

AGAINST MECHANISMIC REALISM

A STUDY ON MECHANISTIC EXPLANATION

BY

MÜNİR CEM KAYALIGİL

Submitted to the Central European University

Department of Philosophy

in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Supervisor: Ferenc Huoranszki

Budapest, Hungary

July 2016

ABSTRACT

This study attempts to counter a specific brand of scientific realism: the not-much-explicated metaphysical stance of the proponents of mechanistic philosophy of science (the mechanistic philosophers, better known as “the new mechanistic philosophers”). These philosophers aim to put philosophical flesh on scientific mechanism-talk that is widely observed in life sciences and some other high-level (non-fundamental) domains of scientific inquiry. They advance a particular understanding of scientific explanation based on mechanism-talk, asserting that the explanation of high-level phenomena by describing the mechanisms underlying them is a non-lawful and non-reductive practice. On their line of argumentation, such practice sets high-level sciences apart from fundamental physics, and it accordingly demands a philosophical treatment that is different than the canonical approaches to scientific explanation.

Most of the charges brought against mechanistic philosophy of science are concerned with its proponents' overestimation of mechanistic explanation. In the final analysis, I intend to be siding with these philosophers challenging the mechanistic philosophers. However the line of attack that I pursue here is different from theirs (the former) for not focusing on the supposed non-lawful and non-reductive character of mechanistic explanation. Instead I am concerned with the metaphysical foundations of mechanistic explanation, and the implications of the mechanistic philosophers' assessment of mechanistic explanation. In particular I argue against the ontological stance with regards to mechanisms construed as causally organized token occurrences which are found relevant to the investigated phenomenon. I claim that, as opposed to being the subjects of models, mechanisms *themselves* are the models of real systems. This position, which I label as “mechanistic non-realism”, can be shown to be compatible with scientific antirealism. Nevertheless I purposefully avoid this, since I am a scientific realist. Hence the challenge for me is accounting for what is out there that mechanistic explanations are about. I suggest that dispositional realism is a viable position to adopt in this context. On the metaphysical approach I defend, mechanisms represent the repeatable patterns in the causal structure, underlain by the causal properties conferring dispositions on their bearers to behave in certain ways. I develop a meta-ontological argument to show that causal properties enjoy an ontological priority over mechanisms.

This dissertation contains no materials accepted for any other degrees in any other institutions. It contains no materials previously written and/or published by another person, except where appropriate acknowledgment is made in the form of bibliographical reference.

M. Cem Kayalıgil

ACKNOWLEDGMENTS

Writing this dissertation has not been easy at all. The research process itself has been emotionally challenging at various levels, but this is not much surprising given the investments involved in and the life plan I have woven around this study. At all stages, I was lucky to have my sweetheart Pinar next to me. She did not only motivate me in every respect, but also turned this process into something enjoyable, into another step in the path we are sharing. I first thank her, of course, for her dear presence as an everlasting, passionate embrace.

For the last couple of years that overlapped with the greater part of my studies, the country I am a contingent citizen of and morally linked with has been passing through a traumatizing transformation. This, to put it very briefly, made this very period difficult (and at times, even painful) to live through. I know that I wouldn't have been able to maintain my stance in life and hold on to my ideals in a setting like this, if I hadn't had the support of my family. My mother Ayten, my father Sinan, and my grandmother Akile backed me in the best way that parents can. My parents-in-law Alev and İhsan Karababa joined them in fully trusting that I would prevail, no matter what. I thank them all.

I am grateful to my supervisor Ferenc Huoranszki for his genial support and help throughout my studies. I am indebted to Maria Kronfeldner for courteously reading the drafts, and for motivating me to give the final shape to my argumentation (possible mistakes and gaps are all due to myself, of course). I am also indebted to Arnon Levy for the invaluable insights he provided on the defended dissertation text.

Thanks to the research grant awarded by the CEU Budapest Foundation, I found the chance to study with Michael Baumgartner at the University of Osnabrück (Germany) between November 2012 - January 2013. I am indebted to him for giving me the friendly-professional push I needed at that point for formulating my thesis question. It took an embarrassingly long time for me to be able to do that, but probably even he can't imagine how critical his contribution has been.

I want to thank Beate Krickel for her constructive criticism of a paper I presented that was based on an earlier draft of Chapter 2. This had the effect of clarifying the overall line of thought.

My great friend Cem Kamözüt also read parts of the dissertation, and showed sincere professional interest in my study. Having mentioned him, I'd like to thank the other members of the "crew" (Hasan, Erdem, Cem Türkeş) for always creating the most delightful moments in Ankara. Thanks also to Johannes, Zsuzsanna ("Suzi"), Işık, Kamuran, Deniz, Damir, Hywel, Sinziana and the members of the Philm Club for entering my life in the Budapest chapter of my life; to our flatmates in Hegelst. 24 for sharing that serene year we had in Weimar with us; to Klara for making my days in Osnabrück less depressing; and to Cem Taşan for providing havens for me at two critical points during this period. Last but not the least, thanks to Selin for the matchless friendship that I know not many people are fortunate enough to experience.

TABLE OF CONTENTS

INTRODUCTION	1
CHAPTER 1	
THE MECHANISMIC DISCOURSE AND ITS FIVE TENETS	6
1.1 PRECURSORS OF THE MECHANISMIC DISCOURSE	9
1.2 MDC's "THINKING ABOUT MECHANISMS"	18
1.3 TENETS OF THE MECHANISMIC DISCOURSE	24
CHAPTER 2	
DISAMBIGUATING "MECHANISM"	35
2.1 A LOOK AT CHARACTERIZATIONS OF "MECHANISM"	35
2.2 CRAVER'S ACCOUNT OF CONSTITUTIVE MECHANISTIC EXPLANATION	42
2.3 THE (CAUSAL) MECHANISM HARBORED IN THE SYSTEM	69
CHAPTER 3	
CAUSAL MECHANISMS AS IDEAL MODELS	74
3.1 MECHANISMIC REALISM	74
3.2 THE HH MODEL OF ACTION POTENTIAL AND CRAVER'S EVALUATION	80
3.3 APPRECIATING THE CAUSAL CONTENT IN THE HH MODEL	86
3.4 MACHINE-LIKENESS AND EXPLANATION AT THE TYPE LEVEL	94
3.5 ABSTRACTION, IDEALIZATION, AND MECHANISMS AS MODELS	105
3.6 NATHAN'S ACCOUNT OF CAUSATION BY CONCENTRATION	116
3.7 MECHANISMIC NON-REALISM	120
CHAPTER 4	
A FAVORABLE METAPHYSICS FOR THE MECHANISMIC NON-REALIST	124
4.1 REGULARITY CONSTRUED AS REPEATABILITY	127
4.2 AN AGNOSTIC VIEW: NATHAN'S SIMULACRUM ACCOUNT ...	133
4.3 THE NEED FOR A METAPHYSICAL WARRANT: DISPOSITIONAL REALISM	137

4.4 EXISTENTS AND REALIST COMMITMENT: META-ONTOLOGICAL CONCERNS.....	144
CONCLUSION	154
REFERENCES	157



INTRODUCTION

This is a critical study of a philosophical paradigm, namely *the mechanistic philosophy of science*, which is concerned with life sciences in particular (and in some occasions, high-level [i.e., non-fundamental, “special”] sciences in general). At the core of this paradigm lies a specific understanding of causal explanation in these sciences. According to its proponents, the *mechanistic philosophers*, life sciences are demarcated from fundamental physics by the lack of appeal to natural laws in the explanations for regularities. Instead of making generalizations in terms of law statements, it is argued, life scientists are after discovering *mechanisms*—that is, to give an informal definition, the recurring causal organizations of occurrences constituting the life phenomena of interest. Protein synthesis, gene expression, neurotransmitter release, or generation of action potential in excitable cells are examples of types of such phenomena, the explanations for which involve providing a causal narrative characterized by likening real systems to the inner workings of machines. This is done by distinguishing the parts of the system that are relevant to the phenomenon of concern. Consider, for instance, how a coffee is made in a moka pot: The artifact has a lid and a handle, which are irrelevant to the making of coffee. Omitting them, one can causally explain *this behavior of the machine* by *how* the water put in the bottom chamber is made to rise through the pipe of the basket above, wherein the ground coffee is put, so that the coffee drink is collected in the collecting chamber. It is observed that the scientific explanations given for many life phenomena appeal to a similar narrative: a step-by-step mental construction of how the phenomenon is produced globally within a system, by weaving the details of what goes on at the local level. We shall call this narrative the *mechanistic explanation* of the investigated phenomenon, and hold the *whole* that it describes as the explanans: the *mechanism for the explanandum phenomenon*. This is to say, what does the explaining is the mechanism described by mechanistic explanation. Mechanistic philosophers think that mechanisms exist in the world and are harbored within the causal structure that is the subject of empirical investigation. The present study is an attempt to counter this very brand of scientific realism.

The notion of “mechanism” is observed to enjoy a great popularity in the scientific discourse, and we can take this fact to be the natural start point for the mechanistic philosophy of science. In brief, the latter can be regarded as an effort to put philosophical flesh on scientific mechanism-talk. By mechanistic philosophers, I am referring to people who have advanced the “mechanistic

discourse” from the year 2000 onwards. The specifics of this discourse is the topic of Chapter 1, where I spell out how it was shaped with the publication of “Thinking About Mechanisms” by Machamer, Craver and Darden (2000). Considering the period before 2000, Kauffman (1971), Wimsatt (1976) and Brandon (1984) have been cited¹ as the precursors to the mechanistic philosophers for having observed the common use of “mechanism”. It has further become a convention to refer to Bechtel and Richardson's *Discovering Complexity* (1993) and Stuart Glennan's first paper on causation (1996) as the forerunners of a new series of philosophical analyses.² Instead of assessing how the philosophical perspective on mechanisms has changed since 1970's (which I leave to the following chapter), I would like to give space to two quotes from Francis Crick, as vivid examples for a scientist's reflections on mechanism-talk in biology. I believe that despite the lack of philosophical sophistication in his evaluation, Crick portrays the ground of discussion mechanistic philosophy is about, in a way that nicely fits the purposes of this introductory chapter.

Here is what he says about the split between physics and biology:

There is really nothing in biology that corresponds to special and general relativity, or quantum electrodynamics, or even such simple conservation laws as those of Newtonian mechanics: the conservation of energy, of momentum, and of angular momentum. Biology has its “laws,” such as those of Mendelian genetics, but they are often only rather broad generalizations, with significant exceptions to them. The laws of physics, it is believed, are the same everywhere in the universe. This is unlikely to be true of biology. We have no idea how similar extraterrestrial biology (if it exists) is to our own. . . .

What is found in biology is mechanisms, mechanisms built with chemical components and that are often modified by other, later, mechanisms added to the earlier ones. (Crick 1988, 138)

I shall make a slight detour now to underline two interrelated points made by Crick, with the intention to introduce the backbone of the discussions to follow. The first point is his claim that generalizations in biology—unlike those in physics—are liable to exceptions. He sees the absence of universally valid laws in biology as an implication of this. The second point is the centrality Crick attributes to mechanisms in life phenomena. These mechanisms, he says, change over time and affect each other (in fact, as implied in the quotations above and below, *evolve* due to natural selection). Examples of such mechanisms that Crick himself mentions include the protein synthesis mechanism and mechanisms controlling it (1988, 110), and a mechanism in the brain correcting the

¹ See, e.g., Machamer et. al. 2000, 3; Nicholson 2012, 156-157.

² See, e.g., Darden and Craver 2002; Tabery 2004; Bechtel and Abrahamsen 2005; Craver 2007 (p.5 n. 6); Illari and Williamson 2012.

memories that is suggested to cause dreams (pp. 161-162). He claims that sex may also be a mechanism, a “trick” that is possibly “used to make for smoother and more rapid evolution.” (p. 31) Above all, according to him, there is the principal mechanism of natural selection that not only governs all of these incidental mechanisms but also makes the scientific discipline unique:

[I]t is this basic mechanism that makes biology different from all other sciences. . . . Most surprising, however, are the results of such a process, acting over billions of generations. It is the general character of the resulting organisms that is unexpected. Natural selection almost always builds on what went before, so that a basically simple process becomes encumbered with many subsidiary gadgets. . . . It is the resulting complexity that makes biological organisms so hard to unscramble. Biology is thus very different from physics. (Crick 1988, 5 [original emphasis])

What Crick says amounts to an argument along the following line: *Due to evolution any life phenomenon is contingent and transient. Biologists are well-aware of this, so their scientific activities are aimed at understanding the specific and incidental results of a continuing process. These results are mechanisms: they are both products and constituents of evolution.* Now, I do not want to create the impression that this argumentation is endorsed by the mechanistic philosophers—giving evolutionary processes such a central role in the philosophical study of mechanisms is an uncommon (if not exceptional) practice in their writings, I could say. But they certainly maintain the difference of the subject matters of biology and physics on the ground that in the former, it is quite less expected to come across universal and ahistorical regularities. It is found as a consequence of this that lawful generalizations are less employed in life sciences. (This is open to debate, however. See, e.g., Waters 1998, or Mitchell 2000 for different counterarguments.) Moreover, mechanistic philosophers argue (e.g., Craver 2006; Craver and Kaiser 2013), even when such generalizations figure in biological models and theories, they are not explanatory. For instance, in ATP synthesis the ratio of oxygen molecules consumed to ATP molecules produced is 1:3. This is known as a reliable generalization that stands for a regularity; but as Bechtel and Abrahamsen note, the ratio per se cannot explain any regularity. Instead, they argue, life scientists take the generalization itself to be something that requires explanation: Asking *why* it holds amounts to asking *how* it holds, the answer to which involves (in most cases) describing the respective *mechanism*—that is, providing a mechanistic explanation (2005, 422). Mechanistic philosophers interpret this as the reason why mechanism-talk is so prevalent in life sciences.

It is a true observation that life scientists do talk about mechanisms, and it *might* turn out to be true that life scientists talk *more* about mechanisms than they do about laws. Given these, how plausible is it to presume that there are mechanisms in the world, independently of scientific

theories, and epistemic and practical concerns? Generally speaking, mechanistic philosophers are content with making this presumption, with the basic scientific realist motivation to include such scientific units in our ontology when scientists successfully work on and theorize about them. In the context of ontological commitments, the case with “mechanism” can thus be likened to the cases with some other units such as “species” or “gene”, in the following sense: Arguing from the prevalence of references in scientific contexts (to such units) to their existence is, to say the least, a *naive* realist move. Therefore, just as anyone needs additional *metaphysical* arguments to make the commitment to “species” or “gene” plausible, mechanistic philosophers ought to eke out their realist stances concerning “mechanism”. Yet as a review of the literature should expose, not much effort has been spent on this (but see Illari and Williamson 2011).

Truth be told, I believe that this brand of scientific realism—which I label as “mechanistic realism” and which finds its most clear exposition in Craver's writings³—is an untenable position, and the arguments to support the ontological commitment to mechanisms are bound to fail. As a scientific realist I hold that mechanisms are *about* the objectively existing repeatable patterns in the causal structure. This aboutness, though, should not be understood in terms of a *similarity* between the configuration of mechanism components described in a mechanistic explanation and the real causal organizations of token occurrences. It should rather be cashed out in terms of models representing real systems by *denoting* that there are recurring concrete patterns of token occurrences (necessary for the instantiation of the token explanandum phenomenon). Put differently, unlike a mechanistic philosopher who would take the configurations described in mechanistic explanations to be *models* (or descriptions) *of* mechanisms, I regard the models themselves as mechanisms. These *do* model the recurring patterns in the causal structure, but those patterns must be quite different than the configuration of components described in models.

I elaborate and defend this position in the last two chapters of my dissertation, after some stage setting which can be summarized as follows. In the first chapter I provide an almost-historical overview of the philosophical works on “mechanistic explanation”, “mechanism” and related notions. Different philosophers are motivated to solve different philosophical issues by the use of the notions in quotes. It is due to Machamer et. al. that an overall “gestalt shift” (Craver and Kaiser 2013) in philosophy of life sciences has been proposed. I outline what is on offer by the mechanistic philosophy of science in this chapter, formulating the proposal in terms of five tenets that typify the mechanistic discourse. In Chapter 2, I delve into the metaphysics of mechanistic

³ This is why, after a certain point in the dissertation, I focus almost solely on Craver's position.

explanation. Here I attempt to clarify the metaphysical picture that a mechanistic philosopher (Craver in particular, due to his 2007) has in mind, with regards to wholes, behaviors and mechanisms. Importance is given to disambiguating the two senses of “mechanism”, that is, distinguishing *causal* mechanisms from *machine* mechanisms.

After these expository chapters (where I turn down my critical voice), I start Chapter 3 with the discussion of a mathematical model of a cellular phenomenon—namely, the Hodgkin-Huxley model of the action potential in neurons. The discussion is intended to show that models which are accompanied by scientific mechanism-talk can idealize and abstract away from real systems in ways that are not appreciated by the mechanistic philosophers. From here I move on with elaborating the modeling practices by which models *deviate* from truth. I regard abstraction in mechanistic explanations as one practice that make the mechanisms different from token causal chains they are about. At first this might seem as a trivial claim, but I think type vs. token distinction is quite applicable in this context with some interesting metaphysical implications that are overlooked by the mechanistic philosophers. Furthermore, I indicate that mechanistic explanations *idealize* real systems by constructing the causal narrative in terms of the local mechanical interactions between token objects. I draw on Nathan 2012 (and also Love and Nathan *forthcoming*) to show that why the actual causal interactions in cellular systems (or real systems with a certain character) are irreducible to these interactions. On this basis I argue that mechanistic explanations generate *fictional models* of real systems. This is how I justify *mechanistic non-realism*, the position I think a scientific realist must adopt with respect to mechanisms.

The last chapter is my attempt for a positive, constructive realist argument. Here I promote *dispositional realism* to replace mechanistic realism, bringing the work of Chakravartty (2007) into the context of the mechanistic explanation of actual regularities. I assert that it is not primarily mechanisms that underlie the regularities of scientific concern, but the coherence of the genuinely existing (occurrent) dispositions of objects—that is, the causal properties due to which objects behave the same in the obtaining of the same causal setting. Finally I develop a meta-ontological argument to show that causal properties enjoy an ontological priority over mechanisms.

I do not say much on *how mechanisms explain* these regularities. That is an issue that concerns the epistemology of explanation and how models represent the world. But I do hold that scientific mechanism-talk is worth to be subjected to metaphysical analysis, even if it may not turn out to be as central (for the scientific community) as mechanistic philosophers take it to be.

