp. 58–59). The driving forces in the model are grouped into three main categories: household development, fuelwood demand, and fuelwood growth and harvesting (ibid). Household development includes such things as growing in size due to births, shrinking due to deaths or emigration, and the founding of a new household or the dissolution of an old one. A submodel, based upon field research in Wolong (An et al. 2003), represents the variables that influence household development (An et al. 2005, pp. 61–64). Similarly, there are submodels for fuelwood demand and fuelwood growth and harvesting also based upon Wolong research (An et al. 2005, pp. 64–66). Some key variables depend on culturally specific aspects of the Wolong population. For instance, being the youngest of several siblings is an important consideration in deciding whether to found a new household; having an elder in the household significantly increases the demand for fuelwood for heating; more cropland results in greater fuelwood demand because grain and potato crops are cooked and fed to pigs whose meat is sold to local restaurants and hotels (An et al. 2005, pp. 58–59). The output of the model was also tested against data from the years 1997 to 2000 for empirical validation (An et al. 2005, p. 67).

Another model based on Wolong data examines the effect of social norms in reenrollment in the Grain-to-Green Program, which pays farmers to convert cropland on sloping hillsides to grass or forest cover (Chen et al. 2012). For example, this model found that higher reenrollment at a given payment level could be achieved through reenrolling landholders in waves, as this provided more opportunities for landholders to learn of the decisions of others prior to making their own decision (Chen et al. 2012, p. 7).

One important question for this chapter is what CHANS researchers take the purpose of models such as those described above to be. Both of the articles cited above explicitly address this issue. The An et al. article states the following:

> Using this combined model has enabled us to develop a better understanding of the relationships between people and panda habitat in Wolong, which may, in turn, help to develop environmentally sound policies in the reserve. . . . This framework is a powerful means for integrating data and models across varying scales and disciplines and shows promise for many human-environment studies. (An et al. 2005, p. 77)

Chen et al. state:

> In this study, we focus on explaining the effect of descriptive social norms on decisions regarding reenrollment in PES [Payment for Ecosystem Services] programs, as well as on understanding the effect of different PES program designs on the emergence of descriptive social norms. (Chen et al. 2012, p. 2)

Note the closely linked emphasis on understanding and policy in both of these passages. This fits nicely with the manipulation account of explanation, which links intellectual interest in explanation to practical concerns about how to effectively attain desired ends. Integration of empirical data from a variety of sources is a second motivation emphasized in the first of the two quotations above. That is, ABM provides a format in which incomplete knowledge about causal relationships can be integrated to guide interventions. In Sect. 15.4, I will discuss how this approach connects uncertainty, robustness, and intervention. For now, let us turn to Woodward's manipulation theory of explanation.

15.3 Invariance, Intervention, and Explanation

According to Woodward, a generalization is a potential basis for causal explanation just in case it is invariant under some range of interventions (Woodward 2003, pp. 12–17). Generalizations that are invariant in this sense indicate variables that can be used as "levers" to manipulate an effect and can be used to answer what Woodward terms *what-if-things-had-been-different* questions (2003, p. 191). A genuine explanation "can be used to answer a range of counterfactual questions about the conditions under which their explananda would have been different" (ibid.). For example, the Chen et al. model described above could, if invariant, be used to answer questions about how reenrollment rates in the GTG program would vary depending on whether individuals were reenrolled all at once or in waves (and if in waves, how many and at what time intervals). This example also illustrates the connection between what-if-things-had-been-different questions and practical matters of designing interventions to promote desired outcomes. Despite his emphasis on intervention, Woodward insists that causation is objective: it is a fact of the world that some generalizations are invariant under interventions, while others are not, and facts of this kind are of great practical importance to humans and many other organisms (Woodward 2003, pp. 119–123). Let us, then, take a closer look at Woodward's concept of intervention.

Woodward approaches the definition of intervention by first defining what he calls *an intervention variable*, which I quote in full:

> Let X and Y be variables, with the different values of X and Y representing different and incompatible properties possessed by the unit u, the intent being to determine whether some intervention on X produces changes in Y. Then I is an intervention variable for X with respect to Y if and only if I meets the following conditions:
>
> **(IV)**
>
> I1. I causes X.
> I2. I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of the other values of other variables that cause X and instead depends only on the value taken by I.

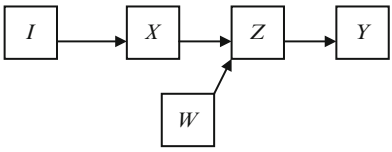


Fig. 15.1 The intermediate Z has a cause other than X

> I3. Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the I-X-Y connection itself; that is, except for (a) any causes of Y that are effects of X (i.e., variables that are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.
> I4. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X. (Woodward 2003, p. 98)

This definition is intended as an abstract description of the type of circumstances that would be attained in an ideal randomized controlled experiment. Unfortunately, however, I4 contains an error, and it is moreover precisely this part of the definition that is the focus of the ensuing discussion. I4 is intended to ensure that the intervention is exogenous, in the sense of being unaffected by things that influence Y. But, as stated, I4 entails that interventions are impossible in the normal circumstance in which intermediate causes between X and Y have causes other than X. An example of this is given in Fig. 15.1, wherein $I \rightarrow X \rightarrow Z \rightarrow Y$, and $W \rightarrow Z$. In this case, Z is a cause of Y and Z is on a directed path that does not go through X (i.e., $W \rightarrow Z \rightarrow Y$). So, I4 requires that the intervention I be statistically independent of Z. But that is clearly inappropriate as I is an indirect cause of Z: if X causes Y, then I and Z would be expected to be probabilistically dependent in a properly conducted experiment.[1] However, a corrected version of I4 can be formulated as follows:

> I4*: I is not an effect of Y and there is no common cause of I and Y.[2]

A common cause of I and Y would be a variable positioned like so: $I \leftarrow \ldots \leftarrow Z \rightarrow \ldots \rightarrow Y$ (i.e., non-overlapping causal paths from Z to I and from Z to Y). Note that if I4 is replaced with I4*, then I is an intervention variable with respect to X and Y in Figs. 15.1 and 15.2. Thus, I4* captures the idea that the intervention is exogenous or causally "upstream" of the outcome Y.

[1] A similar error occurs in Craver's (2007, p. 96) adaptation of Woodward's definition of intervention (although Craver reorders Woodward's four criteria, so Craver's I3 corresponds to Woodward's I4).

[2] Given Pearl's concept of d-separation (2009, pp. 16–17), I4* could be equivalently stated as I d-separates Y from the parents of I.

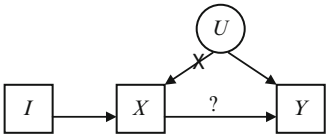


Fig. 15.2 An intervention variable for X with respect to Y

Consider Woodward's definition of an intervention variable in relation to a clinical trial of a pharmaceutical. In that case, values of X could indicate whether the subject received the drug or a placebo, while Y would represent the outcome in question (e.g., whether blood pressure was lowered or not). Thus, assigning each subject to the control or test group constitutes the intervention variable I, and I1 is satisfied so long this has some effect on whether the subjects take the drug. To achieve I2, we would have to ensure that the experimental subjects entirely complied with their treatment assignment. Standard experimental procedures such as double blinds are intended to focus the impact of the intervention solely on X as I3 requires. Finally, random treatment assignment is a means of achieving I4*. Figure 15.2 provides a graphical representation of an intervention variable. The intervention variable I directly targets X and X alone, it eliminates the influence of other causes that would normally affect X (represented by crossing out the arrow pointing from U to X), and it is exogenous since it is not an effect of any other variable in the graph.

An intervention variable, however, is not the same as an intervention. Woodward defines an intervention as a cause that actually sets an intervention variable to a particular value:

> (**IN**) I's assuming some value $I = z_i$, is an intervention on X with respect to Y if and only if I is an intervention variable for X with respect to Y and $I = z_i$ is an actual cause of the value taken by X. (Woodward 2003, p. 98)

I will not describe Woodward's conception of actual cause (see Woodward 2003, pp. 74–86). The actual causes relevant to the examples considered here will be actions, such as designating Wolong a national nature reserve. An intervention variable, then, can be thought of as an experimental setup, while an intervention would be a particular act taken within that setup, for example, assigning a particular subject to the group that receives the active treatment.

A generalization relating variables X and Y, then, is invariant if it continues to hold under some range of interventions on X (see Woodward 2003, p. 250). A few clarifications are in order here. First, it is not required that the interventions in question can actually be carried out. Invariance is a counterfactual property: the generalization would continue to hold where certain (possibly hypothetical) interventions performed.[3] It is also not necessary that the generalization continues to hold under all interventions in all background conditions. A generalization might be invariant only under some restricted range of interventions in a particular time and

[3]Indeed, Woodward does not even require that the intervention be physically possible (Woodward 2003, pp. 127–133).

place. Invariance, then, has both a threshold and degrees: some generalizations are not invariant at all, and among those that are invariant, some are more so than others (Woodward 2003, pp. 257–265). Thus, explanation according to Woodward's theory does not require laws of nature, which makes his approach appealing for fields in which such laws seem to be absent (Woodward 1999, 2001, 2003, pp. 265–279).

However, some critics of Woodward's approach argue that causal generalizations in some areas of science are not invariant under any feasible intervention. This situation arises if (a) the causal relationship in question is highly context dependent and (b) any possible intervention would alter the context in such a way to disrupt that causal relationship. Sandra Mitchell argues that null results of gene knockout experiments illustrate this scenario (Mitchell 2009, Chap. 4). By tracing metabolic pathways in a cell, we might find that a particular gene is involved in producing a specific enzyme, but an experiment in which that gene is knocked out may nevertheless generate no noticeable result if removing the gene activates a backup mechanism that also produces the enzyme. In a similar vein, Julian Reiss observes that, while manipulation conceptions of causation and explanation are common in economics, there are well-known examples of causal generalizations that turned out not to be invariant (Reiss 2009, p. 26). According to Reiss, this occurs because an intervention must inevitably change the context to which the causal relationship is sensitive:

> Since the aggregate relations depend for their existence partly on the economic agents' expectations, and policy interventions may change the expectations, the aggregate relations may be disrupted by policy. (Reiss 2009, p. 32)

The discussion of CHANS in the foregoing section also illustrated the possibility that causal relationships may be highly context sensitive and hence do not behave as expected when an intervention is performed.