CHAPTER 1

THE MECHANISMIC DISCOURSE AND ITS FIVE TENETS

It is common knowledge that the concept of “mechanism” has an extensive use in today's scientific discourse. Examples to the customary uses of the concept in medical-biological sciences include *regulatory mechanism* (for the synthesis of, for instance, a certain biochemical compound), *control* and *feedback mechanisms* (e.g., to stabilize the organism's blood pressure or body temperature), *enzymatic mechanism*, *defense mechanism* (against infection) and *mechanism of action* (e.g., of a drug). In many other scientific disciplines, too, one may come across terms having the “_____ mechanism” form: *reaction mechanism* (in chemistry), *formation mechanism* (in material science), or *social mechanism* (in human and social sciences) to name a few. More typically, though, the concept is used in unary form, having a somewhat less technicality (assuming that terms such as “regulatory” or “enzymatic” are technical terms), as can be seen in the quotation below. For present purposes, what is being stated in the quotation (the scientific significance) can be ignored; what merits attention is the use of “mechanism”—that is, the unary form and the simplicity of the concept, compared to all other terms belonging to the jargon of biochemistry.

The purpose of this article is to suggest that in view of the difficulties confronting the orthodox chemical conception of coupling in oxidative and photosynthetic phosphorylation, one might now profitably consider the basic requirements and potentialities of a type of mechanism that is based directly on the group translocation conception. This type of mechanism differs fundamentally from the orthodox one in that it depends absolutely on a supramolecular organization of the enzyme systems concerned. Such supramolecularly organized systems can exhibit what I have called chemi-osmotic coupling. (Mitchell 1961, 145)

This quotation is exemplary in two respects: First, it displays the use of the concept that one encounters a lot in scientific texts. We must note that a causal explanation is implied with this use. In this particular case, for example, the scientist (i.e., Peter Mitchell) is suggesting an alternative explanation for a phenomenon (“coupling in . . . phosphorylation”¹) which “differs fundamentally from the orthodox one.” Arguably it is more frequent to encounter scientists stating that a mechanism of/for a particular phenomenon is debated, unclear, or not known or understood. It is evident in those cases, too, that what is being admitted is that a satisfactory and/or generally

¹ More properly put, the phenomenon of concern is how the phosphorylation step in ATP synthesis is related with the transmembrane electron-proton transfer.

accepted causal explanation of that phenomenon has not been provided yet. In short, in some sense, “mechanism” and causal explanation are associated.

Second, the quotation also exhibits the *centrality* of the concept for the scientific research. Suggesting a novel mechanism is given as “the purpose” of the article. We may consider the possibility that the scientist might be using “mechanism” simply as a synonym for “explanation”. Then it could seem almost trivial that the suggestion of it is given as the purpose of the article. But even so, the wording is still remarkable, that instead of “explanation” the scientist chooses to use “mechanism”. Besides, to repeat the main point, we observe that this is quite a common use in the scientific discourse, that many scientists do prefer the use of “mechanism” to “explanation”.

These observations are voiced clearly by Machamer, Darden and Craver (MDC), in the very first paragraph of their seminal paper on mechanisms (“Thinking About Mechanisms”, 2000):

In many fields of science what is taken to be a satisfactory explanation requires providing a description of a mechanism. So it is not surprising that much of the practice of science can be understood in terms of the discovery and description of mechanisms. Our goal is to sketch a mechanistic approach for analyzing neurobiology and molecular biology that is grounded in the details of scientific practice, an approach that may well apply to other scientific fields. (MDC 2000, 1-2)

Surely MDC is not the first philosophical work on mechanisms (in science). Noteworthy papers predating MDC's “Thinking About Mechanisms” are Kauffman 1971, Wimsatt 1976, Brandon 1984 and Glennan 1996. Bechtel and Richardson 1993, on the other hand, is a book-length study on how scientists explain complex systems mechanistically. However, MDC's perspective is essentially different from the approaches found in these earlier works.

Shortly put, MDC 2000 is the first work where we see both the deliberate motivation for precise characterization of mechanisms (i.e., how they are understood by life scientists) and the endeavor to draw ontological-metaphysical implications from the relevant scientific practice (i.e., the practice in life sciences). The practice of concern here is a particular way of investigating and explaining recurring/regular phenomena of scientific interest. Examples to such phenomena are how RNA chains are formed according to the nucleotide sequence on the DNA (*RNA transcription*), how an electrical signal is generated in neurons (or other excitable cells) to enable cell-to-cell communication (*generation of the action potential*), or how neurotransmitters are released into the synaptic cleft. It is observed (by MDC, for instance, as seen in the quotation above) that a *mechanism-talk* frequently accompanies the explanations given for such phenomena in the scientific discourse. This is to say, regarding the phenomena just mentioned for instance, we see scientists

talking and writing about the “RNA transcription mechanism”, “the action potential mechanism”, or “the neurotransmitter release mechanism”. It is to MDC's credit that for the last 15 years, there has been a wealth of work by which a philosophical flesh is put on this mechanism-talk.

Put very briefly, the description of a mechanism M for a phenomenon P implies that there is a repeated pattern of causally-related occurrences. For instance, one can take the RNA transcription mechanism to be formed by the following three occurrences: (i-) A transcription factor binds to the nucleotide sequence on the DNA (this sequence is called the “operator”), which mediates (ii-) the binding of the RNA polymerase enzyme to another nucleotide sequence, by which the latter sequence (i.e., the gene that is to be expressed) is used as a template for (iii-) the synthesis of the RNA chain. This is a quite coarse description of the mechanism responsible for the transcription of the RNA. It could be illustrated as in below (the ellipses illustrate the occurrences of concern, and the arrows represent the causal relations between them—that is, which occurrence causes what):

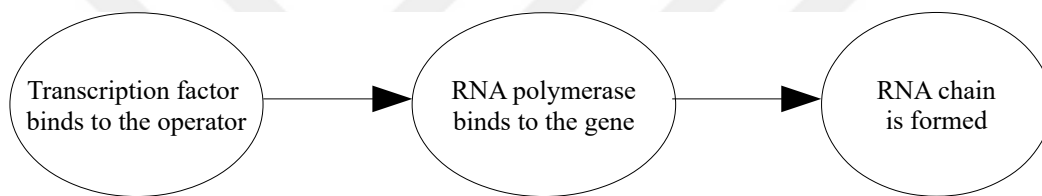


Figure 1-1. RNA transcription mechanism

The main premise here is that, scientific mechanism-talk more or less amounts to an understanding gained by such mental decomposition of the phenomena into the important causal steps of their production.² When a satisfying description of how the steps are related with each other (i.e., how the occurrences are causally organized) is given, we have a *mechanistic explanation* of the investigated phenomenon. This conceptualization will serve as the basis for the discussions in the following chapters. My aim will be to raise objections to a particular philosophical attitude with regards to scientific mechanism-talk and mechanistic explanation. The attitude is displayed in the writings of the “mechanismic philosophers (of science)” who are engaged in the “mechanismic discourse”.³ I hold that the discourse has originated in MDC 2000. A review of the earlier works should let us appreciate the originality of this paper. It will also serve as a sketch of the discussions surrounding the philosophy of mechanisms.

² Nevermind at this point what attaches importance to a causal step.

³ The term “mechanismic philosopher” is due to Nicholson 2012. I will use “mechanismic discourse” almost interchangeably with “mechanismic philosophy of science”. I elaborate on these preferences in the last section of this chapter.

1.1 PRECURSORS OF THE MECHANISMIC DISCOURSE

I start with Wimsatt's and Brandon's papers which are actually on *reduction in science*—a topic which was hotly debated in that period. Both make the point that mechanisms play a central role in biological explanation (Wimsatt 1976, 671; Brandon 1984, 345). Yet it must be underlined that this observation is stated for the goal of elaboration of reduction, that is, *not* for the elaboration of mechanisms. Wimsatt does not discuss what mechanisms are at all. He develops a functional account of reduction on which reductive relations are to be understood between levels of organization (instead of being between theories which was the standard view during 1970s). In this understanding, levels are *real* features of the world; they are characterized both by how entities are composed of lower-level entities and by where regularity and predictability are found maximally (1976, 679-680). A reductive explanation is the explanation of how a phenomenon is produced by the causal interactions at a lower level. Now, given that Wimsatt says that, for biologists what is seen as reduction is *discovering mechanisms* (and not explaining theories by reducing them to some other theories) (1976, 671), it can be argued that he equates mechanisms with the *causal interactions at lower levels* found to be producing the phenomena at higher levels. However, it must be stressed, this is an *implicit* premise in his paper; what a mechanism is is not questioned but implied to be these lower level interactions.

Brandon's paper has the remarkable feature of raising the question of what a mechanism is. The author immediately admits the difficulty of giving a precise definition. For him this difficulty is due to the variety in things which are seen as mechanisms:

But what is a mechanism? Here I cannot be precise. Sometimes old-fashion spring-wound clocks and watches are called mechanical devices. Clearly, I cannot use 'mechanism' in such a narrow sense. Mechanisms may consist of springs and gears, they may consist of computer chips and electrical pulses, they may consist of small peripheral populations and geographic isolating barriers. I cannot delimit all possible mechanisms because it is the business of science to discover the mechanisms of nature. At best I could list the sorts of mechanisms science, or more specifically, biology has discovered. (1984, 346)

In Nicholson's phrasing, here Brandon recognizes the “semantic ambivalence inherent in the biological usage of 'mechanism'” (2012, 157). Nicholson has a specific critical concern that I share, which deserves a separate discussion that can wait until the next chapter. For now, though, it should be enough to note that there are two senses of “mechanism” in today's science: One is due to likening relatively complex systems (the natural phenomena) to artifacts made/preset to perform specific functions—to *machines*, so to say (be they mechanical or electronic, as observed by

Brandon, too). In this sense the mechanism is definitely objectified and spatially localized; the boundaries of the mechanism can clearly let one decide what is internal and what is external to it. In the second sense, “mechanism” amounts to the set of causally-related occurrences producing the (scientifically) investigated phenomenon. It is in this latter sense that a mechanism is held to play its explanatory role. The difference between these senses is easier to see when artifacts are concerned. To use Brandon's example, the springs and gears of a spring-wound clock and their orchestrated functioning constitute the mechanism of the clock: this is “mechanism” in the second sense, and it is this mechanism one must appeal to in order to explain *how* the clock shows the time (*viz.*, the energy stored in the wound spring makes Gear A rotate, thus Gear A transmits the torque to Gear B, etc.). On the other hand, it is also possible to see the clock itself as a mechanism (a *machine* showing the time—“mechanism” in the first sense). Obviously this mechanism, that is, the single object is not an occurrence; it does not produce or (causally) explain any phenomenon by itself (mechanism in the first sense). I shall use the labels “machine mechanism” and “causal mechanism” for the first and the second senses, respectively. These labels are due to Nicholson, and the referents of the machine and causal mechanism *senses* of “mechanism” are to be discussed throughout the dissertation. (I must acknowledge that to some extent, my understanding of machine and causal mechanisms differs from Nicholson's. As it is early for elaborating this, I leave the issue for the time being.)

As noted earlier, Brandon's paper is also on reduction in science. Unlike Wimsatt, Brandon uses the notion of “mechanistic explanation” (nowhere in Wimsatt's paper we see this). Nevertheless, since Wimsatt, too, associates the discovery of mechanisms with accounting for how a phenomenon is produced (1976, 686 and 689), this difference does not amount to a significant disagreement in their perspectives. A more noteworthy dissimilarity is whether the authors see the discovery of mechanisms as a *reductive* task or not. As indicated above, for Wimsatt, explaining through mechanisms is nothing but reduction (where the phenomenon is explained by lower-level causal interactions). Yet Brandon explicitly opposes this; in fact this constitutes the main point of his paper. He acknowledges that sometimes wholes are explained by their parts: “For instance, the behavior of a watch may be explained in terms of the behavior of its springs and gears. This may support gear-spring reductionism, but it does not imply that everything is explainable in quantum mechanical terms.” (1985, 347) However, he observes, “[s]ome mechanistic explanations are given in terms of entities more inclusive than the explanandum entities”. For this he gives the example of gene frequency change in natural selection, in whose explanation whole organisms, structures of populations and properties of groups of organisms could be relevant (1985, 347-348). Therefore in

these cases discovering the mechanisms means finding the causal relations taking place at a higher level, a task whose fulfillment works against reductive tendencies.

As a matter of fact, *Wimsatt* does not reject the case of explaining a phenomenon by higher-level causal interactions either. Nonetheless he maintains that explanatory efficiency dictates an “order of priorities”, according to which explanation in terms of higher-level interactions is much less desirable than a reductive explanation (explanation in terms of mechanisms)(1976, 689).⁴ The reader should be excused if this last point was confusing, so let me recapitulate: Wimsatt associates mechanistic explanation (a term he does not use) with reduction, while Brandon argues against this. The difference between the two lies in their understanding of mechanisms: Wimsatt has a narrower understanding, according to which mechanisms are only the lower-level causal interactions producing a phenomenon. *Their discovery* “is seen by [most scientists] as reduction, or as integrally tied to it.” (1976, 671) Brandon would definitely agree with this much, but his views diverge when it comes to explaining phenomena in terms of the causal interactions at a higher level. Both authors admit that this is a legitimate explanation. Wimsatt says in scientific practice this is less desirable than reductive (mechanistic) explanation, while Brandon says this is mechanistic explanation, too—and since it is so, mechanistic explanation is not reductive.

On the way to showing the originality of MDC 2000, I now move on with giving an overview of Bechtel and Richardson's book *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research* (1993). As displayed by the full title of the book, Bechtel and Richardson's aim is to identify *how scientists* (i.e., life scientists) *think* during their researches: what strategies they use for tackling with the problems they face with in developing the scientific knowledge. The authors take stance against the logical positivist tradition which downgrades the *context of discovery* by regarding it as a “‘merely’ psychological matter having no significant consequences for a theory of scientific rationality.” (1993, 3) Their stance involves taking discovery as inseparable from justification/confirmation, and the recognition of cognitive capacities and limitations. Specificities of human cognition affect the two inseparable modes of the research and thus determine the dynamics of scientific development.

⁴ It must be noted that he says, where possible, the most desirable explanation of a phenomenon is the one which shows how it is produced by interactions at the same level. A reductive explanation has a less priority than this, because the levels of organization are taken as the “local maxima of regularity and predictability” (1976, 689), and the lower level is examined if the scientist wants to account for the *exceptional* cases (why an alternative phenomenon occurs instead of the expected one) and/or regularities (1976, 691).

We operate on the assumptions that psychological constraints are important in understanding scientific change, confirmation, and discovery, and that the sorts of explanatory strategies employed in scientific problem-solving are analogous to the strategies employed elsewhere in human problem-solving. We are concerned primarily with how an explanatory task is seen and how the problem is represented; with how the representation of the problem can influence the character of the resulting theories. . . If these explanatory strategies were not common, our own descriptions might be charged with being no more than post hoc descriptions of cases. The fact that they are robustly represented in human cognition, however, constitutes the case for their reality. (1993, 7)

Bechtel and Richardson give weight to real cases and episodes in history of science to substantiate this last point, the point that particular heuristic strategies are employed vastly in life sciences. They introduce these strategies as “decomposition” and “localization”. Basically, decomposition amounts to the assumption that the whole—be it the nature or the units composing it—is composed of identifiable and isolable parts (1993, 39). The wholes of scientific concern are treated as *complex systems* responsible for producing certain *behaviors*.⁵ Finding which system produces that specific behavior is the first step of the scientific endeavor. The identification of this “locus of control” is followed by specifying one constituent unit of the system as the functional component *individually* responsible of the behavior. Bechtel and Richardson call this step “direct localization”, which is nothing more than “locat[ing] an underlying system within a complex system.” (p. 65)⁶ Direct localization “relocates the problem . . . to a lower level of organization” (*ibid.*); so the research never ends there, but usually reveals instead that what the functional component does is just *one operation among many others* in the system altogether producing the behavior. The ultimate explanation for the behavior requires elaborate investigation of *how* the components of the system *function together*. Now, the authors refers to *this* as “mechanistic explanation”, and they have a reason for this labelling:

By calling the explanations *mechanistic*, we are highlighting the fact that they treat the systems as producing a certain behavior *in a manner analogous to that of machines* developed through human technology. A *machine* is a composite of interrelated parts, each performing its own functions, that are combined in such a way that each contributes to producing a behavior of the system. A mechanistic explanation identifies these parts and their organization, showing how the behavior of the *machine* is a consequence of the parts and their organization. (1993, 17 [only first emphasis is in original])

⁵ Bechtel and Richardson intend to address biology and psychology in particular: “Examples of complex systems we have in mind are the physiological system in yeast that is responsible for alcoholic fermentation, and the psychological system responsible for memory of spatial locations.” (1993, 17)

⁶ Both the identification of the locus of control and direct localization are in fact regarded as applications of the same heuristic (i.e., localization). The authors state that they differ “only in the level of analysis.” (1993, 63)

A conspicuous aspect of this approach to mechanistic explanation is that the labeling is *not* due to the frequent use of “mechanism” concept in life sciences. Nowhere in their book Bechtel and Richardson attempt to analyze or describe this very concept as it is used by life scientists. They rather analyze the scientists' reasoning and problem-solving strategies, and proceed from these to the label of “mechanistic explanation”.⁷ Furthermore, note that they define “machine”, not “mechanism”, due to their judgment that natural “complex systems” are treated (by scientists) as analogous to artifacts. Let us recall here the semantic ambivalence in “mechanism” and remark again that the scientific usage of the term *only sometimes* legitimizes this interpretation (i.e., mechanisms as machines). Other times, the proper philosophical understanding of the “mechanism” notion should be based only on the existence of a set of causal relations (producing the phenomenon of scientific concern). When scientists use the concept in this sense (the causal mechanism sense), the causal relations are there but the set they constitute (the mechanism) has no non-arbitrary (spatial and temporal) boundaries to isolate them from whatever taken to be external to the phenomenon. What sets the boundaries is the explanatory context, that is, what the investigated phenomenon is; mechanisms in this sense are necessarily individuated by the phenomena they explain (see e.g., MDC 2000, Section 4; Craver 2007, Section 4.4). This is one reason why a (causal) mechanism is not machine-like, and why the analogy with machine fails. This seems to weaken Bechtel and Richardson's description of mechanistic explanation by limiting it with cases where the analogy holds.

Let me make this more precise: The problem is with their *description*, with *not* realizing and acknowledging the semantic ambivalence in “mechanism”. Otherwise, the authors' *comprehension* of mechanistic explanation does not exclude the causal mechanism sense. In fact it *cannot* and this is just as expected, for the “actual scientific practice” that their study is grounded in (1993, 17) invokes both senses. A disambiguation is quite essential for an accurate appreciation of how scientists use the concept, but as we will see, the lack of this is common to all philosophical works valuing the scientific mechanism-talk (including MDC's paper).

A very significant feature of the Bechtel-Richardson approach to mechanistic explanation is the importance attached to the *epistemic* aspect of the scientific activity. *Modeling* is the keyword here: Bechtel and Richardson are overt proponents of the tradition which regards scientific theory construction as the development of *mental apparatuses* to *approximate* objective reality. (1993,

⁷ They also indicate that their study is not intended to be exhaustive. They acknowledge that life scientists may be using other heuristics and explanatory strategies, too; but what they present in the book should be seen as an approximation that is best applicable to contexts where problems and criteria for their solutions lack clear statements. (1993, 7; see also p. 233)

232) On this perspective, giving mechanistic explanations of phenomena is equivalent to proposing models which are “partial and abstract representations of the causal mechanisms” (*ibid.*). Their use of “causal mechanisms” here is not because they also write about machine mechanisms. Instead, they are plainly addressing the causal interactions between the parts that make up the investigated system responsible of the phenomenon. In other words, they intend to use “mechanism” in the causal mechanism sense; but having not distinguished the two senses, they need to add “causal” in order to emphasize the appropriate meaning. They could be well off if they did not use “mechanism” (they could use “causal interactions between the parts of the system” instead, for instance), but then they would probably lack the straightforward justification for the labeling “mechanistic explanation”.

In light of these considerations and the descriptions of machine and mechanistic explanation quoted above, the stance in *Discovering Complexity* can be summed up as this: Giving a mechanistic explanation of a phenomenon is to *model* a complex system mechanistically. This system in question is a decomposable whole (where the phenomenon is localized); to model this system mechanistically means to *mentally represent it as analogous to a machine*. Objectively speaking, there may or may not be a single, spatially localized, machine-like unit producing the phenomenon⁸—however the case is, what makes the explanation of the phenomenon *mechanistic* is that the overall behavior of the system is shown to be the consequence of the causal interactions between the parts making up the system. Summarized as such, Bechtel and Richardson's stance illustrates the *epistemic conception* of mechanistic explanation, on which the explanans does not have an objective ontic status. Some of the most important contributions to the philosophical study of mechanisms have been on distinguishing this from the *ontic conception* of mechanistic explanation, on which mechanisms are, qua situated in the objective causal structure, *the* explanantia. The latter is notably championed in Craver 2006 and 2007, and appears to be given a dominance in MDC 2000. The ontic conception deserves special scrutiny due to its problematic implications (with regards to its promoting the ontological commitment to mechanisms); a lot will be said on this in the forthcoming chapters.

In this section I have been reviewing the important philosophical works which predate MDC and analyze mechanisms and mechanistic explanation in science. As indicated above, in neither of these

⁸ “Fermentation is something done by cells; the mechanism is a complex biochemical pathway. Development is something organisms undergo; the mechanism that regulates it is a complex genetic pathway.” (Bechtel and Richardson 1993, 231) These two are cases where the proposed mechanisms cannot be machine-like, because they are “pathways” that span numerous units. Arguably most mechanisms are of this sort, that is, it is dominantly the causal mechanism sense of “mechanism” that applies in scientific contexts. (Note: The authors' intention of citing fermentation and development is—naturally—different than mine. Recall that they do not discriminate the two senses of “mechanism”.)

three studies can we find a precise definition of “mechanism”. The best we have is Bechtel and Richardson's definition of “machine” and it is meant to be a *depiction* of the explanatory model of a phenomenon: According to the authors, the model arrived at by mechanistic explanation is the model of a system *resembling a machine*. This does not provide an adequate characterization of scientific mechanism-talk (not all mechanisms are machine-like [see footnote 8]), but it is important to recall that Bechtel and Richardson do not intend to capture this scientific usage of the concept anyway (unlike Wimsatt and Brandon, and as will be seen, MDC). The case is almost the same with Glennan 1996, which is the last work to review before closing this section: A description of “mechanism” is provided; it is neither completely faithful to the scientific usage of the concept nor meant to be so.

This particular paper by Glennan, like some others he published afterwards, is actually on causation. It is fair to say that Glennan is not much interested in scientific explanation or mechanistic explanation allegedly found in life sciences.⁹ Nevertheless, while not particularly concerned with the scientific practice *per se*, he argues for a distinction to be drawn between fundamental physics and other sciences: Generalizations in the form of laws are found in both, but the “mechanically explicable” ones are only in sciences other than fundamental physics. It is this insight of Glennan (explained below) that places his work next to a body of papers building up on MDC's account of mechanisms. Aside from this, Glennan's standpoint in the philosophical analysis of mechanisms is quite peculiar; it is marked by advancing discussions *other than* those on the implications of the (life) scientists' usage of the concept.

One outstanding point about Glennan's paper is the deliberate demarcation of the two senses of “mechanism”: “The first sense refers narrowly to the internal works of machines, as when one speaks of a clock mechanism. The second refers more generally to complex systems analogous to machines, as when one speaks of a human perceptual mechanism or a market mechanism.” (1996, 51-52) This is followed by the definition of the concept, with a note that it is the second sense that is intended to be captured. (To not confuse: The “second sense” for Glennan is the one I have been referring as the first sense—i.e., the machine mechanism sense). The definition is:

A mechanism underlying a behavior is a complex system which produces that behavior by the interaction of a number of parts according to direct causal laws. (1996, 52)

⁹ This is especially true regarding his 1996 paper, but the point concerns his work in general, too. “Rethinking Mechanistic Explanation” (Glennan 2002) is a prominent exception to this observation. Still, even in this later paper he is not concerned with the mechanism-talk as found in life sciences (or, as I say in the sentence to follow, with the scientific practice *per se*). Thus I see here sufficient reason to regard his work separately from the literature on mechanistic explanation.

For Glennan, showing that two events are connected by a mechanism (defined as above) accounts for their being causally related. This is an empirical task and hence the nicety of his “mechanical theory of causation”: It solves (according to him) the notorious Humean problem of finding the “secret connexion” without resorting to metaphysical reasoning (1996, 64). Glennan gives the familiar example of smoking and cancer: These are correlated, but the existence of a genuine causal link between them cannot be put forth unless “the mechanism by which tar, nicotine, etc., interact with the body to produce cancerous cells” is discovered (1996, 64). In previous pages we see Glennan examining a float valve and a voltage switch as two examples of mechanism. One of the aims of this is to elucidate “mechanism” understood as machines; and since these two systems are artifacts, it is straightforward to grasp the relevance of this definition which is tailor-made to cover these very cases. Nonetheless I find it hard to apply this definition in his mechanical theory of causation. It appears that the possible mechanism “connecting” smoking to cancer should be of the “clock mechanism” sort (i.e., first sense in Glennan, second sense in my study [the causal mechanism sense]): The explanation of how “tar, nicotine, etc. interact with the body” would probably be more like the descriptions of “internal works of machines”. More could be said on this, but I do not want to be repeating the points made before—so here is my quick verdict: It is curious to see how the demarcation of the two senses does not help Glennan himself, for again we observe the semantic ambivalence in “mechanism” blurring the discussion.

Concerning the proposed theory of causation, there are some other problems with Glennan's account, arguably overshadowing the issue of the applicability of the definition. Yet considering them would not add much to this introduction to the philosophical work on mechanisms in scientific discourse, so I do not intend to go deeper.¹⁰ Instead I want to close this overview with a glimpse of the issue of laws in non-fundamental sciences.

One key notion in Glennan's paper is being “mechanically explicable”. Giving the mechanical explanation¹¹ of a behavior involves decomposing the system (the *mechanism* in the sense defined by Glennan) into its parts and showing how the parts causally interact with each other. These interactions may be, but need not be “strictly mechanical” (1996, 51): Parts of a float valve, a mechanical mechanism, move owing to the pushes; but in a voltage switch, which is an electronic

¹⁰ Chapter 3.4 of Craver 2007 offers a good critique of Glennan's theory of causation (see especially pp. 90-93).

¹¹ Glennan does not use the term “mechanistic explanation”.

mechanism, there are no moving parts, at least at the macrophysical level. In either case, as long as systems can be decomposed into robust and real parts, their behavior can in principle be mechanically explicable.

The important (and controversial) feature of this framework is that, laws enter into this picture at two points: They govern the interactions, as stated in the definition of a mechanism (1996, 54), and they themselves might be mechanically explicable as well (1996, 61). The former point received criticism (e.g., MDC, 4; Woodward 2002) on the basis that laws usually do not figure in the mechanism-talk in life sciences (and other special, or high-level sciences), so that they cannot and should not be invoked in the characterization of mechanisms. Glennan must have been convinced by this that we see the replacement of “causal laws” with “invariant, change-relating generalizations” in his later definition of “mechanism” (2002, S344). The latter point (the one about the mechanical explication [mechanistic explanation] of laws), on the other hand, can be read as expressing one of the ideas the mechanistic philosophers generally agree upon—that mechanisms can explain why generalizations given as law statements hold.¹² On Glennan's perspective (which, for the reason just mentioned, does not differ from the mechanistic philosophers') laws are stripped off the metaphysical weight classically associated with them. They are taken as generalizations supporting counterfactuals (Glennan cites Goodman and Nagel as figures he sides with in this issue). Further, Glennan liberally counts any such generalization as a law, however “banal” it could be (1996, 55). Hence, a law as a generalization could be the subject of mechanical explanation, too, given that the behaviors to be explained by mechanisms are regular and can be expressed in the form of such generalizations.¹³

Glennan uses the voltage switch example to demonstrate that some laws are mechanically explicable. The switch is made to produce a certain input/output behavior—that is, it regulates if and how much current passes, depending on the properties of a transistor and the comparison of the voltage differences between the transistor's terminals. Glennan provides the following mathematical expressions of this behavior, which is conditioned on the input voltage V_{in} (1996, 60):

$$\begin{array}{ll} V_{in} \leq V_{off} \Rightarrow (V_{out} = V_{source}) & \text{Switch is off} \\ V_{in} \geq V_{on} \Rightarrow (V_{out} \approx 0) & \text{Switch is on} \end{array}$$

¹² See e.g., Andersen 2011, Craver and Kaiser 2013. More on this in Section 1.3 below.

¹³ Bogen (2005) argues against this and there is a debate on whether there can be one-shot mechanisms. Therefore this should not be read as a claim free of controversy. Yet it is certainly the view of Glennan and the majority, that mechanisms produce regular behaviors.

The only details that concern us are these:

(i-) V_{off} can usually be taken as zero, while V_{on} is determined by the characteristics of the transistor and V_{source} .

(ii-) The current that passes is a function of the difference between the voltages V_{source} and V_{out} ; so there is no current (the switch is off) when the values are equal.

So we have these laws which not only describe the behavior of voltage switches but also can be explained mechanically. For a system described by them is a decomposable one, where causal interactions take place between robust and real parts (resistors, conductors, etc.). Furthermore, such mechanism may itself be a part of a higher-level mechanism (say, an electric circuit); alternatively *a part of it* could be taken as a mechanism to be explained in isolation (whose parts might then be “electrons, molecular lattices, or other such entities”)(1996, 60). It is also true for some well-known laws, such as Hooke's law or the laws of classical genetics, that they can be mechanically explicable (1996, 61).

The picture is, then, of mechanisms within mechanisms and mechanical explanation within mechanical explanation. Yet there is an ultimate level at which there is no further mechanism to explain the behavior, and this is the level of reality (the wording is mine) described by the *fundamental* laws. Maxwell's equations, Schrödinger equations and Einstein field equations are examples to these, which can be seen in contrast to the mechanically explicable, non-fundamental laws mentioned above (1996, 61). Fundamental laws are not mechanically explicable, because they are not about *parts* interacting with each other, or about systems constituted by robust entities.

This last insight by Glennan, the multidirectional mechanical explication ceasing at a fundamental level, is a characteristic view recurring in the philosophical works on the scientific mechanisms-talk. We are now in the position to examine the paper which pioneered this trend in philosophy of science.

1.2 MDC's “THINKING ABOUT MECHANISMS”

So far, my attempt has been to cover the earlier efforts to place the scientific mechanisms-talk into a philosophical framework. We saw that this task is interpreted in not so similar ways. Wimsatt and Brandon acknowledge the usage of “mechanism” in life sciences; developing some specific

views on the scientific explanation characterized by the discovery of mechanisms, their primary aim is to elaborate on reduction. On the other hand, there is no such acknowledgment in Bechtel and Richardson's book *Discovering Complexity* or in Glennan's 1996 paper. When the latter is concerned, this does not come as a surprise. For the mechanism-talk Glennan avails himself of is tangential to its articulated scientific functions and worth. Nevertheless, this is hardly the case with *Discovering Complexity*: Although the observation about the ubiquity of “mechanism” is not explicit (or apparent) in Bechtel and Richardson's work, there is no question that their study *is* on *scientific* mechanism-talk. For the authors are quite aware of (and informed about) the scientific contexts where the concept of “mechanism” is invoked and figures in. The far more critical point in their study is that, for them, what makes an explanation *mechanistic* is due to the *mental image* representing (modeling) the phenomenon. The alternative to this would be to view the explanans for the phenomenon as belonging to the objective, empirical world. As indicated before, this alternative (the ontic account of mechanisms) is notably put forward in Craver 2007.

Among the many original aspects of “Thinking About Mechanisms” (MDC 2000), arguably the most crucial one advanced is the claim for the “ontic adequacy” of the scientific search for (or, scientific research on) mechanisms (Section 3). This claim attaches an importance to *activities* as being a fundamental ontological category. The coupling of these requires them to be examined in tandem, and a good starting point for this is the MDC's description of “mechanism”. Mechanisms for MDC are,

entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. (2000, 3)

Accordingly, a mechanistic explanation is given when the particular organization of entities and activities is revealed. (MDC's paper does not provide a clear picture of what mechanistic explanation is [see footnote 16 below]. In any case, this much constitutes the basic idea.)

Let us take a moment to appreciate the novelties inherent in this description. First, the intention here is to directly grasp the usage of the concept of “mechanism” per se in the scientific discourse—in biology and neuroscience, in particular. Reflecting in a later paper, Darden (the “D” of MDC) underlines the team's motivation:

[T]he MDC *characterization* of mechanisms was not presented as a *definition* to provide necessary and sufficient conditions for the term's usage in all cases. Instead it was a characterization aimed at capturing the way biologists use the term, as informed by our detailed examination of cases from molecular biology and neurobiology, and also informed by philosophical reflection on requirements for *productive* changes. (2008, 960; original emphases)

This should be set in comparison to Bechtel and Richardson's description of “machine” as something the explanatory model looks like, and to Glennan's description which does not attend to the scientific mechanisms-talk.

Second, there is no mention of “complex systems” or “interactions” in the description (as in Glennan 1996), but instead, there are “entities and activities” which are “productive of regular changes”. These are construed quite broadly: Ions, sodium channels, DNA, gene, neuron or an organ of a mammal are examples of entities that can figure in a mechanistic explanation, where the activities can include attracting, pushing, bonding, diffusion of a substance or conduction of electrical impulses. Despite appearances, “entity” and “activity” are *weighty* concepts and MDC are well-aware of this. Even though the same might be claimed for “system” and “interaction”, their use in the mechanisms literature is much more conventional. This is to say, in this particular literature, “system” and “interaction” are usually cleared of their possible philosophical (metaphysical) implications. (More to be said on this below.) However, MDC put effort to elucidate “entity” and “activity”, and they do this with a constructive intention to open the scientific mechanism-talk to a metaphysical analysis. It is not an exaggeration to say that the crux of the MDC perspective of mechanisms is encapsulated in the “productive” nature of “entities and activities”.

MDC explicitly advocate a dualist ontology in their analysis of “mechanism” and endorse “the healthy philosophical intuitions underlying both substantivalist and process ontologies.” (2000, 4) Accordingly, both entities and activities are granted fundamental ontological statuses. They ground this dualism in the scientific practice where, according to their understanding, progress and success lie in the discovery and the correct specification of real entities and the real kinds of activities that these entities engage in. As a case study to substantiate this point, they consider the example of the discovery of the mechanism of protein synthesis (Section 6). In particular, what is needed to be explained was the link between DNA and protein synthesis (how the amino acids were sequenced according to the genetic code). It took almost a decade (*circa* 1954 to 1964) for biochemists and molecular biologists to improve on the explanation that the incorporation of the amino acids into proteins took place on the RNA.¹⁴ Watson's proposal was that RNA had “holes” which were shaped by the surrounding bases. Different bases would mean differently shaped holes, and due to the shape differences these holes accommodated different amino acids. Hence the ordering of the amino acids were accounted for by the sequence of these holes, determining the order of amino acids that “fit” into them.

¹⁴ The idea that protein synthesis takes place on RNA was itself an improvement. Earlier, the synthesis was associated directly with the DNA (MDC 2000, 18-19).

Watson was wrong, though. The critical facts to be discovered would be as follows: RNA has types; there is a type on which the genetic information is *transcribed* (messenger RNA [mRNA]) and there is another type by which the transcribed information is *translated* as the amino acid sequence (transfer RNA [tRNA]). (Besides these two, there are many other types of RNA.) There are no “holes” (of the sort Watson proposed) on either of these types of RNA. Instead, there are *codons* on the mRNA which base-pairs with the *anticodons* on the tRNA. Codons are sequences of three nucleobases that encode specific amino acids and anticodons are the complementing sequence. Once a codon is presented, the tRNA molecule with the required anticodon binds to the mRNA. Opposite to the anticodon loop of the tRNA is situated the site where a specific amino acid is attached (see the figure below). Therefore the codon-anticodon base-pairing (by hydrogen bond) brings a specific amino acid and places it next to the one brought previously by another tRNA. With each amino acid being added (by peptide [covalent] bond), thus forms chains that make up the proteins.

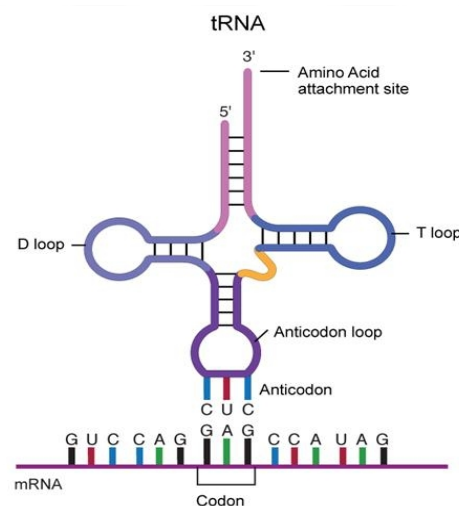


Figure 1-2. Simplified structure of the tRNA, and the codon-anticodon pairing. Taken from Microbe.net (licensed under CCA-ShareAlike 3.0 Unported License).

This case of scientific progress could be seen solely in terms of entities: Some of the proposed *entities* (like Watson's “holes”) are abandoned because it is found that they do not exist, new *entities* are introduced (different RNA types are discovered), knowledge about some *entities* grows and deepens (the role of the RNA family of macromolecules in protein synthesis is understood much better), and so forth. MDC, however, stress that an adequate account of the history of the discovery of protein synthesis mechanism should also include the discovery of new *activities* (2000, 21).

According to them, the obsolete mechanistic explanation of the protein synthesis phenomenon is wrong in its conjecture with regards to the activities, too: Just as there are no “holes”, the amino acids do not engage in the *activity* of “fitting” into them; and just as the true entities are mRNA and tRNA, or codons and anticodons, the true activity is the hydrogen bonding that these entities engage in.