I see two general options for Woodward here. First, he could suggest that his theory provides an ideal model of what a causal explanation should be that can be used as a measuring stick for evaluating imperfect explanations. If researchers in biology or social science or multidisciplinary fields like CHANS have failed to produce invariant generalizations, Woodward might say, "that means they just need to work harder to deliver the goods." Mitchell anticipates this type of response and characterizes it in the following way:

> Perhaps we have not correctly described the causal structure to capture all of its causal relations. A complete description might include not only the causes active in the normal operation of the genetic network or of the nerves and brain network but also those that are possible given any internal or external perturbation to the system. (Mitchell 2009, p. 80)

There is, I think, something right about this line of response, namely, that any model devised to represent an extremely complex system is likely to be incomplete in ways relevant to predicting the consequences of interventions. But the problem, as Mitchell points out, is that including all interacting contextual factors in a representation of a complex system in biology or social science may be a hopeless task (Mitchell 2009, pp. 80–81). And if this is right, then admonitions to "just try harder" are unhelpful, because the pursuit of an unattainable ideal may

be counterproductive when it diverts effort and resources away from achievable endeavors.

The second option for Woodward would be to modify his theory in a way that preserves its core insights about the link between explanation and intervention but which enables it to more adequately treat sciences that study highly complex systems. It is this second option I will pursue here. The modification of Woodward's theory that I propose concerns the concept of intervention described above. There are two general reasons why one might carry out an intervention, practical and experimental. A practical intervention is intended to promote some desired end, for example, to protect panda habitat by designating Wolong a national nature reserve. The purpose of an experimental intervention, by contrast, is to learn about cause and effect, as in a randomized controlled experiment intended to assess the effectiveness of a medical therapy. Practical interventions figure prominently in Woodward's motivating discussion of his approach; according to Woodward, the practical necessity of manipulating our surroundings explains why we have concepts of causation and causal explanation (Woodward 2003, Chap. 2). Yet the definition of intervention that Woodward gives is modeled on experimental rather than practical interventions. Note that Woodward's definition of "intervention variable" quoted above is prefaced with the following: "*the intent being to determine whether some intervention on X produces changes in Y*" (Woodward 2003, p. 98; italics added). In other words, that definition characterizes an intervention whose purpose is not to achieve some practical aim but instead to learn what effect, if any, X has on Y. But if practical interventions are the inspiration for the manipulation theory, then this would seem to be the wrong intervention concept.

Practical and experimental interventions differ in a number of ways, but the most important difference for this chapter concerns the requirement, expressed in I4* of Woodward's intervention variable definition, that an intervention be exogenous. This is a good idea for an experimental intervention. If treatment assignment in a controlled experiment is influenced by factors that also affect recovery, then any statistical association between treatment and recovery is confounded. But exogeneity is *not* a good idea at all for practical interventions. Suppose the question is which treatment should be given to alleviate a patient's illness. In this case, the patient's symptoms *should* influence which medical intervention is performed – obviously, the physician should not assign a treatment at random! Likewise, the decision to designate an area a national nature reserve should be influenced by the features of that area (e.g., that the area contains a valuable yet endangered environmental resource such as a large panda habitat). So, if practical intervention is the relevant concept, then I4* should not be in the picture. Moreover, interventions that violate I4* are especially useful for managing complex systems. In particular, interventions in complex systems can often benefit from being conditional in the sense that the action taken at each step is contingent on the results of prior actions. Pearl (2009, p. 345) refers to such interventions as *conditional plans*, a concept that is very similar to the notion of *adaptive policies*, which are particularly relevant to CHANS as will be discussed below.

15.4 Robustness and Autonomy

This section develops the idea that conditional plans are useful for complex systems because they can make interventions more *robust* in the face of uncertainties, and explores modifications to Woodward's approach that follow from countenancing non-exogenous interventions of this kind. In Sect. 15.4.1, I explain why robust interventions are, in contrast to Woodward's definition discussed in the prior section, often not exogenous. In Sect. 15.4.2, I discuss Pearl's (2009) concept of atomic interventions, which are not necessarily exogenous and which can be used to explicate the notion of a conditional plan. I explain how atomic interventions can be used to define the concept of autonomy, which I argue is a preferable substitute for Woodward's notion of invariance. Finally, in Sect. 15.4.3, I consider this approach in relation to the objection raised by Mitchell and Reiss.

15.4.1 Robust Interventions

Robust interventions rely on causal knowledge to promote a desired outcome, but hedge their bets against uncertainty and "surprises." More exactly, an intervention is robust to the extent that its ability to promote the intended result is insensitive to errors or omissions in the causal model. Given the emphasis on an "intended result," robust interventions clearly fall into the category of practical rather than experimental interventions. Robustness is an important virtue of interventions when two conditions are present: (1) uncertainty about the correct causal model and (2) the system being studied is such that errors or omissions in the model may lead to interventions going significantly awry. The characteristics listed in Sect. 15.2 that make CHANS complex are directly related to the following points:

- *Nonlinearity and Thresholds*: This makes predicting effects of interventions sensitive to getting the threshold right. It may also, for various technical reasons, make it more difficult to infer causal relations from statistical data.[4]
- *Legacy Effects and Time Lags/Indirect Effects*: Time lags make causal inference difficult because they require data sets that cover long expanses of time. They can also make it difficult to predict consequences of interventions, because lags allow ample opportunity for exogenous changes in workings of the system between cause and effect.
- *Heterogeneity*: Distinct subsets of the population respond differently to the same cause and similarly for separate populations. So, an effect in a subpopulation may differ significantly from the average effect in the population overall, and an effect in one population may not be a good

(continued)

(continued)

guide for that in another. This creates difficulties for predicting the effects of interventions if the intervention targets a population distinct from the one from which data was collected (Steel 2008). Heterogeneity is also related to the possibility of conflicting causal pathways emanating from the intervention, some which promote and some which hinder the intended outcome. In such cases, distinct populations may differ as to which pathway is the predominate one.

– *Feedbacks*: Like thresholds, these also complicate causal inference for various technical reasons.[5] Feedback effects can also amplify or accelerate the impact of a cause, which can result in substantial errors in predicting the results of an intervention if the feedback effect is not represented accurately.

None of these circumstances are incompatible with causal generalizations that are invariant in Woodward's sense. For instance, a causal model could have feedbacks and thresholds and still continue to hold under interventions performed on the independent variables. Nor do they pose in-principle barriers to learning causal relationships. But they do make it more difficult to discover the correct causal model, and they can make predicting the consequences of interventions more sensitive to getting the model just right.

So even if invariant causal generalizations exist for complex systems like CHANS, substantial uncertainties may be associated with any particular model that can be devised, and these uncertainties will often have implications for our ability to predict the results of interventions. Consequently, robust interventions are highly desirable for CHANS. The hard question, then, is how to design interventions so that they are more robust. Although the literature on CHANS does not, to my knowledge, use the term "robust intervention," it nevertheless includes some discussion of precisely this issue:

> Despite the obvious importance of using information from CHANS studies for policy making, governance, and management of natural resources, recognizing the incompleteness of knowledge about CHANS and the inevitability of surprises is vital. The negative consequences of inherent uncertainty and the increasing likelihood of surprises can be minimized by 3 approaches: (i) maintaining margins of safety to account for uncertainties (e.g., in calculating fisheries quotas), (ii) factoring in insurance as a hedge against disasters (e.g., adding in a buffer of additional area in calculating the size of marine reserves), and (iii) ensuring adaptive mechanisms. These approaches are all essential elements of a strategy to effectively manage CHANS . . . . (Liu et al. 2007b, p. 645)

[4]For example, thresholds can result in counterexamples to the commonly made assumption that if X is a cause of Y, then X and Y are probabilistically dependent see Neopolitan 2004, p. 99).

[5]For example, the causal Markov condition, a common assumption in causal inferences see Spirtes et al. 2000, pp. 11–12; Pearl 2009, p. 30), can fail if causal feedbacks are present see Steel 2006). See Richardson and Spirtes 1999) for a proposal about causal inference can proceed in such cases.

Approaches (i) and (ii) are similar and can be usefully condensed into one idea, that is, when it is worse to err in one direction than another, one should "overshoot" somewhat towards the less bad mistake. Approach (iii) concerns adaptive policies, that is, policies that are not implemented all at once but which consist of a succession of interventions, each of which is contingent on the outcomes of those that went before. Let us consider the relation of adaptive mechanisms or policies to robustness.

This concept is very similar to Pearl's (2009, p. 354) concept of a conditional plan. A conditional plan consists of a series of actions, wherein the choice of action at each stage is contingent on the results of the previous ones and where the choice of the first action may be influenced by observed features of the system. Such interventions are commonplace in medical treatment. A patient presents certain symptoms to the physician who then prescribes a particular therapy and requires a follow-up examination. Depending on the patient's symptoms at the follow-up, the physician may recommend a continuation of the original therapy, switching to a distinct therapy, or discontinuing treatment altogether (e.g., if the problem has gone away). Given the presence of uncertainties concerning diagnosis and the precise conditions of the patient, conditional interventions are more robust than "once and for all" interventions that implement a fixed plan of action. For example, suppose that the physician's choice is between two treatments: one which is more effective but which sometimes has severe side effects and a second that has no side effects but which is not always effective. Moreover, suppose that there is no way to know in advance which treatment will work best for the patient. Then there is no robust fixed treatment, since the treatment will fail to promote the patient's recovery if the physician guesses wrong about which treatment will work. In contrast, the success of a conditional intervention is much less dependent on having the right diagnosis from the start, since the patient can be switched to other therapy if the initially chosen one does not produce positive effects.

Consider this idea more precisely. Let the variable X indicate which treatment is administered, where x_1 is the treatment with the lower chance of recovery but with no side effects, while x_2 is the one with a higher chance of recovery but the possibility of harmful side effects. Let Y be a variable for recovery (y_1 if the patient recovers, y_2 if the patient does not), and let S be a variable for side effects (s_1 if the side effects occur, s_2 if they do not). Suppose that x_1 is effective if mechanism M_1 is present in the patient, while x_2 leads to side effects and hence is ineffective and positively harmful, if mechanism M_2 is present. As we are uncertain about the causal processes at work in the patient, we do not know whether these mechanisms are present, but we judge their probabilities to be as given in Table 15.1. Given this table, consider the probabilities of success (i.e., of resulting in y_1) resulting from the fixed treatments x_1 and x_2. Since treatment x_1 is effective when and only when M_1 is present, the probability of y_1 given a fixed treatment consisting solely of x_1 is 75 %. Similarly, since x_2 is effective when and only when M_2 is absent, its probability of resulting in recovery is 90 %.