MDC's general point is, the mechanisms discovered and described by scientists “do things” and the mechanisms' being “active” rests in the activities—*not* in the interactions of entities (cf., Glennan 1996), changes in the entities' properties, or manifestation of capacities (powers) of entities (2000, 5; see also p. 3). In their words, “much of what neurobiologists and molecular biologists do should be seen as an effort to understand these diverse kinds of production [in the activities] and the ways that they work.” (2000, 7) Still, entities and activities need each other; entities and their properties determine which activities can take place, and conversely the activities determine which entities with which properties can engage in productive relations (2000, 6). This being so, understanding the activities is never independent of finding out about the entities, and vice versa (as demonstrated by the example of protein synthesis mechanism, according to MDC).

From the ontological viewpoint, it should be noted that the *fundamentality* of activities and entities is meant to be *relative* to the scientific domain and/or research interests. The components of protein synthesis mechanism include macromolecules and chemical bonding. These comprise a mechanism at a certain *level*: Below and above this, there are other mechanisms—below, for example, there are mechanisms of the formation of the relevant macromolecules; and above there are mechanisms where the synthesized proteins themselves act as the entities. That mechanisms stand in such *nested hierarchy* was an insight we first saw in Glennan 1996, and becomes a central idea after MDC. Nevertheless, MDC's approach differs sharply on the issue of “bottoming out”—that is, on the issue of the fundamental level where no further mechanism is found below.¹⁵ Whereas in Glennan's analysis of “mechanical explanation” this fundamental level is always where fundamental physics is set, for MDC there is no such ultimately distinguished level.

Bottoming out is relative: Different types of entities and activities are where a given field stops when constructing mechanisms. . . . The fields of molecular biology and neurobiology, in 1999, do not typically regress to the quantum level to talk about the activities of, e.g., chemical bonding. . . . Levels below molecules and chemical bonding are not fundamental for the fields of molecular biology and molecular neurobiology. But remember, what is considered the bottom out level may change. (2000, 13-14)

¹⁵ I want to note that the phrase “bottoming out” is a quite popular term belonging to the jargon of the mechanisms literature. It should not come as a surprise that the first use of the term is found in MDC 2000. I take this as just another sign of the influence of MDC.

The difference between Glennan's and MDC's approaches should be attributed to their respective motivations. Let us recall: Glennan's primary aim is to develop a theory of causation based on the “mechanical explication” of the phenomena. In this respect the fundamental level constitutes the limit for his theory (beyond which the theory is not claimed to apply). What MDC is after, on the other hand, is to describe and interpret the scientific usage of the concept of “mechanism”. Hence the relativity of fundamental level (or of bottoming out) is based on MDC's agreeable observations regarding scientific practice (as presented in the quotation above).

In respect of deliberately aiming at addressing the scientific practice in life sciences, MDC's analysis can be likened to Bechtel and Richardson's. In both works mechanistic explanation is brought to the foreground—it may appear that Wimsatt or Brandon's paper does an equivalent thing, but let us not forget that in both of these earlier papers the major problem is reduction in science, and they give attention to mechanistic explanation for supporting their arguments in this particular issue. A further similarity between the two accounts is that both regard mechanistic explanation as an alternative to lawful (nomological) explanation. Bechtel and Richardson only *suggest* this point, at the end of their book within the final discussion of their analysis of heuristic strategies employed in life sciences (1993, 231-232). They indicate that the cases they examine (the cases from the history of science) rarely involve explanation by laws; and that instead of laws and theories, there are “explanation by composition of causes” (Mill and Cartwright are referred for this phrasing) and relatedly, models of mechanisms underlying phenomena. The point (that mechanistic explanation is in contradistinction to lawful explanation) is more pronounced in MDC 2000 (Section 3.2). Observing that laws as universal stable regularities have few applications in life sciences, MDC assert that they are substituted by mechanisms. MDC is confident that this practice is philosophically justifiable, because the activities of entities of a mechanism bring about the phenomena regularly and non-accidentally. Furthermore the regularities explained by mechanisms support counterfactuals, too (which is usually seen as a mark of lawful necessity)—the example given by MDC is: “[I]f this single base in DNA were changed and the protein synthesis mechanism operated as usual, then the protein produced would have an active site that binds more tightly. This counterfactual justifies talking about mechanisms and their activities with some sort of necessity.” (2000, 8) I believe MDC's readers need to see a stronger argument here, or perhaps a more elaborate example. Unfortunately this is all we are provided in their paper. (Later individual works by Machamer, Darden and Craver improve on this basis, though.)

These similarities and differences aside, the patent distinction of MDC's approach is its doing away with the “interactionist tradition” (2000, 4). MDC rightly cite Wimsatt, Bechtel and

Richardson and Glennan as belonging to this tradition, for whom the behavior of a mechanism is due to *whatever* is happening between its constituent parts (entities). In contrast, on MDC's account, *activities* relate entities and so, entities enter into diverse, distinct and irreducible *productive* relations that cannot be glossed over. I would also like to note another originality in the MDC paper, which is its inclusiveness. My present concerns allowed me to go over only about half of the issues covered by MDC, which included assertion and justification of a dualist ontology for life sciences, the nested hierarchy of mechanisms and their relative bottoming out, and adherence to the non-lawfulness of mechanistic explanation. Even if these comprise the most interesting and important parts of the paper, the other half of the issues brought forward by MDC deserves at least a mentioning. Here I do this by listing them: How scientists describe the set-up and termination conditions for the changes produced by mechanisms; the role played by idealizations, abstractions and sketches in developing a mechanistic explanation; the epistemic functions of a mechanistic explanation (how mechanistic explanations make phenomena intelligible¹⁶); and lastly, the multi-level character of mechanistic explanation that defies reductive understanding (of ontology, theory change and explanation). (Some of these will be addressed in the following discussions.)

I believe this much suffice to illustrate the richness of the paper by the MDC team. It also offers an impressive reading upon seeing how these issues and ideas are elegantly pieced together. Acknowledging these should allow us to appreciate MDC's influence on the subsequent philosophical musing on mechanism-talk, which I survey in the next section.

1.3 TENETS OF THE MECHANISMIC DISCOURSE

In the previous sections I attempted to make the case that the earlier works lack MDC's deliberate, forceful and comprehensive arguments for the (supposed) philosophical worth of the scientific mechanism-talk. I now want to contextualize these arguments and view the bigger picture—namely, the “mechanismic discourse” which owes its conception to MDC 2000 and has been advanced since then.¹⁷ The discourse reflects a certain stance taken with respect to the canonical

¹⁶ “Mechanism descriptions show *how possibly*, *how plausibly*, or *how actually* things work. Intelligibility arises not from an explanation's correctness but rather from an elucidative relation between the explanans (the set-up conditions and intermediate entities and activities) and the explanandum (the termination condition or the phenomenon to be explained.” (2000, 21) Let us acknowledge that the explanans-explanandum fit is itself an issue in epistemology. Some later works on mechanistic explanation problematize this, but MDC do not engage in this discussion.

¹⁷ A short list of notable works that shape and exemplify the mechanismic discourse: Darden and Craver 2002, Bechtel and Abrahamsen 2005, Craver and Bechtel 2007, Craver 2006 and 2007, Bechtel 2008, Craver and Kaiser 2013, Illari and Williamson 2011 and 2012, Andersen 2012.

form of lawful and reductive scientific explanation: Mechanismic philosophers argue that high-level phenomena (investigated in special sciences) are explained neither by appealing to laws nor by privileging the contributions of the lower-level states of affairs. They think that the scientific analysis of phenomena through mechanisms is a separate practice that must be given its own due. To this, of course, there have been objections raised on the grounds that mechanistic explanation is not free of reductive explanatory practices (e.g., Soom 2012) or cannot do away with laws (e.g., Leuridan 2010). Furthermore, mechanismic philosophers are challenged also from a broader perspective, on which the search for mechanisms is not as central as they think, and life sciences is a reductive enterprise in the final analysis (e.g., Bickle 2006; Weber 2005 and 2008).

These are certainly interesting topics for discussion, but I will have very little to say on these. In any case, the present section is the best place to enlarge upon the (purported) non-lawful and non-reductive character of mechanistic explanation. I shall do this with an *uncritical* reproduction of the discourse, for this seems to be the most appropriate way to portray the presuppositions that are in play in the advancement of mechanismic philosophy of science. In what follows, I give a sketch of the backdrop against which the mechanismic philosophers take position. To state briefly at the outset, the main line of debate concerns how different the subject matter of life sciences is (in comparison with that of lower-level sciences), and how the theories about these subjects are structured/presented. Considering these, I will simply present the mechanismic philosophers' views, just to give a picture of the ground they share. I will associate three tenets with this ground. My actual critique of the mechanismic discourse, on the other hand, will be about the two other typifying tenets.

Writing in 2001, Paul Thompson mentions five issues of importance in philosophy of biology. These are—in his phrasing—*the structure of theories, laws in biology, explanation and prediction, reductionism and fitness and selection* (2001, 17-20). It can be said that except the last one (which is a topic native to the field of biology), these topics concern inquiries in general philosophy of science brought into the context of biological sciences (cf., Griffiths 2011, Section 2). It is equally possible and plausible to question, say, the status of laws *in general* (regardless of the scientific field); but the claim is that at the level of complexity biology is concerned with, the issue of laws calls for a different treatment. Hence laws *in biology* becomes one of the major topics here. The same goes for *the structure of theories, reductionism and explanation and prediction*: They are perennial issues overarching the fields of science, but—it is claimed—the particularities in biology invite a different philosophical inquiry.

Undoubtedly these particularities are life phenomena and the evolutionary processes affecting and shaping them. In some other scientific enterprises which are usually considered as more fundamental, such as physics or chemistry (and the related disciplines, e.g., material science, physical chemistry) the objects of studies are completely immune to the workings of evolution. Presumably this entails *universality/generality* of the theories and the findings in these sciences, and the lack of them in biology. Let us recall what Crick says (previously quoted in the Introduction): “Biology has its 'laws' such as those of Mendelian genetics, but they are often only rather broad generalizations, with significant exceptions to them. The laws of physics, it is believed, are the same everywhere in the universe. This is unlikely to be true of biology. We have no idea how similar extraterrestrial biology (if it exists) is to our own.” (1988, 138) This remark concerns a difference between the *scope* and *strictness* of generalizations in biology and more fundamental sciences. (The extended version of this contrasts special sciences [including, for instance, chemistry, geology, economics and social sciences] with fundamental physics.) An earlier (probably the earliest) and classical elaboration of this argument is by Smart (1963). Smart draws a fairly intuitive comparison in terms of the *simplicity* and *homogeneity* of the objects of research. He observes that atoms (as subjects of physics) are unlike mice, diploid cells or chromosomes, because they are composed of simple constituents (electrons, protons and neutrons) which are ubiquitous in the universe. Living organisms (and their constituents), by contrast, are “vastly complicated and idiosyncratic structures.” (1963, 55) Smart further claims that biological theories (he mentions the theory of evolution and ecology) are not

typically “scientific” in nature. They are quite obviously “historical” subjects. They are concerned with a particular and very important strand of terrestrial history. No doubt there are analogous histories on remote planets, but in the theory of evolution we are concerned with the hereditary relationships of terrestrial creatures only, and so we are not concerned with laws in the strict sense. (1963, 59)

A later and more sophisticated argument against the existence of strict laws in biology is due to Beatty (1995). His “evolutionary contingency thesis” asserts that generalizations about the living world are either plainly mathematical, physical or chemical generalizations (or deductive consequences of these) or “distinctively biological”, meaning that they are about the historical/accidental/contingent products of evolution (1995, 46-47). Quite expectedly the philosophical literature on whether there are laws in biology (or on how to formulate the distinctness of biology—if it is really distinct) is rife with arguments advanced against these above¹⁸—but none of these will concern us in the present study.

¹⁸ Major papers include Sober 1997, Waters 1998, Mitchell 2000, Elgin 2006.

As is well-known, the early bits of philosophizing about science did not involve a distinguished perspective on biology. The philosophical picture of science at that time (1950s) was essentially modeled after physics; biology had to wait until the chemical structure of DNA was discovered and the methodology of the discipline thus became comparable to that of the fundamental science. (When J. J. C. Smart, for example, wrote in early 1960s that there were no laws in biology, this was in fact a swift philosophical reaction to the apparent change of status of this science.) It did not take long time for philosophers to recognize that biology was certainly left out of frame in this panorama of science. In fact physics, too, did not fit well into this framework of ideas about scientific theories, the framework that Putnam named as the “received view” (Thompson 2001, 17).¹⁹

On the received view, theories are axiomatic systems; they are *linguistic structures* and the explanations they provide are *sound deductive arguments*. The explanandum, which can be either singular events (e.g., the accelerated motion of a particular object in a certain direction) or regularities (e.g., accelerated motions of objects), logically follow from law statements coupled with premises that serve as correspondence rules between theoretical terms and observed entities and their attributes. The classical statement of this *covering law model of explanation* is presented by Hempel (1965; cf., Hempel and Oppenheim 1948). According to this model, the phenomenon is explained by the argument's providing an answer to a *why*-question: “[T]he argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred.” (Hempel 1965, 337 [original emphases]) The first emphasis in the quotation marks the *nomoc expectability thesis*, the distinguishing element in Hempel's model. It has also been a target for criticism, as we will see in passing in Chapter 2.

While the received view is centered on the model of scientific explanation (and its derivative, *prediction*) associated with Hempel, it is more truthful to view it as a broad paradigm extending to other aspects of the scientific enterprise (and, of course, goes well beyond Hempel's own opinions). Let me briefly mention the other important points about the paradigm. First, the laws of the covering law model of explanation are strict laws: Given that explanations are held as arguments, these must be interpreted as true *statements* of universal and ahistorical validity. (It is worth mentioning that Hempel himself adopts a more liberal stance with regards to laws. He is satisfied with the generality of “lawlike statements”, such as “All robins' eggs are greenish-blue.”) Second, theory-making is believed to start from empirical observations, leading to inductions and

¹⁹ The framework is given other names, too, including the “orthodox view” and the “syntactic view”. See (Craver 2002, 74, n. 3 and 4) on this, and for a brief list of the exemplars of the paradigm.

hypothesized generalizations tested over time, which—on the condition that they turn out as true—give way to formulations as laws (i.e., law statements). As a consequence of this, the influence of the theory on observation (theory-ladenness) is overlooked. Third, in Craver's words, the received view²⁰ “enforce[s] a stratigraphic picture of science and of the world – a picture in which ontological levels map onto levels of theory which in turn map onto fields of science. . . . On this caricature, theories at each level develop in relative isolation until it is possible to derive the higher level theory from the lower.” (2002, 57-58) This last point is also closely related with the belief that theory *follows* observation (the previous point). This is why theories (are believed to) develop in isolation: Layers of objective reality are subject to observation that is *unaffected* by any theoretical outlook. It later becomes possible to *reduce* one theory to another and this corresponds to a progress in science. On Nagel's (1961) classical account (an exemplar of the received view), intertheoretic reduction is taken as *explaining* the reduced theory by the reducing theory (e.g., explaining thermodynamics by statistical mechanics) and here, too, explanation is understood to operate as a formal derivation (following the rules of logic).

Starting from 1960s, the received view of science was challenged on all fronts; advocating the models of explanation and reduction, and the theory dynamics related with this framework evidently started to look as an unsustainable endeavor. Some of the attacks on the received view concerned the scientific enterprise in general (Kuhn 1962 and Feyerabend 1975 are the most well-known critiques), but even so, physics served as the main domain for the debates. Nevertheless there have always been critiques centered on the inapplicability of this framework to biology and other disciplines in life sciences. The pathbreaking advancements in molecular biology and neuroscience have certainly been quite influential on this, urging philosophers to ponder on the implications of these new developments in science.

We can thus view Wimsatt 1976 and Brandon 1984 (the two papers predating MDC 2000, which are mostly on reduction in science) against this background and regard them as contributions to the scrutiny of the received view. In writing that scientific explanation in life sciences has much to do with how phenomena are causally produced, they actually voice a dissatisfaction with interpreting explanations as arguments. More exactly, like most other philosophers of their times and today, they do not see the relation between explanans and explanandum in logical terms. Instead they emphasize the causal and constitutive aspects in this relation. The later contributions by MDC and the other mechanistic philosophers are in line with this account of explanation. The improvement

²⁰ Craver in fact labels it as the “*once* received view”, emphasizing the outmodedness of the paradigm (2002, 74, n. 4).

is the deliberative widening of the horizon for the philosophy of life sciences. In fact, as we are about to see, the mechanistic discourse that MDC have initiated is overtly set against the received view in virtually every respect. Seen from another angle, challenging the received view of science *still* functions as the context for the debates the mechanistic philosophy of science is engaged with.

No wonder, then, we observe Bechtel and Abrahamsen confronting the received view in the opening sentences of their “Explanation: A Mechanist Alternative” (2005), one of the major papers following the opinions in MDC 2000. The authors begin by underlining that “most *actual* explanations in the life sciences do not appeal to laws in the manner specified in the D-N model [deductive-nomological model, the essential version of Hempel's covering law model of explanation]” (2005, 422 [emphasis in original]) This is followed by the example of the *explanatory uselessness* of a law statement in biochemistry—the one concerning the ratio of oxygen molecules consumed to ATP molecules produced during ATP synthesis (it is 1:3). Bechtel and Abrahamsen assert that this generalization itself is a phenomenon requiring explanation, and life scientists answer *why* this occurs by explaining *how* it is produced.²¹ The authors also point out “the ubiquity of references to mechanism in biology, and the sparseness of references to laws”. (2005, 423) The prevalence of “mechanism” over “law” is emphasized by some other philosophers embracing the mechanisms-talk, too (Thagard 2003; Illari and Williamson 2012); but Bechtel and Abrahamsen is unique for offering a quantitative fact in support of this: They say the search for “mechanism” and “protein synthesis” in the titles of journal articles produces 302 results, whereas “law” and “protein synthesis” generates no result.

The claim that in special (high-level) sciences laws are not invoked as much and as centrally as envisaged in the received view is defended (by the mechanistic philosophers) on practical grounds. This fact is cited also by Illari and Williamson as the first reason for why they see mechanistic explanation “far more promising [compared to the covering law model] as an account of explanation *across much of the sciences*.” (2012, 2 [emphasis added])²² The second reason they bring forward rests on the idea that these sciences in fact lack laws, and have done quite well without them:

²¹ This is directly set against Hempel's above quoted assertion that explanations answer why-questions. Thagard makes the same point with Bechtel and Abrahamsen: “The primary explanations in biochemistry answer how-questions rather than why-questions: How do cells get energy? How do cells divide? How do cells form tissues?” (2003, 238-239)

²² With respect to the scope of applicability of mechanistic explanation, Illari and Williamson are perhaps the most ambitious among the mechanistic philosophers. They think the core ideas of mechanistic philosophy applies not only to special sciences in general, but also to cosmology and solid state physics (see fn. 1 in their paper).

Those “laws” that do exist in the special sciences are nothing more than generalizations backed up by an understanding of the mechanisms underlying those laws. . . [A] reductive faith that there will ultimately be discoverable laws everywhere has been seriously empirically undermined by the huge success of science without laws – such as the extraordinary amount now known about the detailed mechanisms of protein synthesis – and should be abandoned. (2012, 3)

The “no laws in special sciences” claim is a thesis going beyond the practical fact of the scarcity or the unimportance of the *use* of laws in special sciences. One important challenge to this is proposed by Mitchell (1997 and 2000) who adopts a pragmatist understanding of laws. For her, generalizations about regularities with sufficient stability can be counted as laws, as long as they play roles in explanations, predictions and manipulations. Leuridan (2010) uses Mitchell's account of laws in order to ground her critique of the mechanistic discourse. In the title of her paper she asks “Can mechanisms really replace laws of nature?” Her negative answer to this is based on the argument that both MDC's and Bechtel and Abrahamsen's characterization of mechanisms ontologically depend on regularities, and these regularities are of the sort Mitchell embraces as laws. In response to this, the mechanistic philosophers hold that Leuridan confuses the explanatory order (Andersen 2011, Craver and Kaiser 2013): Mechanisms explain the regularities in a manner alternative to laws are supposed to exhibit. The formulation by Andersen is: “Regularities are the explananda for which mechanisms are the explanantia.” (2011, 5) (The explanation of regularities by mechanisms will be a topic in Chapter 4.)

With this short sidetracking from Illari and Williamson's list of reasons to support mechanistic explanation, I intended to have a glimpse of the current and active debates with the mechanistic discourse. Yet we did not depart much, as the third reason in their list is the explanatory superiority of mechanisms over laws. Illari and Williamson join MDC who write that laws do “no philosophical work” (2000, 8): “[I]t is hard to see how they could explain in the way that mechanisms explain.” (2012, 3) They are also in agreement with Bechtel and Abrahamsen, Andersen, and Craver and Kaiser on the idea that laws themselves await explanation, and this can be and is done by using mechanisms. This idea should naturally remind Glennan's position (1996), but Illari and Williamson note where they diverge from it: As we saw, for Glennan what are “mechanically explicable” are the non-fundamental laws, and the fundamental laws of physics sets the limit for mechanical explanation. Illari and Williamson, on the other hand, state their doubt about there being the lowest, fundamental level of reality (2012, 3).

Given these consistent emphases, the proposed non-lawful character of the mechanistic explanation can be put as one of the tenets of the mechanistic discourse. I now turn to two other

tenets which are also characterized by the mechanistic philosophers' dissatisfaction with the received view. These concern *the structure of theories* and *the issue of reduction*.

Bechtel and Abrahamsen argue against the understanding of *theory* in the received view by noting the non-linguistic structure of biological theories. They highlight the use of “visuospatial representations” of the described and/or explained phenomena: “a photograph showing an apparatus, a micrograph revealing a subcellular structure, a graph summarizing the relation between two or three variables in an experiment, and so forth.” (2005, 427) When mechanistic explanation is considered in particular, Bechtel and Abrahamsen observe that *diagrams* do a prominent job by “provid[ing] a vehicle for keeping in mind the complex interactions among operations [of mechanisms].” (2005, 427) According to the authors, the frequent use and the success of diagrams in representing mechanisms is due mostly to the spatial conveyance of information (which is unavailable to linguistic representation). Diagrams can display the spatial relations within mechanisms (i.e., spatial organization of the parts) and the temporal order of the operations (the arrow signs may be used for this), both of which are critically important in mechanistic explanation. Colors and shapes can also be visualized by the help of diagrams (2005, 428). Idealizations and the omissions of details are also (usually) part of the practice of visually representing mechanisms. MDC introduced the term “mechanism schema” to label these and proposed that scientific progress in life sciences resided in improving on the schemata (2000, Section 5.3). Watson's diagram for the central dogma is the example they use in their paper (taken below). They view “transcription” and “translation” in this diagram as “schema terms”—terms that stand for activities in higher-level mechanisms. On their view the schema abstracts away the underlying mechanisms producing these activities (2000, 16). MDC use a further label, *mechanism sketch*, for representations where the bottom out entities and activities are not known. They contrast schemata and sketches on an epistemic ground (2000, 18); with scientific progress, models of mechanisms (such as diagrams²³) approach to *complete* descriptions, where no detail of how a mechanism produces a phenomenon is left out. The content of the representational relation between the model and the objective, *real* mechanism is discussed by Glennan (2005). Characterization of the criteria for the explanatory adequacy of mechanism schemata constitutes Craver's major philosophical endeavor (2006 and 2007), which I cover in the chapters to follow.

²³ Of course it is not claimed that diagrams are the only means of modeling mechanisms. For instance, a later paper by Bechtel and Abrahamsen examines computational modeling of mechanisms with differential equations (2010).

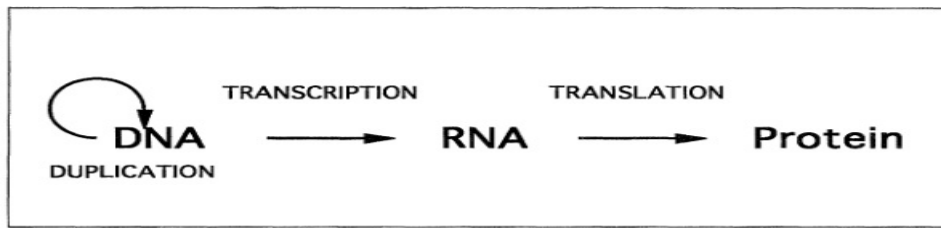


Figure 1-3. Diagram showing The Central Dogma (Taken from MDC 2000, 16)

With the exception of Bechtel and Richardson 1993, support for the mechanistic explanation has always involved a special concern for the issue of (explanatory) reduction. As we saw, reduction was *the issue* in the early exposition of the importance of mechanistic explanation (Wimsatt and Brandon—but let us recall, their views differed on whether to count mechanistic explanation reductive or not). Bechtel and Richardson, on the other hand, only *touch upon* the subject in the first edition of their book, as one of the last concluding remarks (1993, 242-243). They write that the strategies of decomposition and localization they explore throughout the book could be seen as suggestive of reduction (due to the practice of analyzing a system by its *lower-level* components). Yet they immediately acknowledge that this is not the reduction as understood in the traditional sense (i.e., in the received view of science), because it lacks the supposed lawful connection between what is reduced and what is reducing (the higher and lower levels).

Here again it is MDC's paper that first pronounced the distinctive feature of the mechanistic explanation (as characterized in the mechanistic discourse) regarding reduction. The keyword is the explanation's being “multi-level”, meaning that mechanistic explanation does not privilege any layer but may span many of them in both directions. MDC's example is the activity of the sodium channel as being one component of the action potential generation mechanism in the neurons. (This mechanism is a popular example in the mechanistic literature and will be examined extensively in the next chapter.) MDC write that this activity cannot be understood properly unless the higher-level phenomena (the release of neurotransmitters, the transmission of signals) are also understood: “Higher-level entities and activities are thus essential to the intelligibility of those at lower levels, just as much as those at lower levels are essential for understanding those at higher levels.” (2000, 23) This is echoed by Craver (2007) and in Bechtel 2011: “Mechanistic explanations are always multilevel accounts, integrating information about parts, operations, and organization within the mechanism with characterization of the phenomenon exhibited by the whole mechanism.” (2011, 538) What is more, while they do not change their original view in the main text, Bechtel and

Richardson also adopt this approach to reduction in the comprehensive introduction to the second edition of their *Discovering Complexity* (2010, xxxix-xl). Bechtel and Abrahamsen, on the other hand, voice the slightly different observation that traditional reduction has no use in mechanistic explanation. For whereas in the former, states of affairs at the lowest level account for *all* phenomena (happening on higher levels), each successive mechanism (described in mechanistic explanation) accounts for *different* phenomena (2005, 426).

Apart from the views concerning laws in explanation, the structure of theories and reduction, there are two other tenets of the mechanistic discourse that I want to mention in brief. These are separate from the three I have been examining, for they do not involve a position to be taken with respect to the received view. As indicated earlier, it is these two that this study is concerned mostly with. We are familiar with the first of these: it is the motivation of the mechanistic philosophers to capture the *manifest scientific practice* itself. They give an importance to the frequent and putatively noteworthy use of the notion of “mechanism” in high-level sciences, which makes their philosophical endeavor descriptive in general. But it is important to differentiate the two senses of this notion, and the respective referents (i.e., machine and causal mechanisms). This is the main issue I deal with in the next chapter.

The other tenet (the final one to be mentioned) concerns the realist tendencies of the mechanistic philosophers. Among them, eminently Craver and Illari and Williamson believe and assert that mechanisms are objectively existing, real features of the empirical world, that can be discovered. I label the brand of scientific realism exhibited by these philosophers as “mechanistic realism”, and in this very study I attempt to show why it is wrong to be a mechanistic realist. Their position has been subject to *scientific antirealist* charges (Colombo, Hartmann, and van Iersel [2015]), but mine will be a realist critique at the end of the day. Naturally I have a lot to say on mechanistic realism, so let this much serve just as an indicator of what is to come.

Taking stock, we observe that five tenets typify the mechanistic discourse:

1. The claim that mechanistic explanation is non-lawful (and is an alternative to the lawful explanation model of the received view),
2. The claim that mechanistic explanations provide understanding by non-linguistic entities,
3. Characterization of mechanistic explanation by its multi-level scope, defying the traditional understanding of reduction(ism),
4. Deliberate attempt to capture the manifest scientific practice (in high-level sciences),

5. The scientific realist stance (mechanismic realism)

For the purposes of the critique to be developed, it will be important to distinguish “mechanismic” from “mechanistic”. It has become customary to call mechanismic philosophy as “the new mechanistic philosophy” after Skipper and Millstein 2005, and to label the mechanismic philosophers as “the (new) mechanists” (or by variants of these).²⁴ But this creates the false impression that the *mechanismic* philosophy of science stems from the earlier *mechanistic* approaches in biological thought.²⁵ As Nicholson indicates, the earlier approaches bore a reductive and materialistic perspective with regards to life phenomena, which plainly has no relevance on the mechanismic discourse (2012, 153-154). It is for this reason that Nicholson avoids the usage of “mechanistic”, and instead prefers the labels “mechanismic program” and “mechanismic explanation” (2012, 154). Although I find these as sound choices, I cannot comfortably adopt Nicholson's labels. For “program” could be the right term if it was the case that the totality of the philosophical output in the mechanismic literature involved a pre-arranged individual contributions. The collection of works, though, appears to be more loosely organized (if at all). This is why I prefer the notion of “discourse” instead, which implies the stabilization of certain ideas through their repeated (but not organized) forwarding by different authors. Construed as such, “mechanismic discourse” can be used *almost* interchangeably with “mechanismic philosophy of science”. I cannot say that they mean exactly the same thing, but I admit the difficulty to put the difference in words (which may not be very important for our purposes).

On the other hand I will not be using “*mechanismic* explanation”. According to the view proposed here, *mechanistic* explanation and *its mechanismic interpretation* are two different things. The latter concerns a particular (and problematic) understanding of scientific modeling through mechanisms. Accordingly I will use “mechanismic *modeling*” for the mechanismic philosophers' view that models of mechanisms represent real mechanisms harbored within scientifically investigated systems. On this view, a scientific realist attitude towards mechanism-talk entails mechanismic realism. In what follows I will work towards countering this idea.

²⁴ See e.g., Fazekas and Kertész 2011, Wright 2012, Craver and Kaiser 2013, Illari 2013, Colombo et. al. 2015, Kaiser and Krickel *forthcoming*.

²⁵ It must be noted that except for Bechtel (e.g., 2008, 10) no mechanismic philosopher regards *mechanismic* philosophy as a continuation of the earlier *mechanistic* philosophy. (See Nicholson 2012, 158 for a critique of Bechtel.)

CHAPTER 2

DISAMBIGUATING “MECHANISM”

2.1 A LOOK AT CHARACTERIZATIONS OF “MECHANISM”

In the previous chapter I introduced the mechanistic discourse with its five typifying tenets. My main concern in the present and the following chapters will be the metaphysical discussions this discourse calls for. The primary issue that I take up in Chapter 2 and 3 is the ontological natures of mechanisms.

In the present chapter I mainly scrutinize Craver's account of mechanistic explanation put forward in his *Explaining the Brain* (2007). My most important concern here is to take the first steps toward criticizing the mechanistic discourse for its “ontologization of causal mechanisms” (Nicholson 2012, 158). Craver being a very good case in point, mechanistic philosophers have the tendency to regard mechanisms as belonging to the “furniture” of the world, awaiting to be discovered. For them, as relevant scientific investigations give their fruits, models of mechanisms get better and better in truly representing the causal structure underlying the empirical world. An ideal (or “complete”, in Craver's terminology) model of a mechanism, so to say, is the one that exposes the phenomenon it explains as situated within this structure. It can do so, because the mechanism also belongs to the causal structure, by being the relevant localized portion of it where the phenomenon is produced. This is the line of thought that is targeted in this dissertation (i.e., *mechanistic realism*). My arguments in this chapter are intended to be mostly expository: That is, in essence, they are not characterized by objections to the claims made within the mechanistic discourse; instead, they aim to offer a critical yet sympathetic reading. In the end, hopefully, they should take us to a clear depiction of the position of the mechanistic philosophers (that I do not endorse). This attitude will change in the next chapter, where my position countering the mechanistic philosophers' will be much more pronounced.

To get the gist of the matter, let us start with considering three extant characterizations of “mechanism” :

Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. (MDC 2000, 3)

A mechanism is 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)

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)¹

Each of these definitions given by mechanistic philosophers can more or less be regarded as an improvement on the preceding one (owing to the fruits of the debates made in the decade following the publication of the MDC 2000). Setting the differences between them aside for a moment,² we see that mechanisms are commonly identified by the phenomena they produce (or are responsible for). One thing to note is, in MDC's definition, the “regular changes” that mechanisms are productive of “from start . . . to finish . . . conditions” stand for such phenomena. This characterization has been found narrow by some and wrong by some others. As MDC themselves acknowledge (2000, 11-12), the start and finish conditions of mechanisms are not completely objective but depend on practical-epistemic concerns; their descriptions involve idealizations and the privileging of certain objects³, properties and states (cf., Darden 2008, 960). However, regardless of these conditions' ontic statuses, the “start” and “finish” impose a sort of linear understanding of mechanisms, which does not apply well to the wealth of cases. Illari and Williamson remark that

‘start’ and ‘finish’ conditions are not even an aspect of all of our mechanism descriptions, far less of all mechanisms. They are pragmatic aspects of the descriptions we give of some mechanisms – but not all. Cell mechanisms such as the Krebs cycle are cyclical. They are continuing, having no real start or end. (2012, 122)⁴

This is the narrowness objection raised against MDC's definition. There is also the objection to the regularity requirement found in this definition, voiced by Bogen (2005). On Bogen's account, mechanisms cannot be characterized by some “regular changes” they are said to be productive of.

¹ The definition found in (Craver 2007, 6) is not much different than this: “[M]echanisms are entities and activities organized such that they exhibit the explanandum phenomenon.” Whether the entities and the activities consisting in mechanisms “exhibit” or “are responsible for” the phenomenon does not raise an issue of significance. On the other hand, the nature of the relation between mechanisms and the phenomena they explain will be of interest in the forthcoming sections.

² I will not be concerned with whether to view what the entities do as “activities” (MDC, Illari and Williamson) or “operations” (Bechtel and Abrahamsen). See Tabery 2004, Illari and Williamson 2012 and 2013 for the discussion. Except while discussing Bechtel and Abrahamsen's approach, I will prefer the use of “activities” over “operations”.

³ Hereafter, “object” will be used as a substitute for “entity” as understood in the mechanistic discourse. The preference is due to “entity” being a much more loaded notion than “object” (as noted in Kaiser and Krickel *forthcoming*, Section 2.3). Despite this, the mechanistic philosophers' term-of-art “component entity” will be kept (I will never write about a “component *object*”).

For many of the mechanisms investigated in high-level sciences do not regularly produce the phenomenon; the failure of the mechanism for a phenomenon is often the rule rather than the exception. Bogen's own example is the firing of a post-synaptic neuron due to the electrical activity in the pre-synaptic neuron (i.e., the etiological mechanism for neurotransmitter release), which is a stochastic process where “failures are by no means uncommon” (p. 412, fn. 20). He takes this to mean that there are many cases where what is *regular* for a mechanism might be its *not* producing the phenomenon. But given that mechanisms do explain phenomena, his argument is for (1) dropping the regularity requirement for a good characterization of mechanisms to reflect the success of mechanistic explanations, and (2) embracing a non-regularist (i.e., Anscombean instead of Humean) account of causation that disentangles causal *production* from causal *regularities*. Bogen's critique has initiated a fruitful debate on the mechanistic philosophy, which I survey in Chapter 4. Suffice it to say here that due to his contribution, the general attitude among the mechanistic philosophers has become looking for a more refined sense of “regularity” to implement in the mechanistic discourse.⁵

There is also a contrast to note between Bechtel and Abrahamsen's and Illari and Williamson's definitions: Whereas in the latter a mechanism is construed to be responsible for *the* phenomenon (singular), according to Bechtel and Abrahamsen, the mechanism can be responsible “for one or more phenomena.” As we shall see below, a mechanism is characterized by a *single* phenomenon; so on a stringent reading, the mechanism cannot be responsible for more than one phenomena. (Hence the use of the singular noun in Illari and Williamson's definition.) But here, Bechtel and Abrahamsen rather intend to emphasize the embeddedness of mechanisms within larger ones. Therefore on a more liberal reading, in virtue of being a part of larger mechanisms (this is their stipulation), mechanisms might be taken to be responsible *also* for the phenomena that actually the larger ones are supposed to explain. The example provided by Bechtel and Abrahamsen is the heart's pumping blood: They take the heart *itself* as a mechanism with the function of pumping blood, and remark that this organ “is itself part of a larger mechanism, the circulatory system, that includes such parts as veins, arteries, and the blood itself.” (2005, 424) Accordingly, for them, the heart is responsible both for pumping the blood and also the various phenomena associated with the circulatory system. Furthermore, as a heuristic move to demonstrate how their definition applies to cases, the authors mention the *parts* of the heart (chambers and valves) and the *operations* these

⁴ See also (Bechtel and Abrahamsen 2005, 423, fn .5).

⁵ Machamer's (the “M” of MDC) voicing a dissatisfaction with the emphasis on “regular” is noteworthy in this respect (2004, 37, n.1).

perform (contraction, relaxation, prevention of backflow of blood). It is intuitive to regard these plus their organization as “a structure performing a function”, namely, a *mechanism* according to their definition. Yet there is a perplexity with this conceptualization that I want to address below, in the rest of the section. (This should work as an introduction to the main problematic of this chapter.)