In contrast to these fixed strategies, consider the following two conditional plans, both of which involve two stages: an initial treatment recommendation (either x_1 or x_2), followed by a "checkup" at which time the treatment may remain the same or

Table 15.1 Uncertainty about mechanisms

M_1	M_2	Probability
Present	Present	.05
Present	Absent	.7
Absent	Present	.05
Absent	Absent	.2

be switched. Conditional plan 1 starts with x_1 but switches to x_2 at the checkup if the patient shows no signs of recovering. Conditional plan 2 starts with x_2 but switches to x_1 at the checkup if harmful side effects are present (i.e., s_1 is observed). The probability of recovery given these conditional plans depends on the reliability of the observations at the checkup. Consider the idealized case in which these observations are perfectly reliable (i.e., signs of recovery are observed at the checkup if and only if the patient will in fact recover, and side effects are observed if and only if they exist). In that situation, each conditional plan succeeds in rows 1, 2, and 4 of Table 15.1, and hence the chance of recovery for both is 95 %. In contrast, the fixed treatments succeed in only two out of the four possibilities, resulting in a lower probability of recovery. When the factors that guide the decision whether to switch treatments are less than perfectly reliable, conditional plans will continue to do better than fixed treatments so long as these indicators are reliable enough. Conditional plans, then, take advantage of the fact that an action may generate data that would not otherwise exist, where these data reduce uncertainty about the causal processes at work in the system in question. In such circumstances, conditional plans will often promote the desired result in a broader set of possible causal scenarios than any fixed strategy, making them more robust.[6]

The medical example just described is similar in some respects to the Wolong case, wherein uncertainty about side effects of an intervention was also a challenge for designing an effective policy. Moreover, adaptive policies have been employed in real cases relating to CHANS, for instance, China's Grain-to-Green (GTG) Program which pays farmers not to cultivate land on steeply sloping hillsides so as to reduce erosion and flash flooding. The program was initially implemented for a specified period, followed by an assessment of its effectiveness, which in turn determined whether to continue the program (Liu et al. 2008). A manipulation approach to explanation that hopes to be relevant to sciences that study very complex systems, then, should be able to explore relationships between explanation and a type of intervention that is especially important for such systems, namely, conditional plans. However, in its present form, Woodward's theory is incapable of this for the simple reason that it judges conditional plans not to be interventions at all.

[6]For more on robustness and decision making, see Lempert et al. 2003, 2006; Popper et al. 2005. Several philosophers have also provided accounts of the notion of robustness (e.g., Henderson and Horgan 2001; Woodward 2006), but not to my knowledge in relation to robust interventions, which is the focus here.

15.4.2 Autonomy

Adaptive policies or conditional plans would not count as interventions according to Woodward's definition because they do not satisfy condition I4*, which requires that the intervention be exogenous, that is, not influenced by factors that affect the outcome. A medical therapy, for instance, can be influenced by symptoms presented by the patient that are affected by earlier stages of the therapy and by the underlying disease, both of which affect the patient's chance of recovery. However, concepts of intervention exist that do not include some analogue of I4*. In particular, consider Pearl's notion of an "atomic intervention" (2009, p. 70). An atomic intervention sets a single variable to a specific value but does not otherwise affect causal relationships. In the example about the choice of medical treatment, the variable could represent which treatment is prescribed, and the intervention could consist of the physician's act of prescribing one of the available options. Pearl uses structural equations to represent causal relationships, in which the variables on the right-hand side of each equation are the direct causes of the variable on the left-hand side. In this framework, an atomic intervention on X is represented by "wiping out" the right-hand side of the equation for X and replacing it with a constant that indicates the value to which the variable has been set. The key idea is that the atomic intervention targets one and only one variable, say X, and wholly determines its value. This entails I1, I2, and 13 of Woodward's definition of an intervention variable, but not I4* (or the original I4). So, the key difference between Woodward's and Pearl's intervention concepts is that the first requires that interventions be exogenous while the second does not. Consider how an adaptive intervention, such as the ones considered in the medical example discussed above, would be represented given Pearl's approach. Pearl uses the notation $do(X = x_1)$, typically abbreviated $do(x_1)$, to indicate the atomic intervention of setting the random variable X to the specific value x_1. (The "*do*" in "$do(x_1)$" is for *doing*, that is, it says that the value of X is determined by action rather than being passively observed.) In Pearl's terminology, a *conditional action* would be represented as $do(X = g(z))$, where X is a variable indicating which treatment is prescribed, g is a function, and z are the particular values of the set of variables Z that determine the choice of treatment (2009, p. 113). A *conditional plan*, then, consists of a sequence of conditional actions, wherein the values of the variables in the set Z that determine the action taken at one stage can be influenced by prior actions (see Pearl 2009, pp. 354–355). In the medical example, variables in the set Z would represent various symptoms or results of diagnostic tests, which in turn would be influenced by the underlying disorder or disease, which is not directly observed. Hence, the intervention in such cases is typically not exogenous, since it is indirectly influenced by factors that also affect the outcome.

Woodward considers Pearl's intervention concept, but argues that it is not appropriate for his project of identifying the distinctive feature of a generalization that enables it to generate causal explanations (2003, pp. 110–111). The concern is that omitting the requirement that the intervention is exogenous could result in admitting correlations due to confounding as explanatory causal generalizations.

However, I show that the concept of *autonomy*, definable via Pearl's notion of atomic interventions, does not have this shortcoming and, furthermore, better suits the emphasis on practical interventions that motivate the manipulation theory of explanation.

Autonomy is a property of a set of structural equations (see Pearl 2009, pp. 27–29). For example, consider these:

$$x = f_1(u_1)$$

$$w = f_2(u_2)$$

$$z = f_3(w, x, u_3)$$

$$y = f_4(z, u_4)$$

Given the convention that variables on the right-hand side are direct causes of those on the left-hand side, this set of equations corresponds to the diagram in Fig. 15.1. The u_i's represent unmeasured causes and are associated with a joint probability distribution $P(u_1, \ldots, u_n)$.[7] A set of causal equations S is *autonomous* if and only if, for any atomic interventions on members of any subset of S, all the other equations in S are true (see Pearl 2009, p. 22, p. 28). For example, if the set of equations above is autonomous, then f_2 through f_4 continue to hold true given an atomic intervention that "erases" $f_1(u_1)$ and replaces it with the constant c.

A generalization $y = g(x)$ is invariant according to Woodward's definition if and only if it continues to hold true under some interventions that satisfy the definitions discussed in Sect. 15.3 (Woodward 2003, p. 250). From a structural equations perspective, Woodward's definition is restricted to the special case in which the system of equations has only two members, $x = g_1(u_1)$ and $y = g_2(x, u_2)$. Then invariance just means that $y = g_2(x, u_2)$ continues to hold true for some exogenous atomic interventions that set x to some constant.[8] Given this it is easy to show that if exogenous atomic interventions exist, then autonomy entails invariance. For suppose the set of equations $\{x = g_1(u_1), y = g_2(x, u_2)\}$ is autonomous. Then $y = g_2(x, u_2)$ continues to hold under *any* atomic intervention that replaces the right-hand side of $x = g_1(u_1)$ with a constant, *whether that intervention is exogenous or non-exogenous*. Given that exogenous atomic interventions exist, $y = g_2(x, u_2)$ continues to hold under some exogenous atomic interventions on X and hence is invariant. Thus, if exogenous and non-exogenous atomic interventions are possible, then autonomy is logically stronger than invariance, since it requires that the

[7] Pearl's interpretation of structural models, therefore, presumes determinism. However, it is possible to interpret structural equations in a manner that does not require this assumption (see Steel 2005).

[8] Recall that Pearl's atomic interventions differ from Woodward's interventions only in not necessarily being exogenous.

model continues to hold under both kinds of intervention, while invariance only requires that it holds under exogenous ones. In addition, if the practical necessity of intervening in our surroundings is the fundamental motivation for the manipulation theory of causal explanation, then it is clear that autonomy is preferable to invariance as a basic concept. That is, a generalization that held true *only* under exogenous interventions (e.g., *only* within a randomized experiment) would be of little use for guiding practical interventions that, for the reasons given above, are typically *not* exogenous.

15.4.3 Explanation Without Invariance

Finally, let us reconnect this discussion with the objection, raised by Mitchell and Reiss, that for some especially complex systems, a causal model can be explanatory and yet fragile rather than invariant under intervention. This objection clearly presupposes that invariance under intervention is not the only possible grounds for judging that a causal model to be accurate. For instance, it may be possible to trace a mechanism from X to Y even when no intervention on X has been carried out. Such reasoning might also be combined with analysis of statistical data from non-experimental studies. Suppose that such means have produced a causal model M, and it is hoped that M could be useful for designing effective policies, as in the case of the CHANS models described at the end of Sect. 15.2. Suppose however, that there are real concerns that interventions being contemplated might alter background conditions in unexpected ways so as to make M an unreliable predictor of the results of the intervention. In other words, M might not be invariant under the contemplated interventions. In its original form, Woodward's theory of explanation could say little more than, "Well, M is explanatory if it is invariant, but not if it isn't." In this section, I suggest how a manipulation approach can be extended to draw distinctions between more and less explanatory causal models in situations wherein all of the models in question fail to be invariant.

To see how this could work, recall Woodward's key idea that invariant generalizations are explanatory because they enable us to answer a range of *what-if-things-had-been-different question*s. If the independent variable were set to that value, then the outcome would be this; if it were set to this other value, the outcome would be that instead, etc. But suppose that the generalization is not invariant because the interventions in question would alter background conditions upon which the truth of the generalization depends. In this case, the generalization can be conceived of as being associated with an "unless" clause: outcomes depend on these variables in this way, *unless* the intervention interacts with contextual factors in the set C. Such generalizations correspond to what are typically referred to in the philosophical literature as "ceteris paribus laws." Since I do not wish to engage in discussions

about laws of nature here, I will use the term "ceteris paribus model," leaving aside the question of whether the models in question deserve to be classified as laws.[9]

A ceteris paribus model may fail to be invariant yet be useful for guiding conditional plans. Doing this involves two things: (1) the associated unless clause lists the most likely contextual factors with which the intervention can interact and thereby disrupt the generalization and (2) the generalization, or model, indicates some format that could potentially incorporate those factors and examine their effects. Consider a first step in a conditional plan that generates an "unpleasant surprise," that is, a result that was contrary from what was expected and hoped for, and suppose that this has happened because the intervention unexpectedly altered some contextual factor. If (1) obtains, then it is likely that this contextual factor is included in the purview of the associated unless clause, and if (2) obtains, the causal generalization or model upon which the original action was premised suggest a means for explicitly including this omitted contextual factor into the analysis. In this case, a model may fail to be invariant with respect to a particular intervention, but may include resources for using the information inherent in the "unpleasant surprise" to make a model that is invariant with respect to some further set of interventions or at least more nearly so.