Mechanismic philosophy attends to mechanisms because of their explanatory significance. Sticking with Bechtel and Abrahamsen's terminology, mechanistic explanation works by describing how the component operations of the component parts are (causally) organized. Now, this should amount to the description of a particular mechanism *M*. On the other hand, such description is provided with the intention to account for a phenomenon *P*; the phenomenon is *individuating* of the mechanism in this sense. As Craver says, “[t]he boundaries of mechanisms—what is in the mechanism and what is not—are fixed by reference to the phenomenon that the mechanism explains.” (2007, 123) This is why “nothing is a mechanism *tout court* - mechanisms are mechanisms for phenomena.” (Illari and Williamson 2012, 130)⁶ Therefore, referring back to the example at hand, the organized operations of the parts of the heart (call it Φ ⁷) make up a mechanism *M* for a particular phenomenon *P*. But then, properly speaking, this *M* cannot be the heart's pumping blood, as Bechtel and Abrahamsen supposes. For the heart's pumping blood is what Φ explains; it is the *explanandum phenomenon* accounted by the “orchestrated functioning” of the *parts* of the heart which is the *explanans*. What must be clarified and stressed here is that, Φ is this very orchestrated functioning: In no sense is the mechanism *M* distinct from Φ , in no sense is there a mechanism *and* its orchestrated functioning, because a mechanism is already the organized operations of (component) parts. I shall dwell more on this in what follows, by attending to the semantic ambivalence in the notion of “mechanism”. That discussion should make it clear that the mechanisms are explanatory only when construed as organized activities/operations of parts of a whole. They are not explanatory if they are construed as an assembly of objects.

The historical and rather established use of the notion refers to machines (mechanical or electronic artifacts), which are wholes with objective boundaries, made up of parts for a known, designed purpose. Machines themselves are objects that are usually active: their parts perform actions and the machine is of use due to the organization of these. An inactive machine is also easily

⁶ In another paper, Illari and Williamson use this for contrasting mechanisms with processes: “[B]iochemists and evolutionary biologists consistently use ‘mechanism’ to describe only something that has a function, using the word ‘process’ when there is no function. So mechanisms are mechanisms *for*, while processes are just processes in themselves.” (2011, 826 [original emphasis])

⁷ This is a notation introduced in (Craver 2007, 212) and it stands for “the property (or family of properties) of being a given components-plus-organization.” By “components” we should understand the entities engaged in activities (Craver's and Illari and Williamson's terminology), or the parts performing operations (Bechtel and Abrahamsen).

conceivable, though—think of a stopped clock, for instance, or a moka pot not put over a heat source. One can still talk of and possibly “handle” (all senses of the word implied) a machine even when it is inactive. Now, some of the mechanism-talk in science is based on likening natural entities (or their collections) to machines, where “mechanism” is used in this particular sense. Here are two examples:

DNA is a mechanism created by the cell to store information produced by the cell.
(Commoner 2014)

Is it really true that the heart works like a pump? A pump sucks in fluid from a reservoir, which is a hydrostatic system and not a hydrodynamic one. In the circulation, on the other hand, not only is blood ejected from the heart, but it flows into the heart. The heart is a mechanism inserted into the blood circuit, and so it is a very peculiar kind of pump.
(Manteuffel-Szoege quoted in Furst 2015)

The DNA and the heart being objects (or at least, concerning the former, object-like entities), made up of parts and having biological functions, a “machine image” (Levy, 2013b) lurks behind the use of “mechanism” these quotations exemplify. Hence I label this sense as the “machine mechanism” sense. Hereafter, I will use the shorthand notation “Machine-M sense”; this should not be confused with “machine mechanism”, which I use as the referent of this sense. Although I take the term from Nicholson's paper (2012), it should be noted that he uses “machine mechanism” with a different meaning. In his lexicon, the machine mechanism is the *inner workings* of machine-like structures—that is, the operations by which a machine performs its function. According to Nicholson, mechanistic explanations work by likening real systems to machine mechanisms (2012, 154).⁸ According to the terminology to be preferred in my study, though, real systems cannot be likened to machine mechanisms, because the latter are just aggregates of objects that need not exhibit behaviors, whereas the former are wholes made up of occurrences. (But I grant that mechanistic explanations work by likening real systems to the *inner workings of machines*, for which I do not propose a specific name.)

Having said these, the arguably much more widespread use of “mechanism” in the scientific context refers to something different than a machine mechanism. The three definitions given for “mechanism” cited above aim at capturing this second sense of the notion (hereafter, the “Causal-M sense”, which I take to refer to *causal mechanisms*⁹). This we can say without doubt, because “mechanism” *cannot* refer to a *mechanism-as-explanans* when it is used in the Machine-M sense. A

⁸ This can be compared with what Levy (2013a) terms as “strategic mechanism”.

⁹ Again, the labeling is based on Nicholson's articulation.

DNA, a heart, or any object (object-like entity) *per se* cannot be the explanans in an explanation; it can at most be a part of the explanans. (Same goes for the explanandum: No object *per se* can be the subject of an explanation.) An explanans (or an explanandum) might be a token- or type-level event, state, situation, process, etc., but never an object. Here, of course, I appeal to the commonsensical distinction between ontological categories that we can think of as the causal relata and objects, which are not the causal relata. After Kaiser and Krickel *forthcoming*, I will use “occurrence” for anything that can be a relatum in causal relations. The term Kaiser and Krickel actually use is “object-involving occurrent”. However I want to reserve “occurrent” as an adjective for properties (see Chapter 4), and I think we can do away with “object-involving” as long as we keep the following in mind: that there can be no bare occurrences, and that *many* objects are involved for there to be an actual occurrence. On this perspective, then, an “activity” as understood by the mechanistic philosophers is an occurrence. In accordance with my preference for “object” and “component entity” (fn. 3 above), hereafter I will be using “occurrence” and “component activity”.

The “mechanism”s in the quotations below are all exemplars to the notion's scientific use in the Causal-M sense, and it is clear that they do not refer to objects. Their explanatory roles are subtle (or somewhat concealed) in the particular contexts, so let me reveal them in turn. The first quotation is from an article named “The Mechanism of Leaf Morphogenesis”. The author A. J. Fleming discusses the causal processes underlying plant leaves' developing their shapes—in particular, how and whether cell division is involved in leaf morphogenesis (because this is a debated issue in plant developmental biology). Fleming's argument is, leaf morphogenesis is due to both cell division and the physical extensibility of the cell wall (i.e., a phenomenon independent of cell division). Reading the following excerpt with this background information should make it clear that Fleming is *not* discussing *machine* mechanisms. For it is not particular objects (or a group of objects) within the plant that explains leaf morphogenesis, but the causal relations and the causal organization they are involved in. I will dwell on this later (as a part of my main thesis), saying that causal mechanisms explain phenomena by *representing* these relations and the organization. Nevertheless, in the discussions to follow, it is much simpler to ignore the representational aspect of causal mechanisms. I will argue *as though* I think causal mechanisms are made up of the relevant links on causal chains scientists are interested in (put differently: as though they are harbored within the scientifically investigated real systems). This is an intentional mistake that I will not correct until next chapter.

At the organ level, cell division-dependent and cell division-independent mechanisms of morphogenesis are co-ordinated to generate structures of “appropriate” size and shape. By balancing these mechanisms the plant can both co-ordinate growth rates with

environmental/nutritional status and can provide a failsafe mechanism by which if one element is removed or not functioning properly, a compensatory mechanism can cut in to ameliorate the situation and maintain at least the basic morphogenic patterns required for plant function[.] (Fleming 2002, 20)

The second excerpt is from a lecture given by an eminent medical doctor, Louis N. Katz, known for his work in cardiovascular research. The lecture, entitled “The Mechanism of Cardiac Failure”, was delivered at a time when the research in the field had just been rewarding. We should note how the two mechanisms he mentions are associated with occurrences, and not with objects.

Low output of the heart may ensue for a number of reasons and need not be primarily cardiac. When, however, in the course of *shock* or *peripheral vascular collapse*, or any other extracardiac mechanism leading to low output, incompetence of the heart ensues, then the latter can contribute to the low output.

High output of the heart is a compensatory mechanism which enables the body to get an adequate supply of oxygen for the tissues, when this is impossible at normal levels of output. (Katz 1954, 663-664 [emphases added])

Similar uses of “mechanism” in Causal-M sense are quite ubiquitous in scientific research. To reiterate, the Causal-M sense is more popular than the Machine-M sense in the scientific mechanism-talk. This is an assertion made with an educated guess (for it undoubtedly demands a herculean effort to survey the whole scientific literature), relying on the observation that a machine mechanism per se is not explanatory and therefore, usually cannot be the referent of “mechanism” in explanatory contexts.

Recalling from the previous chapter, mechanistic philosophers observe the extensive scientific use of “mechanism” and state this as their motivation to articulate on the notion. However, they do not seem to pay an equivalent attention to differentiate the senses involved. This is a remarkable mistake when the metaphysics of mechanisms is of concern. For a machine mechanism is an assembly of objects whereas a causal mechanism is the causal organization of occurrences which involve these objects (in mechanistic discourse: entities engaged in activities).¹⁰ Not appreciating this results in supposing the explanans to be an autonomous object that looks like a machine (like the DNA or the heart is) capable of some functions in a higher-level system (the cell, or the circulatory system). (This autonomy is to be elaborated later in the chapter.)

The main content of the present chapter is based on a variant of this confusion found in Craver's *Explaining the Brain* (2007), which I take as the best exemplar of the mechanistic philosophy.¹¹

¹⁰ In Chapter 3, I will argue for a difference in their ontic statuses, too.

¹¹ Although I mostly draw upon the book, in several places I also refer to Craver 2006 and 2008. These are published papers whose contents are covered in the book to a great extent, but where (occasionally) some ideas are given preferable expressions.

Below, I first overview Craver's construal of (constitutive) mechanistic explanation. Importance is given to his motivation for a normative account of scientific explanation in high-level sciences. The overview also includes the introduction of the phenomenon of action potential and its currently accepted explanation. The action potential is a case that I frequently use throughout the dissertation. It is also a case by which Craver makes the point that good scientific explanations (in neuroscience) are good because it is usually observed that they describe mechanisms. The critical discussion of this very point will be left to the next chapter. The discussion in this chapter (regarding the explanation of the action potential) will instead function as a stepping stone *en route* to enlarging on key notions in mechanistic explanation.

2.2 CRAVER'S ACCOUNT OF CONSTITUTIVE MECHANISTIC EXPLANATION

There are several book-length studies building on the mechanistic discourse. Bechtel 2008, Steel 2007 and Tabery 2014 could be cited alongside Craver's book in this respect. However, among these, only Bechtel does a job akin to Craver's project, which is the straightforward characterization of mechanistic explanation (in neuroscience). Let me justify my choosing Craver's book (instead of Bechtel's) for scrutiny, then. Craver's project should be seen as a natural follow-up to MDC 2000, and as the perfect exemplar of the mechanistic discourse. Accordingly, it is quite apt to focus on his book, given that my study is a critique of this discourse. Bechtel 2008 reproduces the discourse, too, especially with regards to the explanatory framework advocated. (Bechtel is not concerned with the metaphysical implications of the discourse—unlike Machamer, Craver or Illari.) The distinguishing marks of Bechtel's book are the emphases on the epistemic conception of mechanistic explanation (i.e., how scientists *get to know* about the “mind-brain” [2008, x])¹² and the avoidance of explicit commitment to scientific realism. As I have already indicated several times, the problem I want to expose in the mechanistic discourse is the scientific realist attitude it professes. Given this, Bechtel's approach falls outside of my critique. Craver, on the other hand, defends a more controversial position (ontic conception of mechanistic explanation and mechanistic realism); in fact his book can be read as furthering the mechanistic discourse for an articulate and resolute defense of this very position. This is why he represents the unmistakable target of my criticism. It would not be an exaggeration to claim that Craver 2007 is the embodiment

¹² A slightly more detailed look at the epistemic vs. ontic conceptions can be found in the opening paragraphs of that chapter.

of the mechanistic discourse. The book and its author has been consistently addressed in the literature on mechanistic philosophy of science; I take this as the concrete justification for my focusing on *Explaining the Brain*.

Craver's main purpose in *Explaining the Brain* is to develop and promote a “mechanistic model of constitutive explanation” (2007, 111). In a nutshell, constitutive (mechanistic) explanation is provided by showing the relation between the behavior of the whole (the explanandum phenomenon *P*) and the behaviors of its parts (the components, whose causal organization is said to make up the mechanism *M*). The relation is exposed by showing the “constitutive relevance” (a term of art) of each component in *M* to *P*; it is this exposition that amounts to the explanation for *P*. I must note that the explication of “constitutive relevance” is crucial for the model of mechanistic explanation advanced by Craver. It is subjected to considerable scrutiny in the literature, mostly because this relevance is needed to be set apart from *causal* relevance, and Craver's approach has been attacked on the ground that this is a metaphysically untenable task.¹³ As I see it, none of the discussions I am concerned with hinges on the metaphysical issues encountered here. Therefore I shall only enlarge on the motivation to distinguish constitutive and causal relevancies.

Mechanistic explanations (ME) fall in two types: etiological and constitutive. The most conspicuous distinction between them is based on the temporal aspect of the *M-P* relationship. In etiological ME, *P* is the end result of *M*, that is, it temporally follows the latter. The release of a neurotransmitter (*P*), for instance, is an outcome of the transmission of the signal (the action potential wave) along the neuron's axon; it occurs after the signal reaches the neuron's end, and what explains *P* is the series of events (the etiological mechanism *EM*) all occurring before it. Put in other way, *EM* is the causal chain leading to *P*. As can be inferred from these, the explanatory relation between *EM* and *P* is causal in nature (*EM causes P*). In constitutive ME, though, the explanandum phenomenon *P'* temporally overlaps with the mechanism (the constitutive mechanism *CM*) explaining it. If we take *P'* to be the transmission of the signal, none of the occurrences in this causal chain (active and passive transportation of various ions) is actualized *before P'*; rather, the organized occurrences *altogether realize P'*. As intimated by Craver (2007, 153), there is a logical dependence of *P'* on the components of *CM* in many such cases. Since (he thinks) this dependence is non-causal,¹⁴ Craver cannot couch it in terms of causal relevance. Hence the concept of constitutive relevance enters the picture. It is the relation between a component of *CM* and the *P'*,

¹³ See Leuridan 2012, Baumgartner and Gebharder 2015. See also Ylikoski 2013, which can be read as a study on how causal and constitutive explanations should be distinguished.

¹⁴ This should not be confused with the interrelations *between the components* being causal.

and—ignoring the technicalities¹⁵—Craver puts this as a part-whole relationship where the relata are *mutually manipulable* (2007, 154). His analysis of the mutual manipulability requirement is in essence the application of Woodward's theory of *causation* (2003) to phenomena located at different “levels of mechanism”.¹⁶ The aforementioned critics have been prompted by this part of his theory (unsurprisingly) which I will not be concerned with. I will instead concentrate on the clarification of the correct part-whole relationship that is inherent in mechanistic explanation. Before moving on with the overview of Craver's work, let me state that etiological ME will also be out of my scope. This is because, either, the debatable aspects of the mechanistic discourse (*i-*) are solely due to the way constitutive ME is elaborated, or (*ii-*) can be shown to be relevant both in etiological and constitutive ME. Therefore etiological ME is not a type of mechanistic explanation that requires distinctive treatment. It is common among both the mechanistic philosophers and their discussants to mention etiological ME just in passing, if at all. Hereafter, then, by “mechanistic explanation”, I will be referring to the constitutive type.

Craver wants his account of constitutive explanation to be a *normative* model of mechanistic explanation, that is, a model that distinguishes good and bad explanations.¹⁷ A good explanation that conforms to Craver's account must provide the description of a *real* mechanism, which is “all and only the component entities, activities, properties and organizational features that are relevant to the multifaceted phenomenon to be explained.” (2007, 111) In particular, the given account of explanation should be able to distinguish conjectures about how the phenomenon might be produced by a mechanism which may not really exist (i.e., *how-possibly* models) from descriptions of *real* mechanistic components, activities, and organizational features (i.e., *how-actually* models). Furthermore, a mechanistic model for the phenomenon is good only if it is *complete* (in contrast to a *sketch*), meaning that, it should involve no gaps in the explanation or filler terms that stands for non-understood processes (2007, 112-114; see also Craver 2006, Section 6).

Craver sees the task as improving on Cummins's functional analysis (1975; 1983) within the systems tradition (2007, 110). The trademark of this tradition is to regard the *explanandum* phenomenon as a result of the specific organization of the parts making up a system. According to Craver, the systems tradition is more favorable than its rival, the *reductive tradition*. The latter is based on the (logical) derivation of the *explanandum* from true statements about more basic

¹⁵ See (Craver 2007, 154-160).

¹⁶ “Levels of mechanism” is the technical concept reflecting the premise in the mechanistic discourse that there is a nested hierarchy of mechanisms until they bottom out. I touch on Craver's explication of the concept (2007, Section 5.3) in a section to come.

¹⁷ It should also be able to distinguish explanations from “merely phenomenal models” (Craver 2006).

phenomena (i.e., inter-theoretic reduction—recall the deductive-nomological [covering law] model of explanation). Craver acknowledges that this derivability of the *explanandum* (the rigorous, law-involving exhibition that the occurrence of the phenomenon is *as expected*) bestows a “regulative ideal” upon explanations within the reductive tradition (2007, 161). This is to say, the rival tradition of constitutive explanation involves an inherent normative criterion by designating if the explanandum follows from laws or less strict generalizations.

One of the extensive discussions Craver engages in in *Explaining the Brain* is based on Hodgkin and Huxley's mathematical model of the action potential (hereafter, the HH model), by which he reviews the explanatory import of the derivability criterion for the practice in neuroscience.¹⁸ The mathematical model itself will be examined in the next chapter, but I will touch upon it very briefly here. And this requires the introduction of the phenomenon the HH model aimed to explain. So I now present the action potential, with its currently accepted explanation. This will be followed by Craver's evaluation of the HH model, which he interprets as showcasing the insufficiency of the regulative ideal of the reductive tradition.

2.2.1 The Need For a Regulative Ideal

Action potential is a phenomenon occurring in neurons and other excitable cells (muscle and secretory cells). It is an electrical signal that can be transmitted along the axon of the cell, from the cell body to the axon terminal. When the signal reaches the end of the axon, it causes neurotransmitters to be released, which bind to the receptors of the next neuron. Thus, it principally enables neuron-to-neuron communication. It is caused by inward and outward ionic currents through the cell membrane, resulting in a local change in the electrical potential, which in turn causes ionic movement in the adjacent area (hence the transmission of the signal along the axon). This is why the biophysicist Hodgkin states (in his Nobel lecture summarizing his own work on the action potential), “[it] is not just an electrical sign of the impulse but is the causal agent in propagation [of impulse].” (1963, 32)

In its resting state, the neuron cell membrane is highly polarized, due to differences in concentrations of positively and negatively charged ions and molecules on the two sides. During this state, basically, the extracellular environment has high concentration of positively charged

¹⁸ The discussion is meant to be extended to practice in life sciences in general, the domain mechanistic philosophy of science appeals to. Craver's 2006 and 2008 are his paper-length works devoted specifically to this discussion.

sodium ions (Na^+) while the intracellular environment has high concentration of negatively charged proteins (and also low concentration of chloride ions, Cl^-). The potassium (K^+) concentration inside the cell is also high compared to its concentration outside, but it is not sufficient to overcome the dominance of the negative charge due to the proteins. Therefore there is an electrical potential difference (voltage) across the membrane, and in this particular state it is called the *resting potential*. The resting potential for mammalian neuron cells is measured to be between -60 to -70 millivolts (mV).