A pair of contrasting examples may be helpful for illustrating the idea here. Consider the causal generalization that presumably motivated the decision to make Wolong a national nature reserve: panda habitat degradation is largely due to commercial logging. This generalization may have been true to a certain extent, but it did not lead to accurate expectations about the results of declaring Wolong a national nature reserve. In this case, the generalization was apparently not associated with any "unless" clause that mentioned the potential effect of tourism, and even if it was, there was no format for incorporating that factor, once recognized, into the analysis. Compare this to the agent-based CHANS models described above. Consider (2) first. The agent-based models developed by CHANS researchers are capable of extension and further elaboration through the addition of further causal factors. In addition, by painstakingly reconstructing an interconnected web of causal mechanisms, these models suggest relevant factors that may alter causal relationships. For instance, the total size of the local population has an effect on the quantity of wood harvested, but the size of that effect is crucially dependent on a number of factors, including the number of individuals per household. The An et al. model described above includes submodels that represent factors that influence individual decisions about whether to found a new household. Such a model, therefore, draws explicit attention to the possibility that policy actions which inadvertently promote factors that encourage the establishment of new households can result in increased wood harvesting.

[9]Woodward critiques the notion of a "ceteris paribus law" and suggests that it can be replaced by his concept of an invariant generalization (Woodward 2002). As should be clear from this section, I regard that position as a mistake.

Tracing an interacting web of mechanisms can make a model better suited for guiding conditional plans, even when that model fails to be invariant under a contemplated intervention. In this situation, I suggest that models satisfying (1) and (2) should, other things being equal, be regarded as more explanatory from a manipulationist approach than those that do not. Since conditional plans are a type of intervention, this suggestion fits the core manipulationist emphasis on the close link between intervention and explanation. There is also an interesting connection here with mechanism approaches to explanation (Salmon 1984; Machamer et al. 2000; Craver 2007). Such approaches typically emphasize the value of highly detailed depictions of complex chains of causal interactions. But why is more fine-grained detail better for explanation? Philip Kitcher, for instance, argues that too much "gory detail" can obscure rather than illuminate key explanatory factors and relationships (Kitcher 1984). The expanded version of the manipulation approach just suggested provides one suggestion for how additional detail can improve explanation. The additional details can suggest what sorts of changes in background conditions might alter the causal structure, and constructing a model that include so many details often involves devising a format capable of incorporating further factors and sub-processes, which in turn can be the basis for a type of intervention that is particularly important for complex systems, namely, the conditional plan.

In sum, I suggest that the concept of a ceteris paribus model can be useful for thinking about how a manipulation approach to explanation can be useful for cases in which genuine scientific doubts exist about whether a seemingly well-confirmed causal model is invariant under a contemplated intervention. Given this approach, the two models might both fail to be invariant, yet one might provide a better causal explanation than the other from a manipulationist perspective.

15.5 Conclusions

In this chapter, I have proposed that a manipulation approach can more adequately account for explanations in complex systems, such as CHANS, if it takes a practical rather than an experimental concept of intervention as its basis. This approach is more in keeping with the motivation given for the manipulation theory that emphasizes the connection between explanation and practical action. Finally, it enables the relationship between scientific explanation and conditional plans to be explored from a manipulationist approach, thereby extending the relevance of that approach to cases in which doubts about the invariance of causal models are a common concern.

Acknowledgments I would like to thank Jim Woodward for helpful correspondence, and Lindley Darden and other members of the University of Maryland philosophy department for helpful comments on an earlier draft.

References

An, L., Mertig, A. G., & Liu, J. (2003). Adolescents' leaving parental home: Psychosocial correlates and implications for biodiversity conservation. *Population and Environment: A Journal of Interdisciplinary Studies, 24*(5), 415–444.

An, L., Linderman, M., Qi, J., Shortridge, A., & Liu, J. (2005). Exploring complexity in a human–environment system: An agent-based spatial model for multidisciplinary and multiscale integration. *Annals of the Association of American Geographers, 95*(1), 54–79.

Chen, X., Lupi, F., An, L., Sheely, R., Viña, A., & Liu, J. (2012). Agent-based modeling of the effects of social norms on enrollment in payments for ecosystem services. *Ecological Modeling, 229*(24), 16–24.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Henderson, D., & Horgan, T. (2001). Practicing safe epistemology. *Philosophical Studies, 102*(3), 227–258.

Kitcher, P. (1984). 1953 and all that: a tale of two sciences. *The Philosophical Review, 93*(3), 335–373.

Lempert, R., Popper, S., & Bankes, S. (2003). *Shaping the next one hundred years: New methods for quantitative, Long-term policy analysis*. Santa Monica: RAND.

Lempert, R., Groves, D., Popper, S., & Bankes, S. (2006). A general, analytic method for generating Robust strategies and narrative scenarios. *Management Science, 52*(4), 514–528.

Liu, J., Linderman, M., Ouyang, Z., An, L., Yang, J., & Zhang, H. (2001). Ecological degradation in protected areas: The case of Wolong nature reserve for giant pandas. *Science, 292*(5514), 98–101.

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H., & Taylor, W. W. (2007a). Complexity of coupled human and natural systems. *Science, 317*(5844), 1513–1516.

Liu, J., Dietz, T., Carpenter, S. R., Folke, C., Alberti, M., Redman, C. L., Schneider, S. H., Ostrom, E., Pell, A. N., Lubchenco, J., Taylor, W. W., Ouyang, Z., Deadman, P., Kratz, T., & Provencher, W. (2007b). Coupled human and natural systems. *AMBIO: A Journal of the Human Environment, 36*(8), 639–649.

Liu, J., Li, S., Ouyang, Z., Tam, C., & Chen, X. (2008). Ecological and socioeconomic effects of China's policies for ecosystem services. *Proceedings of the National Academy of Sciences, 105*(28), 9477–9482.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

Mitchell, S. (2009). *Unsimple truths: Science, complexity, and policy*. Chicago: University of Chicago Press.

Neopolitan, R. (2004). *Learning Bayesian networks*. Upper Saddle River: Prentice Hall.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge: Cambridge University Press.

Popper, S., Lempert, R., & Bankes, S. (2005). Shaping the future. *Scientific American, 292*(4), 66–71.

Reiss, J. (2009). Causation in the social sciences. Evidence, inference, and purpose. *Philosophy of the Social Sciences, 39*(1), 20–40.

Richardson, T., & Spirtes, P. (1999). Automated discovery of linear feedback models. In C. Glymour & G. Cooper (Eds.), *Computation, causation, and discovery* (pp. 253–302). Menlo Park: AAAI Press.

Salmon, W. (1984). *Explanation and the causal structure of the world*. Princeton: Princeton University Press.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Cambridge: MIT Press.

Steel, D. (2005). Indeterminism and the causal Markov condition. *British Journal for the Philosophy of Science, 56*(1), 3–26.
Steel, D. (2006). Comment on Hausman and Woodward on the causal Markov condition. *British Journal for the Philosophy of Science, 57*(1), 219–231.
Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford: Oxford University Press.
Taylor, P. (2005). *Unruly complexity: Ecology, interpretation, engagement*. Chicago: University of Chicago Press.
Woodward, J. (1999). Causal interpretation in systems of equations. *Synthese, 121*(1–2), 199–257.
Woodward, J. (2001). Law and explanation in biology: Invariance is the kind of stability that matters. *Philosophy of Science, 68*(1), 1–20.
Woodward, J. (2002). There is no such thing as a Ceteris Paribus law. *Erkenntnis, 57*(3), 27–52.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.
Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology, 13*(2), 219–240.

Chapter 16
Biology and Natural History: What Makes the Difference?

Aviezer Tucker

Abstract The distinction between the historical and theoretical sciences runs through rather than between academic disciplines. Some branches of biology, like evolutionary biology, genetics, and phylogeny, are historical. Other branches that study types of biological objects in theoretical contexts independent of space and time, like biochemistry, anatomy, and cell biology, are theoretical. The historical sciences infer origins, common causes of information-preserving effects in the present: phylogeny and evolutionary biology infer the origins of species from homologies, genome sequences, and fossils. Historians of humanity infer past events and processes from their effects in the present, documents, material remains, visual depictions, and recordings. I explicate the ontological distinction between types and tokens that lies at the basis of the distinction between the historical and theoretical sciences. Then, I demonstrate how this distinction leads to different epistemic methodologies for the historical and theoretical sciences. Finally, I address the heuristic issue of explanation in the historical and theoretical sciences. I distinguish two senses of explanation in the historical sciences: a strict one that explains the evidence and looser and context-dependent one that explains representations of historical events. Explanations in the second sense, that is, explanations of events, explain the evidence in the first sense. The evidence and events that the historical sciences explain are tokens. The theoretical sciences, by contrast, are not interested in token evidence and events, but in types of replicated evidence and repeated events.

Keywords Philosophy of historiography • Type-token distinction • Historical sciences • Theoretical sciences • Explanation

A. Tucker (✉)
University of Texas, Austin, The Energy Institute, Flawn Academic Center, FAC 428, 2 West Mall C2400, Austin, TX 78712, USA
e-mail: atucker@energy.utexas.edu

M.I. Kaiser et al. (eds.), *Explanation in the Special Sciences: The Case of Biology and History*, Synthese Library 367, DOI 10.1007/978-94-007-7563-3_16,

16.1 Introduction

In this chapter I clarify the distinction between the historical and theoretical sciences and show that biology is composed of branches that are historical and resemble other historical sciences, like the study of the human past (historiography) and historical linguistics, other branches are theoretical and have more in common with theoretical physics than with phylogeny and evolutionary biology. I first explicate the ontological distinction between types and tokens that lies at the basis of the distinction between the historical and theoretical sciences. Then, I demonstrate how this distinction leads to different epistemic methodologies for the historical and theoretical sciences. Finally, I address the heuristic issue of explanation in the historical and theoretical sciences. I distinguish two senses of explanation in the historical sciences, a strict one that explains the evidence and a looser and context-dependent one that explains representations of historical events. Explanations in the second sense, that is, explanations of events, explain the evidence in the first sense. The evidence and events that the historical sciences explain are tokens. The theoretical sciences, by contrast, are not interested in token evidence and events, but in types of replicated evidence and repeated events.

I argue that the distinction between the historical and theoretical sciences runs through rather than between academic disciplines. Some branches of biology that concern the past, like evolutionary biology, genetics, and phylogeny, are historical. Other branches that study types of biological object in theoretical contexts independent of space and time, like biochemistry, anatomy, and cell biology, are theoretical. The historical sciences infer origins, common causes of information-preserving effects in the present: phylogeny and evolutionary biology infer the origins of species from homologies, genome sequences, and fossils. Historians of humanity infer past events and processes from their effects in the present, documents, material remains, visual depictions, recording, and so on. Darwin himself compared species to languages and phylogenetic inference to the inference of ancestral languages (cf. Tucker 2011).

I show that all the historical sciences attempt to infer consecutively the existence of common cause tokens, the information transmission nets that connect them with their present effects that preserve the information, and the properties of these common causes-information sources. Information transmitted from common origins may be lost through an evolutionary process and may be mixed with a good deal of noise. When information is lost, there is not much that the historical sciences can do. Their art is finding it when it is tacit or nested and separating it from the accompanying noise that original information signals accumulate over their long route from transmitting event to received evidence. One of the basic tasks of historians like Darwin is to distinguish patterns that result from natural regularities, such as homoplasies, which have separate causes, from those that result from preserving information from common causes, such as homologies. The first is noise; the second is a valuable information signal. This holds true irrespective of the aspect of the past the historian attempts to infer, whether it is human history, natural history,

evolutionary history, geological history, the history of languages, the history of the universe, or the history of a village. Archaeology infers the common causes of present material remains, and cosmology infers the origins of the universe.

By contrast, the theoretical sciences are not interested in any particular token event, but in types of events: physics is interested in the atom, not in this or that atom at a particular space and time; cell biology is interested in the cell or in types of cells (brain, skin, muscle, etc.), not in this or that token cell; physiology is interested in the study of different species, not in their historical family relations. Generative linguistics studies "Language," not historical languages that were spoken by particular groups of people in particular places at particular times. The theoretical sciences are interested in inferring regularities of types from replicated experiments.

16.2 Types and Tokens

I argue that the historical sciences are distinctly interested in inferring common cause *tokens*. The theoretical sciences, by contrast, are distinctly interested in inferring common cause *types*. These mutually exclusive, though not exhaustive, scientific goals necessitate entirely different methodologies.

Pierce's distinction between types and token has been useful for the analysis of myriad philosophical problems in metaphysics and the philosophies of mind, language and science, ethics, and aesthetics (Wetzel 2009, pp. 1–2). For current purposes, it is sufficient to make the fairly uncontroversial claim that, as particulars, tokens necessarily occupy a unique spatial-temporal location, whereas types, as abstracts, do not. "A token event is unique and unrepeatable; a type event may have zero, one or many instances" (Sober 1988, p. 78). For example, cell, mitochondria, and chromosome are types. This cell in my body and the mitochondria and chromosomes in it are tokens of these types. Cell as a type does not exist in space and time. The brain cells that participate in the generation of this sentence now had a beginning and will have an end and they exist in particular locations in my skull. If we have the same type of parents, we are human; if we have the same token parents, we are siblings.

Historical events, like the extinction of the dinosaurs, the genetic mutation that created Homo sapiens, the fall of the Roman Empire, and the American Civil War, were the common causes of myriad effects in the present that preserved information about them, such as fossil records, geological structures, DNA sequences, population distributions, documents, and material remains. Historians attempt to infer representations of these events and processes from their contemporary information-preserving effects.

Most historical events are also explosions of information; they transmit information mostly unintentionally in all directions. Much of this information decays quickly and, *ceteris paribus*, the more time passes the greater is the decay. The

historical sciences have developed a toolbox of methods to interpret information signals from the past that reach the present and infer from them representations of their common causes. By contrast, theoretical scientists attempt to generate theories about the causes and effects of types such as evolution, genes, revolutions, empires, and Civil Wars. Since types and tokens do not have to share properties (Wetzel 2009, pp. 118–119), theories about types do not have to be about their tokens, for example, cell theory does not have to be about a particular cell and evolutionary theory does not have to be about a particular event in the history of evolution. Vice versa, particular tokens can illustrate discussions of their type, but cannot directly confirm or refute theories whose building blocks are types.

There is a high correlation between the goals, methods, practices, and paradigmatic success stories of the historical and theoretical sciences and, respectively, the inferences of common cause tokens and types. The historical evidence for common cause tokens is made of other tokens, for example, fossils, mineral deposits, sequenced genomes, documents, and material remains, all existing here and now in space time. The theories that connect the evidence with the inferred representations of their information sources are information theories about the transmission of information in various media in time.

By contrast, the theoretical sciences infer common *types* of causes from *types* of effects. Often these common cause types are "hidden" or are not obvious. Their inference is therefore a discovery. Common cause types can be theoretical entities, such as universal constants, types of particles or molecules, types of cells, types of viruses, types of xenophobia, or types of rent seeking. Types of things like standardized and replicated results of laboratory experiments, medical symptoms, and social pathologies are explained as the effects of these types of causes. As theoretical types, neither the causes nor their effects have specific space or time. The theoretical sciences are not interested in the results of any particular experiment, only in the type of standardized results that replicated experiments typically generate. Scientists build theories from correlations between such types.

Philosophers who worked on the problem of the inference of common cause have only rarely noted the crucial distinction between the inference of common cause types and tokens. Arntzenius (1992, pp. 230–231) correctly stressed the frequent confusion between types and tokens in philosophic discussions of the inference of common cause. Reichenbach's (1956) examples of inferences of common causes are of tokens. But his characterization of correlations between effects of common causes as correlations between statistical frequencies is of types. Reichenbach's seminal influence on the framing of the philosophic discussion of the problem of inference of common cause and his influential "principle of the common cause" that conflated types with tokens have led to a great deal of confusion in subsequent discussions and consequently, in my opinion, to further difficulties in distinguishing the historical from the theoretical sciences (Tucker 2007). Within the context of analyzing inferences of common cause tokens in cladistics, Sober (1988, 2001, p. 339) cleared some of the conflations by clearly stating that he was discussing the inference of common cause tokens. Cleland (2002) distinguished, like me, the historical sciences by their study of tokens rather than of types. I also argue that

distinguishing the historical from the theoretical sciences requires the presentation of a clear distinction between the inference of common cause tokens from tokens of effects and of common cause types from types of effects.

16.3 The Historical Sciences

Some processes tend to preserve, in their end states, information from their initial state more than others. The historical sciences first look for the results of such information-preserving processes and within them for the properties of evidence that tends to preserve information about their origins. Information theories assess the extent to which different properties tend to preserve more information than others and try to determine their approximate *reliability,* in its probabilistic sense, or *credibility,* the same concept as used in jurisprudence (Friedman 1987), or *fidelity*, the term favored by textual critics to figure the reliability of texts (Maas 1958). Information theories also help to extract nested information from evidence: if some information is *nested,* it can be inferred only with the aid of theories that link properties explicit in the information signal with information that is "nested" in it (Dretske 1981, pp. 71–80). For example, historians and detectives use information theories to infer token events from what is *missing* from the evidence, such as the dog that did not bark in the night. Schatzki (2006) considered what I call information theories too commonsensical or trivial to be considered as such. Yet, these simple theories about the transmission of information in various media over time, its admixture with noise and mutations, have been widely known and utilized for more than two centuries. If they seem trivial to us today, it is a reflection of how entrenched they have become, not of their lack of theoretical vigor and rigor. The scope of change that such simple theories have brought about in the historical sciences is truly amazing and is just as profound as that of any of the major scientific theories that are associated with scientific revolution and paradigm change. Just consider how different is the received view of the history of humanity, the founding texts of the great religions, language, the species, planet Earth, and the cosmos itself today in comparison to what the most rational people believed at the dawn of the information revolution in the historical sciences around 1780. All these revisions in our view of the past, and consequently in our view of ourselves and our place in history, are almost exclusively due to mostly simple information theories.

The estimation of the reliability of evidence may also involve the examination of evidence for the information transmission chains that transmitted information to the evidence. If testimonies about historical events are separated by time or space from those events, historians look for independent evidence for links on the information transmission chains that may have connected the events with them. The selection of historical evidence is, according to its information-preserving qualities, theory-laden, and it is bootstrapped by historiographic knowledge of the chains that transmit information in time. For example, phylogenic inferences are made on the basis of nonfunctional parts of the genome that do not affect fitness and survival, are

therefore not subject to evolutionary pressures, and are more reliable. Evolutionary biologists also look for intermediary forms in the fossil records to connect sets of contemporary species with their hypothetical ancestors.

Evidence in the historical sciences always includes correlations or similarities between reliable documents, testimonies, languages, material remains, species, genomes, and so on. The first stage of inferring representations of the past is the comparison of the likelihoods of the units of this evidence given *some common cause token* and given *separate causes*: *The common cause token hypothesis* asserts that the information-preserving properties of the evidence preserve information about some common cause or causes *without specifying the properties of that common cause*, for example, that the similarities between all the species that belong to the monkey family, including us, preserve information about some common ancestor species without specifying the properties of that species or speculating about the family relations of the various species of monkeys. For the token common cause hypotheses to be accepted, the likelihood of the evidence, given some token common cause, must be higher than its likelihood, given separate causes. Usually, the properties of the separate causes are specified, unlike the properties of the common cause. For example, suppose that the evidence consists of a set of testimonies that share telling that a certain king murdered his father to inherit the throne. The token common cause hypothesis would simply claim that the testimonies preserve information transmitted from a common cause event without specifying its properties; it does not claim that the king indeed murdered his father or that a group of false witnesses colluded to frame the orphan, just that some common cause was at work. The separate causes hypothesis specifies the separate causes, for example, that one witness hated the new king, that another wanted to put on the throne the next in line of succession, yet a third had an interest in discrediting the dynasty, a fourth lost his job as a result of the succession, and so on, as long as there was no single event that all the evidence preserves information about. Often the separate causes are tokens of the same type. For example, if we consider the likelihood of a set of species that share some characteristics, the common cause hypothesis would assert that some common ancestor is the common source of the similarities; a separate causes hypothesis would suggest these were all separate adaptations to a similar type of environmental conditions in different places at different times, that is, different tokens of that type. For example, various species that have no fur lost it in separate adaptations to warming climates.

It is often difficult but also unnecessary to assign *precise* quantitative likelihoods to the evidence given token common or separate causes because the common cause and separate causes hypotheses are exhaustive and mutually exclusive, proving that the evidence is highly unlikely given that one of the hypotheses implies that the other must be the case.[1]

[1]Formally, if E_1& E_2 & … E_n are units of evidence that share certain properties and C is some common cause, the likelihood of the evidence given the common cause hypothesis is:

$Pr(E_1 \& E_2 \& \ldots E_n|C) = Pr(E_1|C) \times Pr(E_2|C) \ldots . \times Pr(E_n|C)$.