This picture changes when the neuron is stimulated: The stimulus has the effect of *depolarizing* the membrane, that is, making the intracellular environment *less* negatively charged. If depolarization is strong enough, more and more extracellular sodium ions are allowed to enter the cell, rapidly making the interior *more positive* than the exterior (corresponding to the firing of the action potential spike). The membrane potential thus rises up to +35 to +40 mV. Adjustments in ion concentrations results in first a rapid drop in membrane potential (due to the disallowance for sodium inflow and the allowance for potassium outflow, it drops even below the resting potential, to approx. -90 mV) and then a comparatively slower reestablishment of the resting potential. The whole phenomenon takes less than one millisecond and its graphical representation has a typical shape as shown (approximately) in Figure 2-1.

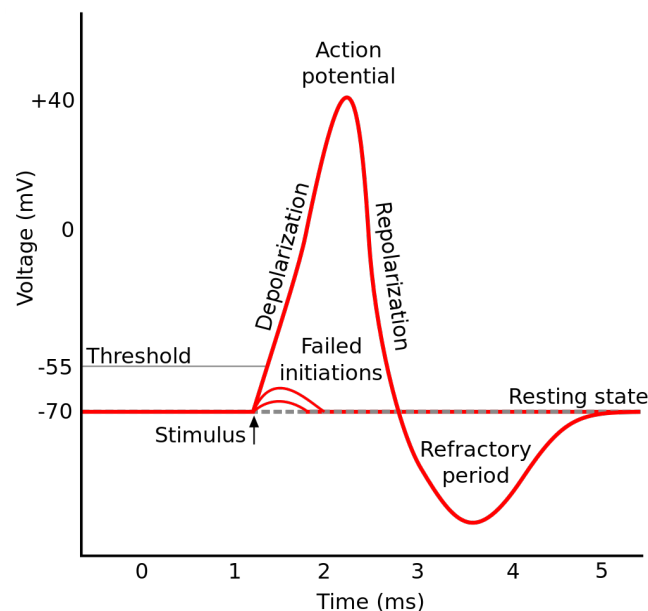


Figure 2-1. The action potential (wave)

Taken from Wikimedia Commons; licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license

The scientific work on the membrane's ability to generate the action potential spans many decades. The first important step was taken when Hodgkin and Katz established that an electrically excited membrane becomes “highly and specifically permeable to sodium” (1949, 38). The earlier theories had suggested that excitation would result in a temporary yet *total* breakdown of the membrane, making it permeable to *all* ions. But if this was the case, then the magnitude of the action potential could not exceed that of the resting potential. Besides, these theories could not account for the reversal of the sign of the voltage across the membrane (the change from minus to positive). With Hodgkin and Katz's work, the inward sodium current has been given the key role in the generation of the action potential. Few years later, Hodgkin and Huxley's mathematical modeling of the action potential provided very accurate quantitative descriptions of the voltage and time-dependent changes in the membrane permeability to sodium and potassium ions (i.e., sodium and potassium *conductances*¹⁹).

The consequent problem has been understanding how these changes take place in the membrane, and despite the great discoveries along the way, the explanation for the action potential is still not complete. Scientists now think that the sodium and potassium fluxes take place through selective voltage-gated ion channels. These channels are transmembrane proteins, whose structures and compositions are revealed to a great extent. They respond to changes in ligand concentration and transmembrane voltage by changing their conformations, letting the ions traverse the membrane down their respective electrochemical gradients. For instance, while the membrane is polarized (i.e., when it is in its unexcited, resting state) a helix-shaped segment of the sodium channel (the S4 segment) is kept in a fixed position. This is made possible by the formation of ion pairs between the positively charged amino acids of the helix and the negatively charged residues in the adjacent segments. Upon activation (i.e., initial depolarization of the membrane), the pull of the negative internal transmembrane electrical field on the helix is eliminated, therefore the latter rotates outward along a spiral path. This bends the ion-selective permeable pores (forming the other segments of the protein: the S5 and S6 segments), so they open up to allow the passage of the sodium ions exclusively (Catterall 2000, 16; Craver 2007, 119). Of course this only accounts for one phase of the generation of the action potential (the rising phase of the signal), which precedes the inactivation of the sodium channel, the activation and finally the inactivation of the potassium channel. But instead of talking about them, I just want to emphasize the point that it is not the whole membrane but only some parts of it which are responsible of the generation of the action potential

¹⁹ When the HH model is of concern, “permeability” and “conductance” can be used interchangeably. See Section 3.2 in the next chapter.

phenomenon—such as the transmembrane proteins making up the ion channels. The membrane has other parts, such as the myelin sheath, which are irrelevant to this phenomenon, but might be relevant to other phenomena (e.g., the myelin sheath is a relevant part for the *transmission* of the action potential).

The landmark work by Hodgkin and Huxley (the HH model published in 1952, for which they received the 1963 Nobel Prize in Physiology or Medicine) gives the mathematical formulation of the changes in the ionic currents across the giant axon membrane of the squid, with an admirable agreement with the experimental data. The model is derived from laws (or “fundamental equations”, cf., Bogen 2008) such as Ohm's Law, Coulomb's Law or Nernst's Equation. This is the basic reason why Craver treats the HH model as “a compelling example of explanation by deductive subsumption under laws of nature.” (2007, 52)²⁰ Yet, in his words, while “these equations [the HH model] characterize how specific ion conductances change with voltage and time . . . *they do not explain why* they change as they do, when they do.” (2007, 59 [emphasis added]) This is a point approved by Hodgkin and Huxley themselves (1952, 541), who were unaware of the voltage-gated ion channels back then. They speculated that some particles within the membrane might be responsible for the changes in ion conductances, but did not shy away from writing “the success of the equations is no evidence in favour of the permeability change that we tentatively had in mind when formulating them.” (*ibid.*) It is also acknowledged that although their model accorded well with the experimental results, alternative (mathematical) formulations of the phenomenon (with equal potential of generalizability) could be offered, too (*ibid.*, and Huxley 1963, 61).

At first glance, it appears that the HH model lacks a causal content.²¹ It allows making some empirical generalizations without giving an account for the obtaining of the causal relations. Craver thus sees it as a “merely phenomenological model” (2007, 59),²² or “[a]t most . . . a 'how possibly' sketch of the action potential.” (2007, 57) In his lexicon, both of these refer to the opposite of a good, complete explanation. But given that the mathematical expressions involved are derived from laws, the HH model satisfies the regulative ideal of the reductive tradition in scientific explanation. In this tradition, which is taken synonymously with Hempel's model of explanation in this context (mechanismic discourse), derivability (namely, “explanation by deductive subsumption under

²⁰ “There is perhaps no better example of covering-law explanation in all of biology than Hodgkin and Huxley's . . . model of the action potential.” (Craver 2008, 1022)

²¹ Weber (2008) and Levy (2013b) disagree with this. Their views are covered in the next chapter.

²² “Merely phenomenal model” in Craver 2006.

covering laws”) *is meant to be* non-causal.²³ Repudiation of causal explanation is an essential point in Hempel (1967): Although “informal, non-technical” explanatory texts might involve causal notions, their “explicit technical restatements” do away with these in favor of laws. Hempel writes:

Generally, it may be said that the technical vocabulary in which scientific hypotheses and theories are expressed does not contain such words as “cause” and “causal factor,” and that an account in terms of causes is acceptable as an explicit explanation in science *only as far as the causal attribution it makes can be substantiated and replaced by statements of corresponding laws*. (1967, 81 [emphasis added])

For Hempel, then, an explanation may *only in appearance* be causal; yet the true technical explanation of the phenomena is unquestionably lawful and *this* lawfulness discards the causal talk. It is also possible that there may be no superficial causal talk to replace, as, for instance, Hempel observes in the case of explanation of laws themselves. Laws like Galileo's and Kepler's or the optical laws of reflection do not involve causal statements but their explanation, too, can be shown in terms of logical-mathematical derivation from other laws (from the laws of mechanics, gravitation, electromagnetic-wave theory of light) (1967, 82). From Hempel's viewpoint, the equations devised by Hodgkin and Huxley would also be laws because they are expressions of regularities concerning the action potential. Accordingly, their derivation presented in the scientists' paper counts as the explanation of the action potential. In this respect, Craver is justified in saying that the HH model is a “compelling example” of Hempel's account of lawful explanation. (Probably Hempel would not disagree with this.)

According to Craver, the HH model's not explaining the action potential showcases the inadequacy of the regulative ideal of the deductive-nomological model of explanation (i.e., derivability). The latter attends to exhibiting how a phenomenon fits within the *nomie* nexus, by setting criteria for the deductions to be made from laws (or law-like generalizations). The HH model *does* meet these criteria. But, Craver thinks, making empirical generalizations about a phenomenon, if not coupled with revealing what actually underlies the success of them, does not yield a real understanding of the phenomenon. Hodgkin and Huxley themselves were aware of this: that is, despite not putting it this way, they were acknowledging that their model did not exhibit how the explanandum phenomenon fit into the *causal* nexus.²⁴ Hence Craver's observation is that

²³ Seen in larger context, the following discussion about Hempel's model of explanation can be read as an instance of the first tenet of the mechanistic discourse—that is, a mechanistic philosopher's (Craver) arguing against the centrality of use of laws in scientific explanation.

²⁴ The expression “fitting into a causal nexus” is quite famous in the literature on explanation and it is due to Wesley Salmon (e.g., 1984a, 9). Salmon's causal-mechanical model of explanation is built on this idea of showing how an event (the explanandum) fits into the causal patterns of the world, that is, the exhibition of the conditions due to which its

meeting the derivability criterion is not what makes a model explanatory, as the HH model illustrates (2008, 1032): When explanations in neuroscience and other life sciences are concerned, some other criteria are required so as to disqualify generalizations like the HH model from being regarded as good and complete.²⁵

Now that scientists know “that conductance changes across the membrane are effected by conformation changes in ion-specific channels spanning the cell membrane” (2007, 58), the relevant discoveries regarding the *molecular mechanisms* represent a progress in the explanation for the action potential (says Craver). This is to say, from Craver's viewpoint, the later models of this phenomenon *are* explanatory (while the HH model is not), because *they do describe mechanisms* by which scientists can knowingly control and manipulate phenomena. The particular comparison of the HH model with the more recent models will be a topic in the next chapter. The generalizing point to be made at this point, on the other hand, is that Craver's account of mechanistic explanation is meant to be “normatively rich” (2007, 61) so as to specify the criteria to compare different models. This can be read as the philosophical articulation of scientific progress in terms of how “complete” mechanistic models are.

To this end Craver first introduces the constraints on causal explanations of particular events and of the activities of the parts of wholes (in the case of constitutive explanation) (2007, 26 and 59-60). I do not need to reproduce the discussion on this matter; it should suffice here to say that the constraints are quite noncontroversial and they point to the need for a proper causal account of explanation that can discriminate explanatory causes (from mere correlates and irrelevant causes). Woodward's (2003) manipulationist theory of causation is of service here for Craver. (Woodward's is, in essence, a *difference-making* theory of causation: The discrimination of the causes from the other factors is based on the coordinated changes in the values of related variables.) A thing worthy of mentioning is that one of the constraints is “causes need not make effects probable to explain them.” (2007, 26) This is directly against the *nomic expectability thesis* emphasized in Hempel's model. As we saw in a quotation in the previous chapter, an explanation conforming to the deductive form of the covering law model exhibits that the phenomenon *was to be expected*, given the circumstances and the laws. But, Craver points out, neuroscience is rife with examples where what is expected does not happen or what looks improbable does happen. (Recall from the opening

occurrence is actualized (or, made more probable than its non-occurrence).

²⁵ That these criteria cannot be based on Friedman and Kitcher's unificationist models of explanation either is elaborated in Section 2.5 of his book (2007, 40-49). We can skip that discussion, though, as it only underlines the need for a causal account of explanation one more time.

section of this chapter Bogen's objection to the regularity requirement in MDC's characterization of mechanisms. I will return to this in Chapter 4.) The rarity of neurotransmitter release upon the generation of the action potential is an example to the first, and opening of ion channels despite unfavorable conditions is an example to the second. (2007, 39) Craver uses these facts to reiterate that what is important for a neuroscientist is how the phenomenon is causally produced whenever it is produced, without any regard to the expectability or the frequency of its production.

Having established these, he continues on to improve a version of the systems approach to constitutive explanation (i.e., Cummins's). The rationale for the “improvement” is explained by the lack of a regulative ideal in this approach, one that is comparable to the derivability criterion of the reductive tradition (2007, 161). The above mentioned constraints are required for qualifying the relevant causes of a phenomenon. On the other hand, in constitutive ME, a further criterion must be met so that the *non-causal* part-whole relationship of constitutive relevance is specified. This is the point where the mutual manipulability requirement is introduced to spell out the regulative ideal that is called for:

Relationships of mutual manipulability can and should replace the requirement of derivability as a regulative ideal on constitutive explanations in neuroscience. One need not be able to derive the phenomenon from a description of the mechanism. Rather, one needs to know how the phenomenon is situated within the causal structure of the world. That is, one needs to know how the phenomenon changes under a variety of interventions into the parts and how the parts change when one intervenes to change the phenomenon. (2007, 160)

A more elaborate characterization of the requirement is:

[T]o establish that X 's Φ -ing is relevant to S 's ψ -ing it is sufficient that one be able to manipulate S 's ψ -ing by intervening to change X 's Φ -ing (by stimulating or inhibiting) and that one be able to manipulate X 's Φ -ing by manipulating S 's ψ -ing. To establish that a component is irrelevant, it is sufficient to show that one cannot manipulate S 's ψ -ing by intervening to change X 's Φ -ing and that one cannot manipulate X 's Φ -ing by manipulating S 's ψ -ing. (2007, 159)

where “manipulating” and “intervening” should be understood according to the framework laid out in Woodward 2003. Mutual manipulability is thus proposed as the constraint put on an activity of the part of a whole, for it to qualify as a component in the mechanism for a certain behavior of the whole (the explanandum phenomenon).

I think this much suffice to lay down the general framework of Craver's take on mechanisms. I begin the next section with articulating his notation, and move on with a conceptual refinement that aims to elucidate the *mechanismic* construal of mechanistic explanation.

2.2.2 Constitutive Mechanistic Explanation and Levels of Mechanisms

The notation used by Craver throughout his book is as follows: when a mechanism is described as “entities and activities organized such that they exhibit the explanandum phenomenon” (2007, 6),

S stands for a *mechanism* as a whole, and *ψ* is the *behavior* *S* is engaged in;

X is any of the component *entities*, and *Φ* is any of the component *activities*, whose organization make up the mechanism;

*X*₁*Φ*₁, *X*₂*Φ*₂, *X*₃*Φ*₃, ... are certain individual *components* in the mechanism, each of which is (constitutively) relevant to the explanandum phenomenon. (2007, 7 and 189)

The explanandum phenomenon is addressed as “*S*’s *ψ*-ing”. This is “the property or the behavior explained by the mechanism” (2007, 7). Similarly, the components “*X*₁’s *Φ*₁-ing”, “*X*₂’s *Φ*₂-ing”, etc. are also phenomena, each of which can—at least in principle—be mechanistically explained by the components at a lower level. There is thus the hierarchy forming the “levels of mechanism”. According to Craver, this hierarchy reflects a peculiar sort of composition relation.

Levels of mechanism are levels of composition, but the composition relation is not, at base, spatial or material. In levels of mechanisms, the relata are behaving mechanisms at higher levels and their components at lower levels. These relata are properly conceived neither as entities nor as activities; rather they should be understood as acting entities. The interlevel relationship is as follows: *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 in the mechanism for *S*’s *ψ*-ing. Lower-level components are *organized together* to form higher-level components. . . . Organization is the interlevel relation between a mechanism as a whole and its components. Lower-level components are made up into higher-level components by organizing them spatially, temporally, and actively into something greater than a mere sum of the parts. (2007, 188-189 [original emphasis])

In order to enhance visualization Craver provides a figure (Fig. 5.7, p. 189). The one I present below is based on this and another one provided at the beginning of the book (Fig. 1.1, p. 7). The figure shows three levels of mechanisms where arrows show causal relations.

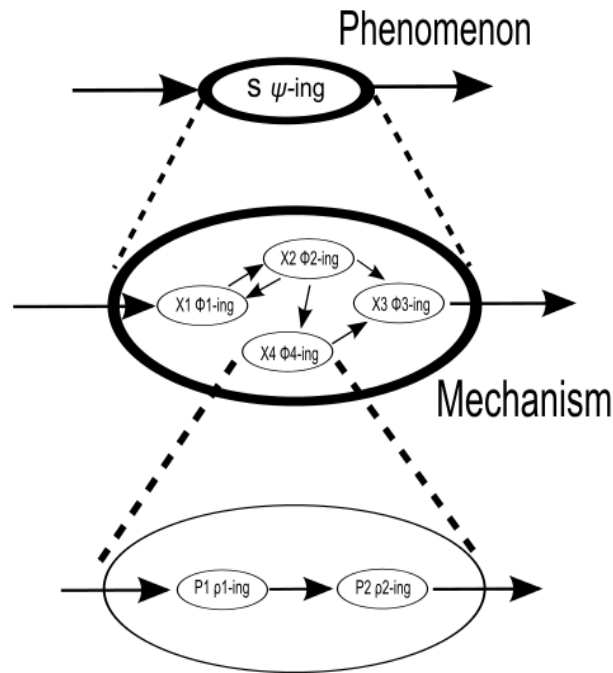


Figure 2-2. Levels of mechanisms and interlevel relations. (Adapted from Craver [2007, 7 and 189]). For both simplicity and argumentative consistence, I made changes regarding causal relations (shown with arrows). I also ignored some components (at the lowest mechanistic level) and made the contours of two ellipses thicker (to make them stand out). Craver labels the ellipses “phenomenon” and “mechanism” as above in the identical Figs. 1.1 and 4.6 (2007, 7 and 121).