Historians assess the prior probabilities of the token common cause and separate causes hypotheses by looking for independent evidence for whether information transmission chains that extend backwards from the units of the evidence could or could not have intersected. The likelihood of the evidence given separate causes often reflects the functions of the shared properties of the evidence that select them for transmission. Often, the shared properties of the evidence have the same *type* of function and the same *type* of cause. For example, unrelated cultures invented agriculture, domesticated animals, seafaring, and writing in reaction to similar challenges to human existence, and in the case of writing, the need to keep record of taxation in centralized monarchies. Eyes, wings, and fins emerged and then were reproduced several times in the history of life.

Vice versa, evidence that has no function or that is even dysfunctional, such as testimonies that present the witnesses in a negative light or run counter to their interests or political or ideological commitments, usually radically decreases the likelihood of the evidence given separate causes. Evolutionary neutral traits and genes are good indicators of common ancestry in phylogeny. The likelihood that any one species would retain them is low; the likelihood that a set of species would share these rudiments without common ancestry is vanishing.

When the likelihood of each unit of evidence given separate causes is low, the effect of multiple members, such as similar testimonies and species, is to decrease their likelihood exponentially, given separate causes. Therefore, historians and biologists devote great efforts to the discovery of multiple units of evidence.

If the likelihood of the evidence given some token common cause is significantly higher than its likelihood given separate causes, historians attempt to determine how the information was transmitted from the common cause to all the units of evidence. Five alternative types of models are possible:

1. *A single source, like a historical event or species, is the common cause of all the information-preserving evidence* (see Fig. 16.1): The modeling of the history of the transmission of information would be treelike and be composed of V-like intersections.
2. *Multiple ancestral common causes* (see Fig. 16.2). All the units of evidence are the information-preserving effects of the same set of common causes. For example, suppose there are two sources of evidence in the present for an event in ancient history. Both were written hundreds of years after the event by historians who had access to the same two primary sources that had since been

---

Reichenbach mentioned the same equation (1956, pp. 157–167). However, the meaning of C in my equation is of *some token common cause* without specifying its properties, whereas Reichenbach's concept of the common cause was ambiguous, not to say confused, about whether it is a type or a token and whether or not its properties are specified. Formally, assessing the likelihood of the evidence given separate causes (S_1, S_2, $\ldots S_n$), background knowledge B, and their respective prior probabilities can be expressed by the following equation:

$$\Pr(E_1 \,\&\, E_{2\ldots} \,\&\, E_n|S_1 \,\&\, S_{2\ldots} \,\&\, S_n) = [\Pr(S_1|B) \times \Pr(E_1|S_1)] \times [\Pr(S_2|B) \times \Pr(E_2|S_2)] \ldots \times [\Pr(S_n|B) \times \Pr(E_n|S_n)]$$

Fig. 16.1

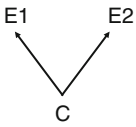


Fig. 16.2

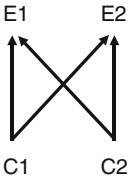


Fig. 16.3

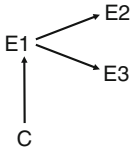
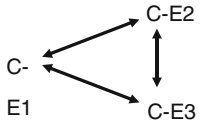


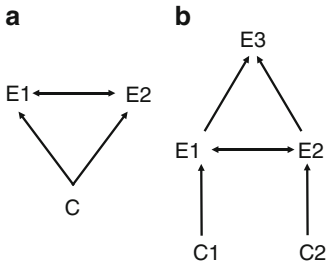
Fig. 16.4



lost. Likewise, a set of contemporary species or languages can be the result of hybridization between the same two or more distinct species or languages. The modeling of the history of the transmission of information would be bush-like with W-like intersections.

3. *The common cause may be one of the units of evidence* (see Fig. 16.3). For example, historians may have three sources for an ancient event. But two of the sources may simply be copies of the third one. An ancestor species may survive to the present day alongside species that descended from it. The model of the history of the transmission of information would be < like.
4. All the units of the evidence may have affected each other, for example, if the witnesses colluded together to produce the same testimony or if all the species could have interbred with each other to generate the similarities we find. The model of the history of the transmission of information would look like a web composed of Δ-like intersections where information is transmitted between all the units (see Fig. 16.4).
5. *Combinations of types 1 or 2, with types 3 or 4* (see Fig. 16.5). The evidence had one or more common information sources, as in type one and two, but the information flows from them intersected on their way to the various units of evidence. For example, an original species may have mutated into several

Fig. 16.5



species that may later have interbred with each other and thereby brought about the current distribution of characteristics across related species. Similarly, several ancient oral traditions were combined to create a unified text. The editors sought not just to preserve the ancient sources but also to create a coherent text that may also have fitted their political interests. The result is a complex text like that of the bible, uniting different sources that shared original sources and then influenced each other. The model of historical information transmission would include H- or X-like intersections.

Another way of saying that the similarities between the units of evidence reflect a common information source or sources but no mutual influences, corresponding exclusively with the first V- or second W-models of information transmission above, is by describing the evidence as *independent*.

Independence of evidence is the absence of intersection between the information transmission flows that connect the units of evidence with their common source or sources.

When independent evidence for the information transmission is scarce, more than one of the five possible information transmission hypotheses may confer equal likelihoods on the evidence. For example, though it is highly likely that the Indo-European languages had common rather than separate causes, the evidence underdetermines whether it was a single language, proto-Indo-European, or whether several languages mutually influenced each other until they became very similar, before spreading around the globe through the Latin, Germanic, and Slavic language families while continuing to influence each other, as in the wave theory of language. Likewise in evolutionary biology, scientists have been able to determine that species must be closely related to each other without having sufficient information to determine how they are related to each other exactly.

Historians are able, in many cases, to infer which of the five possible information transmission hypotheses is the most probable. For example, textual critics proved that the standardized editions of the bible and Homer's epics had initially multiple common causes and then they influenced each other in the process of editing (an A-like version of model 5). In many cases, there is independent evidence for

links on the information transmission chains that connect events with evidence for the presence of a single or multiple common causes. Composite documents may preserve linguistic differences that indicate multiple common causes. Historians and textual critics look for discontinuities in style, conceptual framework, and implicit values, as well as for internal contradictions, gaps in the narrative if there is one, and parts that are inconsistent with the alleged identity of the author. Likewise, geneticists who examine the genome can spot parts that clearly were inserted through hybridization or through some unusual lateral form of transfer.

Frequently, the same information transmission theories that assist in the assessment of the reliability of the evidence also assist in proving whether there were single or multiple common information sources. For example, assuming the theory that the mutation rate of the names of God is lower than those of other words, that the reliability of the names of God is higher than that of other parts of edited documents, it is possible to analyze parts of the bible into its constituent parts, as the first biblical critics did.

When one of the possible five information transmission models clearly increases the likelihood of the evidence more than its alternatives, scientists may attempt to infer the properties of the common causes. The evidence may not suffice to determine the properties of the common causes. For example, from the references in the Bible to two older books, the *Book of the Wars of Jehovah* and the *Book of Righteousness*, it is possible to infer that some of the materials in the Bible preserve information from those books, but there is insufficient evidence to determine hypotheses about them. Likewise, humans and apes had a single most recent common ancestor about six million years ago, but many of the character traits of that ancestor are unknown. Darwin was careful not to speculate about the properties of species that he considered to be the common ancestors of sets of species that shared common properties (Tucker 2011).

When there is sufficient evidence, historians compare the likelihoods of the evidence given competing representations of history. The prior probabilities of these competing specific common cause hypotheses follow their degree of coherence with established historiography and internal coherence (cf. Kosso 2001, pp. 106–108).

16.4 The Theoretical Sciences

The theoretical sciences typically test hypotheses that connect types of causes with types of effects. In the experimental theoretical sciences, the replication of experiments and the standardization of their results transmute tokens into types. In observational, nonexperimental, theoretical sciences, like parts of medicine, astronomy, agronomy, and the social sciences, the transition from tokens to types of causal relations is achieved via the averaging of causal effects. Observational, nonexperimental, scientists measure the average effects of the types of factors they study, not effects on this or that individual unit like a person or a cell or a plant.

Some hypotheses correlate just one type of cause with one type of effect. But since causes usually have multiple effects, more hypotheses connect one type of cause with multiple types of effects. For example, smoking is causally related not just to lung cancer but also to stroke, heart attack, and so on. Medicines, likewise, are connected to effects on hosts of symptoms. Genes are often responsible for more than a single trait, alone or in combination with other genes. Such inferences of common cause *types* proceed in two stages:

In the first stage, theoretical scientists need to prove that the correlations between the types of effects, like elevated levels of cancer, stroke, and heart attack, are more likely given the hypothetical common cause type like smoking than given separate types of causes. Theoretical scientists specify the properties of the common cause type they propose (e.g., gene, smoking, unemployment), but do not specify the properties of the alternative types of separate causes (separate causes should not be confused with confounders whose properties are specified and therefore are alternative common causes types). The method for achieving a significant gap between the likelihoods of the correlations given the common cause type and the unspecified separate causes that may be many, varied and unknown, is the random assignment of members to two populations to make them nearly identical in sharing the same types of (unknown or unspecified) variables with the exception of the common cause type (sometimes called the treatment) that all the members of one group share and none of the members of the other (control) group are affected by. Significant differences between the two populations are likely, then, to be the result of that common cause type. For example, scientists may choose a random sample of a population (of people, animals or crops), divide it into two randomly assigned equal and sufficiently large groups, whose only difference is the presence or absence of the hypothetical type of common cause, like a particular gene, virus, medicine, or a social independent variable like educational level or marriage. Then, theoretical scientists measure the difference in the putative effects between the two populations and see if there is a significant gap between the two groups. If there is such a significant gap, the correlation between the types of effects is more likely given the common cause type than given unspecified separate types of causes.

The goals and methods of the historical and theoretical sciences are the mirror image of each other. The historical sciences are interested in inferring common cause tokens that are information sources from their information-preserving effects. The theoretical sciences infer common cause types whether or not they are sources of information. The historical sciences compare, in the first stage, the likelihoods of information-preserving evidence given a *common cause token whose character traits are unknown*, and separate *cause tokens whose character traits are specified and are often separate tokens of the same type* (e.g., either human or apes had some common ancestor or their common features are the result of separate adaptations to living in similar environments). In the theoretical sciences the character traits of the proposed common cause type are specified. But the character traits of the alternative separate causes are not specified (lung cancer and stroke are the results of smoking or of myriad other unspecified causes).

The historical sciences try to prove that their evidence is more likely given a common cause token than separate causes *using information theories*. By contrast, theoretical scientists use statistical randomization techniques to prove the causal relevance of common cause types. Types of effects in the theoretical sciences need not preserve information about their causes.

Once theoretical scientists establish that the evidence is more likely given the specified common cause type than given separate types of causes, they attempt to find the exact causal relations, which may be complex. One or more of the effect types may affect the others requiring the construction of multicollinear and interactive models. Some theoretical scientists like a number of philosophers of science further demand the discovery of mechanisms as a necessary condition for the inference of causal relations. Suffice it to say that in this second stage, theoretical scientists attempt to use a variety of tools to infer increasingly precise causal nets. Additionally, theoretical scientists need to control for confounding hidden variables types that may cause both the common cause type and its effects. In the experimental theoretical sciences, experimental designs control types of variables to isolate their effects. When such experiments are impossible, the theoretical sciences resort to statistical observational data analysis. If successful, using statistical control techniques to hold different variables constant while measuring others, theoretical scientists conduct multivariate regression analyses that generate equations and multi-equation models, and causal maps that measure levels of causal influence that each variable exerts on the others on a scale from 0 to 1. The posterior probabilities of causal hypotheses need not be affected by the means by which the researcher infers them, whether control of types of causes is achieved in the laboratory by design or through other "natural" experimental methods or by using statistical analysis of observational data.

16.5 Explanations of Evidence and Events

The final question I address is of the analysis of explanation in the historical sciences and its distinction from explanation in the theoretical sciences. The philosophical literature is extensive and rich with different and competing accounts of explanation. Kitcher and Immerwahr (this volume; a similar typology can be found in Kitcher 1989) offered a tripartite division of types of explanation: *strict* explanation (Hempel 1965), *Orthodox* explanation (i.e., answering a why question), and *liberal* explanation (i.e., context dependent and answering many different types of questions). They endorse the third model and apply it to historiographic explanations of events. As they note, what they call the Orthodox model, associated with van Fraassen (1980, Chap. 5), shares with the liberal model its pragmatic sensitivity to context. I agree that historiographic explanations of past events and processes are context dependent and answer different types of questions liberally. Indeed, one of the first things I did as a philosopher of historiography was to apply van Fraassen's pragmatic model of explanation to the explanation of the

representations of historical events (Tucker 1993). Hart and Honoré's (1985) pragmatic account of causation and explanation in jurisprudence has also been very applicable for the analysis of causal explanations in historiography. But neither Kitcher and Immerwahr nor the advocates of a variety of stricter models of explanation in the philosophy of historiography have considered that the explanatory relationship between historiography and its evidence, mediated often by information transmission theories, may fit a stricter model of explanation.

I group philosophic models of explanation in two rough clusters. One cluster includes strict, formal, and context-independent models of explanation. Models of explanation in the other cluster pay attention to the pragmatic, contextual, hermeneutic, or even psychological aspects of explanation, attempting to analyze how people remove puzzlements and conundrums by explaining things. This second cluster is looser and less formal because such models of explanation must consider its context dependency; the same statement can be explanatory or not depending on the context of inquiry, its audience, and so on.

As my argument so far implies, in the historical sciences, the first stricter type of explanation is of the evidence; the second looser type of explanation is of representations of historical events. In the first type of explanation, historiographic hypotheses increase the likelihood of the information-preserving aspects of the evidence in the three stages that I outlined above. The historiographic hypotheses, that are the best explanation of the evidence in the strict sense of increasing its likelihood, may be in some contexts explanatory in the second, loose, sense by answering questions about historical events. For example, the best explanation, in its strict sense, of the geologic and fossil evidence from the end of the Jurassic era, that is, the disappearance of dinosaur fossils from the geological record and the geological composition of the K-T boundary, is that a meteor hit the earth and caused greenhouse conditions that caused the extinction of the dinosaurs. This strict explanation of the fossil and geological evidence is itself an explanation of an event in the loose sense in the context of answering why did the dinosaurs become extinct at the particular time they did, rather than earlier (cf. Cleland 2011)? Similarly, when historians explain, in the loose sense, events in human history by adducing causes, or colligating them, or just by adding information, these explanatory historiographic statements in the loose sense explain historical evidence in the strict sense of increasing its likelihood. For example, if somebody explains why the Renaissance took place in Northern Italy rather than elsewhere in Europe, by mentioning the ubiquitous presence of remnants from antiquity that served as models for emulation, this also explains present evidence in the form of art works from the ancient world and the Italian Renaissance and their similarities as well as documentary evidence from the Renaissance era about renewed interest in classical antiquity during that era. When historians explain the victory of the Greeks over the Persians in the Persian Wars, answering why the Greek defeated the Persians whereas numerous other political entities succumbed to the superior military might of the Persian Empire, by the superior training, morale, and tactics of the Greek citizen-soldier, the hoplite, it is also the explanation of some of the documentary and archaeological evidence for the Persian Wars which is presently available to us. If an art historian

explains, by colligation (cf. McCullagh 2009), some of the properties of a Rothko painting from the 1950s that differ from those of paintings done before Rothko by saying it is an excellent example of abstract expressionism, the historian also explains a broad scope of paintings from that period that are observable now and share abstract expressionist properties.

In a strict sense, explanatory hypotheses in the historical sciences increase the likelihood of the information-preserving evidence more than competing hypotheses. The theories that support the assessment of likelihoods given competing hypotheses are usually information theories. The strict sense of explanation that I deploy is partly overlapping with Salmon's (1971) statistical relevance account of explanation. Since not all statistically meaningful correlations and not all hypotheses that increase the likelihood of the evidence indicate causation or explanation, and since statistical relevance can be a symmetric relation that does not distinguish the explanans (which explains) from the explanandum (what is explained), Salmon suggested first to add a causal-mechanical criterion that requires the transmission of a mark from the explaining cause to the explained effect (Salmon 1984) and then somewhat revised this criterion to that of the transmission of a conserved quantity (Salmon 1994). These criteria will not do for the historical sciences. The causal-mechanical condition is too strong and demanding because in many cases the historical sciences can first prove that there was a transmission of information from the hypothetical source event to the evidence in the present and then infer the existence and some of the properties of the hypothetical historical event, without being able to determine the mechanism by which the information was transmitted. For example, when Darwin made his famous phylogenic conjectures about the origins of species, he knew nothing of the genetic mechanisms through which information is transmitted from ancestor to descendant species. Nevertheless, he was able to prove that the similarities between descendant species are more likely given common than given separate ancestries. Somehow the information was transmitted, but he did not know how (Tucker 2011). To take another example, Polynesians have been cultivating the sweet potato in some pacific islands since prehistoric times. The potato is native to the Andean region of South America. The Polynesian word for sweet potato is very similar to the Andean word for the plant. The explanation of the botanic and linguistic information-preserving similarities between the plants and the words that refer to them is that sometime, way back in the mists of history, there was some contact between some Polynesians and South Americans during which the Polynesians acquired the Andean native plant and learned the word for it. The separate cause hypothesis would be that the sweet potato evolved independently in the Eastern pacific and that, by coincidence, the Polynesian word for sweet potato sounds very much like the Andean word. Since the likelihood of the evidence given separate causes is vanishing, the first hypothesis wins by default. Still, there is insufficient evidence to determine the *mechanism* of transmission of information from South America to the East Pacific. The nearest Polynesian islands to South America are the Easter Islands and they are 4,000 km away. Genetic evidence also suggests that the first Polynesian islands to cultivate the sweet potato, from which it spread to other islands, were even further away. At most, historical scientists can

offer possible mechanisms by which the transmission of information could have taken place (Montenegro et al. 2008). To take a final example, genome analyses have revealed that lateral transfer of genes between species that are not closely related have occurred quite often in the history of life, especially between unicellular organisms. How particular lateral transfer events took place is shrouded in the mists of the past. But the discovery of obviously foreign fragments of genome that resemble those of entirely unrelated species embedded in the genome of another organisms can only be explained by this transfer of information. Historical scientists debate the mechanisms of transfer, how to model these lateral transfers, and which one is the best, but they all agree that they happened (Velasco and Sober 2010).

While the requirement for a causal *mechanism* is too demanding for the historical sciences, a mere causal process that connects past events with present evidence is insufficient, too loose, as a model of explanation because some causal processes do not preserve information about their origins. For example, it is a priori true that all our ancestors lived at least until reproductive age and must have made a living somehow at least until they reproduced. This is a necessary condition of our biological existence. But though we usually receive information about when our immediate ancestors lived and what they did for a living, this information usually decays after a few generations. Though we are absolutely certain that our existence today is the result of a causal process that included some ancestors who lived at least until reproductive age in, say, the twelfth century, our mere existence today does not preserve any information about how our medieval ancestors survived at least until they reproduced. By contrast, if we point to particular information-preserving characteristics that we possess, for example, whether we are lactose tolerant or intolerant, and explain them by the four genetic mutations (one in Europe and three in Africa) that created lactose-tolerant adults or by their absence among our ancestors, we increase the likelihood of this characteristic and thus explain it in the strict sense, as well as infer a representation of some of the characteristics and origins our ancestors.

The main motivation for characterizing strict explanation in the historical sciences as increasing the likelihoods of information-preserving effects is epistemic. This model fits what historical scientists are actually doing to gain knowledge, inferring representations of token common causes from similarities between their information-preserving effects. This model does not ask metaphysical questions about the ultimate constituents of reality, about whether the universe is ultimately mechanistic in some sense, and whether all change is, in some ultimate sense, the transmission of conserved quantities. It does not offer a unified model of strict explanations in general because strict explanations in the theoretical sciences do not have to preserve information. Modestly, it models the epistemic function of strict explanation of the evidence in the historical sciences.

The explanations of historical events cannot be strict. Unlike historical evidence, historical events are never an observable given in the present. Phenomenologically, we may be conscious of some historical events as appearing as "facts" because they are so well entrenched in our web of beliefs. If there had been no Roman Empire, no Renaissance, no Industrial Revolution, or no Two World Wars, the

core of our web of beliefs would have been torn. The scope of evidence that is explained by hypothesizing those events is so vast, broad, and diverse and any alternative hypotheses is of such low prior probability that it is easy, especially for people who are not familiar with the tacit practices of historians, to accept them as simple facts and ask for their description rather than their representation and for a direct rather than inferred explanation. They can ask for the explanation of the end of the Second World War in Europe just as they can ask for directions to the 9th of May street in Paris (take a right as you leave the Gare de l'Est and you will be able to observe it right in front of you). But there are no given historical events as primitive facts that can be observed directly (*Pace* Glennan (2002, p. 350)). Historians cannot choose how to describe the "actual sequence" of historical events like those that culminated with the First World War at various levels and with more or less detail. "The description of a process" is not "guided by considerations of robustness and reliability" (Ibid). There is no history present that can be described. History is not here. Whenever we knock on history's door, we are told that it has already moved away and has at most left some of its belonging behind for us to inspect. A pathologist may choose to describe a cadaver that lies here on various levels and attempt to discover the mechanism that brought about the death. But a historian who wishes to do the same with the remains of Archduke Ferdinand has no cadaver to work with. Should the historian be lucky enough to receive the permission of the present members of the Hapsburg dynasty and the Catholic Church, it may be possible to examine the remains, by now bones, of the Archduke. But the bones are not likely to preserve much information about the process that led to the assassination of the Archduke and the end of the civilized modern era in Europe. Otherwise and beyond that, there are only the testimonies from a hundred years ago. These evidential scarcities and limitations dictate to the historians what can and cannot be known and represented about the past. If the evidence is bountiful and diverse, the historian may be able to represent a historical mechanism that may be necessary or contingent (or robust or fragile in Glennan's (2002) and Kim Sterenly's terminology). But for much of history, there is simply not enough evidence for inferring mechanical sequences of events. Since all historical events took place in the past, none of them is given or observable (with the exception of cosmological events that we see on our planet light years after they occurred). Consequently, explanations of representations of historical events must themselves explain, in a strict sense, historical evidence.

Historical token-token and theoretical type-type explanations in strict or loose senses are necessary for understanding the world we live in. Problems emerge only when philosophers confuse types with tokens in the context of modeling explanation. Three great and influential philosophers of science Gustav Hempel (1965), Hans Reichenbach (1956), and Wesley Salmon (1984) confused the explanation of types with that of tokens. Reichenbach and Salmon did so in their discussion of the inference of common cause. Reichenbach attempted to infer tokens from types (Tucker 2007, pp. 448–450). Salmon (1984, pp. 220–221) likened the strict explanations of tokens by tokens (his example was that of testimonies to murder by

the event of the murder) to explanations of types by types (his example was that of the results of chemical experiments by the Avogadro number), though they are entirely different forms of explanation (cf. van Fraassen 1980, p. 123).

Hempel's mismatch of tokens with types was simpler and easier to understand and therefore has been receiving more attention, not all of which was positive. Even though the critics of his covering law model did not frame their criticism in terms of types and tokens, covering laws are about types. Explanations in the historical sciences are of tokens or their representations. This gap between types and tokens is difficult to bridge: how can laws or theories made of types explain complex token events in the past? The same problem exists also in the experimental sciences, as they can rarely explain and predict events outside of the laboratory, unless they occur in relatively simple and isolated systems, such as the solar system in classical mechanics.

Covering laws or theories may fit perfectly one and only one historical case by transmuting tokens into types through dropping their specific space-time coordinates. Goldstein (1996, pp. 36–44) showed that examples for putative covering laws provided by its advocates did not provide more information than the actual historiographic explanations of representations of events they sought to cover because they merely replaced proper names or concrete terms with generalized abstract terms. Such covering laws are necessarily underdetermined because they have the confirming scope of a single historical case, even before we enter the issue of how historians can know about that single historical event that would necessitate a discussion of the relation between its representation and the evidence.

To increase its scope and standardize the historiographic representations of different events, the theory or law can become vague, so it can be interpreted in different and even inconsistent ways and participate in different historiographic explanations. A theory that is vague or complex and inconsistent may fragment into ad hoc theories each of which again fits one and only one explanation of a representation of a historical event. Such theories fail to satisfy formal semantic requirements for being scientific because they apply only to a single event, or because they are too vague, or because they are internally inconsistent.

Some philosophers and historians suggested that the elusive laws that can be interpreted and implemented consistently to explain a broad scope of historical events, the laws of history, should be imported from the social sciences, or be applied forms of "the laws of human nature." But the actual interaction between historiography and the social sciences is limited. The social sciences may lend historiography some of their theories as applied information theories, about the transmission of information over time rather than the evolution of society. For example, given the social science theory about the causal relation between centralized taxing states and the invention of writing, the discovery of writings from an as yet unknown culture would be interpreted as carrying the information that that culture must have had a centralized taxing bureaucracy during or prior to the production of these writings. Social science theories may also affect the evaluation of the prior probabilities of historiographic hypotheses. For example, the prior probability of the hypothesis that

a particular society had a system of writing at a particular period is low if we know that it did not have a central bureaucracy at that time on the basis of the social science theory that connects writing with taxing bureaucracy. This low prior probability has, for example, some interesting ramification on the prior probabilities of hypotheses about when parts of the Old Testaments, like the Pentateuch and the books of Joshua and Judges, could have been written.

The old dream of enlightenment philosophers like Hume was to turn historiography into an applied science of universal human nature, partly resembling what would later be called psychology. Since human history was mostly made by people, motivational explanations could be the key to understanding history. Several objections have traditionally been raised against this enlightenment vision of discovering the laws of history through understanding universal human nature: the motives that make people act have been evolving in history and are mostly not universal (fear of snakes and love of sex seem to be the only anthropological universals); historians do not study psychology, and psychologists do not possess laws that govern the behavior of actual human beings, except at their pathological extremes, or they would be able to predict it; and the motives of each individual are uniquely complex and cannot be captured by any general law.

Be that as it may, all sides of this debate shared the assumption that historical events, composed of human actions, are given, but that their causal and other explanations are not. This assumption, however, is false; there is nothing given about history. Historians start with the evidence; most of their work takes place in the archives and is concerned with the evidence. Whatever knowledge of human nature historians may have, introspectively from being human themselves, or from studying historical sources, or from their life experience, or from reading psychology books if they do, goes into the evaluation of the prior probabilities of motivational hypotheses. But in the end, motivational hypotheses rise or fall on how well they increase the likelihoods of the evidence. People do things for weird reasons to which historians would have assigned very low prior probabilities; for centuries the Romans enjoyed watching people being killed and killing each other in bizarre ways in the Coliseum. The Romans considered Germans and Jews physically and culturally inferior to them because, unlike the Romans, they did not practice infanticide. The Romans considered physical deformity to be funny. All these weird tastes and motives originate in a culture that, in the larger scheme of things, is extremely close to the Western culture of most historians Rome, since Roman culture is a direct ancestor to the Western civilization. But this does not matter because the Roman bloodthirsty concept of entertainment sport, the taste for killing some of their babies, and the bizarre sense of humor are the best explanation of many Roman documentary sources that comprise the evidence for the Roman culture. Whether or not this fits a particular psychological theory has only a minor effect on the posterior probabilities of historiographic hypotheses. That people anywhere would find physical deformity funny may have a low prior probability. But if this hypothesis is false, there has to be a better explanation of all the documentary evidence to the contrary, and there is none.

16.6 Conclusion

The historical and theoretical sciences are, respectively, sciences of tokens and types. Historical sciences are concerned with inferring token common causes or origins: phylogeny and evolutionary biology infer the origins of species from information-preserving similarities between species, DNAs, and fossils; comparative historical linguistics infers the origins of languages from information-preserving aspects of existing languages and theories about the mutation and preservation of languages over time; archaeology infers the common causes of present material remains; and cosmology infers the origins of the universe. These are the historical sciences, sciences that attempt to infer rigorously representations of past events, processes, and their causal relations from their information-preserving effects. The theoretical sciences are not interested in any particular *token* event, but in *types* of events and in regularities between types. The correlations between distinctions realms of nature and academic disciplines are therefore epistemically and methodologically arbitrary and obsolete. Since, from an epistemic and methodological perspective, historiography has more in common with geology, and the social sciences with agronomy, than with each other, one implication of this chapter is the elimination of a special place for human beings, their societies, and histories in epistemology. Following the Galilean and Darwinian revolutions, it is time to release ourselves from yet another narcissistic belief in our uniqueness, that is, from our alleged epistemic uniqueness.

References

Arntzenius, F. (1992). *The common cause principle* (In: PSA. Proceedings of the biennial meeting of the philosophy of science association, Vol. 2, pp. 227–337). Chicago: University of Chicago Press.

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science, 69*(3), 447–451.

Cleland, C. E. (2011). Prediction and explanation in historical natural science. *British Journal for the Philosophy of Science, 62*(3), 551–582.

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Friedman, R. (1987). Route analysis of credibility and hearsay. *Yale Law Journal, 96*(4), 667–742.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(3), 342–353.

Goldstein, L. (1996). *The what and the why of history: Philosophical essays*. Leiden: Brill.

Hart, H. L. A., & Honoré, T. (1985). *Causation in the law*. Oxford: Oxford University Press.

Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: The Free Press.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (pp. 410–505). Minneapolis: University of Minnesota Press.

Kosso, P. (2001). *Knowing the past: Philosophical issues of history and archeology*. Amherst: Humanity Books.

Maas, P. (1958). *Textual criticism* (B. Flower, Trans.). Oxford: Oxford University Press.

McCullagh, C. B. (2009). Colligation. In A. Tucker (Ed.), *A companion to the philosophy of history and historiography* (pp. 152–161). Malden: Wiley-Blackwell.

Montenegro, A., Avis, C., & Weaver, A. (2008). Modeling the prehistoric arrival of the sweet potato in Polynesia. *Journal of Archeological Science, 35*(2), 355–367.
Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.
Salmon, W. (1971). *Statistical explanation and statistical relevance*. Pittsburgh: University of Pittsburgh Press.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science, 61*(2), 297–312.
Schatzki, T. (2006). On studying the past scientifically. *Inquiry, 49*(4), 380–399.
Sober, E. (1988). *Reconstructing the past: Parsimony, evolution, and inference*. Cambridge, MA: MIT Press.
Tucker, A. (1993). A theory of historiography as a pre-science. *Studies in History and Philosophy of Science, 24*(4), 633–667.
Tucker, A. (2007). The inference of common cause naturalized. In J. Williamson & F. Russo (Eds.), *Causality and probability in the sciences* (pp. 439–466). London: College Press.
Tucker, A. (2011). Historical science, over- and underdetermined: A study of Darwin's inference of origins. *British Journal for the Philosophy of Science, 62*(4), 805–829.
Van Fraassen, B. (1980). *The scientific image*. Oxford: Oxford University Press.
Velasco, J. D., & Sober, E. (2010). Testing for Treeness: Lateral gene transfer, phylogenic inference, and model selection. *Biology and Philosophy, 25*(4), 675–687.
Wetzel, L. (2009). *Types and tokens: On abstract objects*. Cambridge, MA: MIT Press.