I want to be maximally precise at this point, for correct visualization lies at the core of the forthcoming discussion. The dashed lines connecting the contours depict a supervenience relation obtaining between the higher-level components and the behavior associated with the organization of the lower-level components (i.e., the mechanism for the behavior in question). We find the somewhat concealed evidences for this in small mentionings—for instance, in a footnote where Craver's point is to state that constitutive relevance should not be confused with supervenience²⁶:

Supervenience . . . amounts roughly to the claim that there can be no difference in *S*'s *ψ*-ing without a difference in the mechanism for *S*'s *ψ*-ing. Supervenience so stated is a relation between a phenomenon and the *corporate behavior of the organized components*. The relevance relation, in contrast, holds between the phenomenon and *one of the components*. The supervenience claim, note, is not symmetrical [in contrast to constitutive relevance]. (2007, 153, fn. 33 [emphases added])

²⁶ See also (2007, 213).

Now let us be clear with the ellipses: each depicts an “acting entity” (see the above quotation)—a component, so to say. As such, they are causally related with one another. As Craver remarks at various points (see e.g., pp. 122, 136 and 138), it is in virtue of their spatial, temporal and *active organization* that the explanandum phenomenon is exhibited. For an illustration let us recall the explanation for the depolarization of neuron membrane (which leads to the generation of the action potential). This involves, among other components, the outward movement of the helix-shaped S4 segment of the voltage-gated sodium channel. We can think this as the $X_1\Phi_1$ ellipse in the above figure, where X_1 stands for the component entity (the S4 segment) and Φ_1 stands for the component activity of the outward movement. It is X_1 's Φ_1 -ing that causes the ion-selective pores (X_2) to be bent (Φ_2): $X_2\Phi_2$, so to say, is another component relevant for the depolarization behavior of the membrane. Similarly, we can think $X_3\Phi_3$ as sodium ions' rushing into the cell (a component entity [here, X_3] need not be singular). For the sake of simplicity we can assume that these are all of the components relevant for the explanandum phenomenon. Then the figure below illustrates the constitutive relationship between these components and the phenomenon.

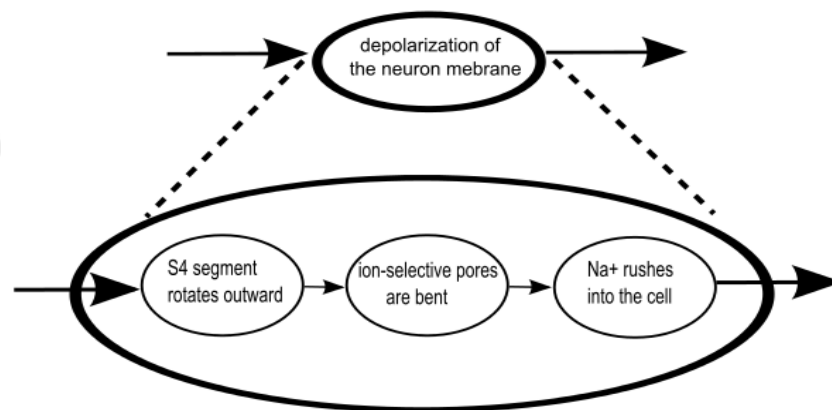


Figure 2-3. Levels of mechanism in the explanation for the depolarization behavior of a neuron membrane.

On Craver's account, the explanation of membrane depolarization is *complete* when all and only the components relevant to the explanandum phenomenon are *fully* described (2007, 111-112). This means, being able to illustrate the constitutive relationship is not the end of the pursuit for mechanistic explanation. One must also be able to see which changes in individual components bring about which changes in the overall behaviour, and vice versa:

That is, one needs to know how the phenomenon changes under a variety of interventions into the parts and how the parts change when one intervenes to change the phenomenon. When one

possesses explanations of this sort, one is in a position to make predictions about how *the system* will behave under a variety of conditions. Furthermore when one possesses explanations of this sort, one knows how to intervene into the mechanism in order to produce regular changes in the phenomenon. (2007, 160 [emphasis added])

As indicated before in this chapter, in developing a model of mechanistic explanation in neuroscience, Craver deliberately contributes to the systems tradition of constitutive explanation. This is why he invokes the notion of *system* (italicized in the quotation above) in relation with the explanandum phenomenon. Hence his preference for the letter “*S*” in his notation for the higher-level object of concern (as in “*S*’s ψ -ing”), which—as he acknowledges—is due to the notation in Cummins’s functionalist account of explanation (2007, 110). Let us remember that “*S*” also stands for a mechanism in Craver’s account (as mentioned in the opening of this section).²⁷ We could grant that a mechanism *is* a system made up of organized components, and find no problem in this dual usage of “*S*”. However we also see Craver using “*S*” for a neuron, as read in this quotation: “The neuron (*S*’s) ability to generate action potentials (ψ) is realized by the fact that the neuron (*S*) has the property (Φ) of being an organized collection of Na⁺ channels, K⁺ channels, a membrane, and so on.” (2007, 213; see also the usage on p. 219) In this latter usage, “*S*” designates an object *as a whole* (i.e., its being made up of parts is underlined). Possibly, then, this very whole and the system/mechanism are related. Given Craver’s spontaneous switch from one usage of “*S*” to the other,²⁸ we might think that they are *very closely* related. As a matter of fact, he does not seem to be concerned by the switch and I think this is quite telling about the essential confusion on the part of the mechanistic philosophers (i.e., confusing machine mechanisms and causal mechanisms). Yet let us pursue this idea for a while—the idea that the whole and the system/mechanism are very closely related. This we shall do for the refinement of the key notions of the mechanistic discourse, and for clearly exposing the metaphysical picture a mechanistic philosopher has in mind.

2.2.3 The Categorical Distinction Thesis

We must be clear with what is being discussed, and clarity begins from notational disambiguation. Therefore I first introduce “*W*” to stand for the object as a whole (whose behavior is investigated) and keep on using “*M*” for the mechanism. An example for a *W* is the neuron

²⁷ See e.g., pp. 7, 145, 216, 219 where Craver is clear that “*S*” designates “mechanism as a whole”. See also pp. 189, 212.

²⁸ The shift is most noticeable when it happens within two consecutive pages (pp. 212-213).

membrane, and if we stick with our last example where the behavior of concern (W 's ψ -ing) is the membrane's depolarization, M is the mechanism by which the membrane depolarizes. The noted switch is thus between W and M : We are given that " S 's ψ -ing" supervenes on M where S is read sometimes as W and sometimes as M . How should we make sense of this? I shall first consider two options regarding the close relation between W and M , so as to ensure that *none* of them could be what Craver has in mind. This is also going to be helpful in demonstrating that W and M are categorically distinct.²⁹

The first option to be considered is taking W and M simply as the same thing. This would entail that " S " always designates a single thing and there is actually no real switch in Craver's writing. (S 's ψ -ing supervening on M would then be an uncontroversial state of affairs.) Yet, here are two reasons why W cannot be M (i.e., why " S " cannot be designating a single thing):

1. It is because M is different than W that scientists are said to be investigating mechanisms. Otherwise "the discovery and description of mechanisms"³⁰ makes no sense: You just *have* the mechanism as you *have* the object as a whole; this whole *itself* explains its own behavior (meaning that the neuron membrane *is* the explanation for its depolarization).

This is obviously unacceptable. Not all parts of the whole are relevant to the mechanism of a behavior. Craver acknowledges this, and this is why he distinguishes a "part" from a "component". More importantly, this is his motivation for developing an account of constitutive relevance, which lies at the core of the model of mechanistic explanation. Here is what he writes at the beginning of the important section on constitutive relevance:

I have not provided an account of constitutive explanatory relevance [yet]. That is, I have not said when a part of S [read: W] is a component in the mechanism [M] of S 's ψ -ing. Not all parts are components. Consider . . . the difference between mechanisms and machines. Machines contain many parts that are not in any mechanism. The hubcaps, mud-flaps, 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. (2007, 140 [original emphasis])

As the quotation shows, there is no doubt that Craver does not equate an object as a whole and the mechanism explaining a certain behavior it exhibits. (I will return to this quotation to draw a

²⁹ Kaiser and Krickel (*forthcoming*) offer an analysis that make similar points with a different concern, which is understanding the ontological nature of the explanandum phenomenon (they analyze the constituted phenomenon as an "object-involving occurrent"). I thank Beate Krickel for letting me know of their perspective, and for her constructive criticism regarding this chapter of the dissertation.

³⁰ I take this phrase from the opening paragraph of MDC's paper (2000, 2). (The full quotation is found in the previous chapter.) Since "discovery" is an important aspect of the scientific practice for a mechanistic philosopher, similar phrases are ubiquitous in Craver's book, too.

comparison between mechanisms and machines. In that context I will refer back to the observation that there is a semantic ambivalence inherent in the usage of “mechanism” [the notion].)

2. There are parts of W which are irrelevant to its ψ -ing and this is the first reason why W is not M . The other reason can be thought as the flipside of this picture: There are cases where the component entities in M are not parts of W . I have been citing one such case—the example of neuron membrane depolarization: Two of the parts of W that are relevant for the explanation are the S4 segment (X_1) and the ion-selective pores (X_2). There is a straightforward composition (part-whole) relationship between X_1 or X_2 and W . There is also the aggregate of the rushing sodium ions that we mention in the explanation, which is the component entity we labeled as X_3 above. The immediately apparent fact is, X_3 is *not* a part of W : the membrane-traversing sodium ions are in no sense a part of the neuron membrane. But all of X_1 , X_2 and X_3 , on Craver's account, are component entities in the *mechanism* M of the explanandum (the depolarization of W). With the respective activities Φ_1 , Φ_2 and Φ_3 (described above) they make up the components in M .

So again we see that W and M are clearly not the same thing. For they are not made up of the same parts. Surely M itself is also a whole (and this is the reason why Craver often speaks of “mechanism as a whole”), for being made up of parts—the causally organized components. But in the final analysis, *it is certainly a different whole*: not W , but, say, W' .

These suffice to conclude that W and M must be distinct. There is one thing I want to note about this *W-M distinction thesis*: While it is *always* the case that not all parts of W are component entities in M , we may come across cases where *all* component entities in M are *also* parts of W . (The moka pot example below provides one such case.) However this must not evoke the impression that the distinction thesis is about intersection sets and formal subset-superset relations between things making up W and M . The real point I want to stress is that, the *W-M relation* that we are exploring is not about the shared or excluded parts (*constituents-talk*). After all, there is no general principled way to fix the conditions of when and which of the M -entities (i.e., component entities: X_1 , X_2 , ...) are among the W -entities (parts making up W). Its absence should motivate us to think that the *W-M relation* cannot be completely covered by the constituents-talk (although the latter has been convenient in the *W-M distinction thesis*). Below I will say more against constituents-talk, with the intention to explore the metaphysical underpinnings of mechanistic explanation.

Now we can consider the second option: W and M could be distinct but still be counted as closely related, just in case they engage in the identical behavior (ψ). There is then both W 's ψ -ing and M 's ψ -ing, and arguably, the switch in Craver's use of “ S ” could be explained by a confusion with

respect to whether W or M is engaged in ψ -ing. One might question the relation between W 's ψ -ing and M 's ψ -ing; she might think that they are identical or she might think that they are different but one *realizes* the other, etc. It would be better for her not to bother with such line of inquiry, though! For the real problem with this option is, given the W - M distinction thesis and the definition of a mechanism, " M 's ψ -ing" is devoid of meaning. One should not forget that "mechanisms are always mechanisms *of* a given phenomenon" (Craver 2007, 123 [original emphasis]). This has two interrelated implications (the second entails the first): There can be no *bare* (causal³¹) mechanism which is not associated with any behavior, and a (causal) mechanism is always associated with the investigated phenomenon.

The inseparability of the mechanism and the phenomenon is implied by definition, by M 's *being* the organized X 's and Φ 's *such that* ψ is exhibited. Recall neuron membrane's depolarization: It is *not* that there is the outward rotation of the S4 segment ($X_1\Phi_1$), the bending of ion-selective pores ($X_2\Phi_2$) and the rushing in of Na^+ ions ($X_3\Phi_3$) happening independently of each other. It is *not* that

(i-) M is the set of independent components $C = \{X_1\Phi_1, X_2\Phi_2, X_3\Phi_3\}$;

and

(ii-) if the elements of this set are organized in a certain way *then* there is ψ -ing.

It is *rather* that

(i*-) the components are already causally organized and this organization itself is M (so M is not C);

and

(ii*-) there is no further condition to meet for there to be ψ -ing, once this organization of components, that is, M is instantiated.

Why so? Because what makes the elements of C the *components* in M is their being causally organized to constitute W 's ψ -ing. Any of the occurrences mentioned above enter into numerous causal relations, besides the ones obtaining between each other. The bending of the pores, for instance, is causally related with the influx of the Na^+ ions *and* (say) the generation of a stress on the membrane, too. There is a good reason why we privilege the former but not the latter—that is, why we count the former as a component while omitting the latter (or many other occurrences). The reason is, the influx of the ions is found causally more important (i.e., causally relevant) to the phenomenon than the generation of the stress; so the elements of C are already picked with the

³¹ We take M to be the mechanism which explains a phenomenon, so it cannot be a *machine* mechanism.

same explanatory purpose. Moreover these occurrences are not only picked individually with a concern for the explanandum phenomenon but also grasped *contextually*, that is, according to the organization of relations they enter with each other—which is, in turn, *nothing but* the mechanism for *W*'s ψ -ing. Any one element in *C* can—in principle—be a component of more than one mechanism, but concerning its relations with the other components picked contextually, they can all produce only one type of phenomenon. This is to say, you cannot get a different phenomenon with the different configuration of the same set components—if they are known to be the components in a certain mechanism, then they are organized within a single configuration.³²

So there is not *M*'s ψ -ing but there is *W*'s ψ -ing: *W* and *M* are not engaged in the same behavior, because *M* itself cannot *engage in* any behavior. No behavior can be *ascribed* to it, but only one behavior is *associated* with it—and that behavior is nothing other than *W*'s ψ -ing.³³ As a consequence, *W*, which is an object (or a group of objects), and *M*, which is an organized set of occurrences, are categorically different. Hence, one makes a *category mistake* if he treats *M* as the sort of whole *W* is. (This is why our motivation to consider the *W-M* distinction in terms of something other than constituents-talk turns out right.) Here below, then, is what I find to be a convenient way to formulate this distinction. Note the introduction of the new symbol **w** that stands for any whole that is made up of parts (not only objects):

The Whole-Mechanism Categorical Distinction Thesis (CD Thesis): It is possible to think of a whole **w** with a constituent (one of its parts) that is *irrelevant* to a behavior associated with **w** (possibly, a behavior *exhibited by w*), but it is impossible to think of a mechanism *M* with a constituent (one of its components) that is *irrelevant* to *the* behavior associated with *M* (this behavior is of course the explanandum phenomenon).

This distinction is important, because the notion of a “whole”, if understood as anything made up of parts, may be taken to refer to a mechanism as well (the “mechanism as a whole”). I needed to

³² Perhaps this is a right spot to remind the reader that within the present discussion, I am staying faithful to the way mechanistic philosophers themselves construe (causal) mechanisms. In line with the “intentional mistake” acknowledged on page 40, I write as if mechanisms are harbored within the real causal structure, i.e., as if the mechanism components are the select links on the investigated causal chains. I endorse this premise for the sake of the argument.

³³ “Being associated with” is an admittedly ambiguous notion. Yet it should work fine in the discussions to follow, where the point is sorting *W* and *M* in terms of the ontological categories they belong to.

As I will briefly mention below, Soom (2012) asserts that *M* itself and *W*'s ψ -ing are strictly identical (at the token-level). If I was to draw on Soom, I could continue this discussion by replacing “associated” with “identical”. But this would require an additional argument that would derail us at this point. Moreover, I have reasons for not adopting Soom's view anyway, that is, for *not* holding that *M* is identical with *W*'s ψ -ing. This is related with my being a mechanistic *non*-realist (to be elaborated in chapter 3).

delegitimize this use of “whole” for the mechanism in the context of this discussion, to be able to identify it only with the thing whose behavior is explained by a mechanism. The formulation above is meant to serve this purpose, by downplaying the mechanism's being made up of parts (the components).

This discussion leads us to the following. To begin with, we see that Craver's *S* (of “*S*'s ψ -ing”) must be a whole rather than a mechanism, because a mechanism cannot be ψ -ing. We can come across many cases where it is possible and legitimate for *S* to stand for an *object* as a whole (our *W*) —e.g., a heart (pumping blood), a neuron (releasing neurotransmitters), a neuron membrane (depolarizing), etc. But it is *safer* to talk in terms of wholes rather than objects, due to the ontological baggage associated with the latter notion: We should be able to regard, say, a *gene* (that is e.g., being transcribed, or expressed [ψ -ing]) as the *S* of concern, without worrying about the gene's ontological status (whether it is an object or not). So far, interpreting Craver's *S* has been fruitful; my suggestion now would be to briefly consider the *system* the behavior is exhibited in. By “system”, I want to invoke not only the investigated whole *S* but also its spatiotemporal vicinity and its causal neighborhood wherein *S*'s ψ -ing is situated. It can be troublesome to clarify or disambiguate what determines the boundaries of the system, but for present purposes it suffices to accept only that there is *more* in the system than in *S*. Part of this “more” is of course the component entities of the mechanism explaining *S*'s ψ -ing which are not parts of *S*, such as the sodium ions in the depolarization example (which are not parts of the neuron membrane), or the water and the ground coffee in the moka pot example below (which are not parts of the moka pot). (Another part of the “more” is the set-up and background conditions required for the phenomenon.)

There are two things to note in connection with the notion of “system”. First, the system is a *causally active* organization of objects/wholes manifesting behaviors; the occurrences within the system which are found relevant to *S*'s ψ -ing make up the mechanism components. Therefore, on the mechanistic discourse that I am elucidating, the mechanism for the investigated behavior is harbored within this system (more on this in the last section). Second, I began the discussion (at the end of Section 2.2.2) by stating that Craver uses the notion of “system” as having *some* relation with the explanandum phenomenon, and I wrote that his preference for the letter “*S*” follows from being affiliated with the systems tradition of constitutive explanation. The whole discussion that followed this was motivated by the need to clarify whether this *S* is a whole or a mechanism. It might be found perplexing that *S* is both a whole *and* the label due to the system this whole's investigated behavior is exhibited within. In other words, *S* is meant (by Craver) to stand for “system”, but it does not, because the system of concern is something larger than the *S* I have been

writing on. I think it is natural for someone to get perplexed at this point, but it seems the problem is solely due to the way Craver writes on mechanisms and mechanistic explanation. Now that we have gained an understanding of the concepts, though, the labels should not matter. All we need to be clear about is the following formulation: A whole exhibits a behavior within a system; what explains how this behavior is exhibited (i.e., how the explanandum phenomenon is produced) is a mechanism that is identified by this very behavior.

These points are all made from the perspective of a mechanistic philosopher, for the sake of clarifying the notions on behalf of him, as much as his analysis of mechanistic explanation requires. Now, let us ask this: Why does Craver himself not distinguish the investigated whole from the mechanism explaining its behavior?

2.2.4 Machine Mechanisms and Causal Mechanisms Re-Visited

In fact, at first glance, Craver seems to distinguish the two. He does this when he states why *machines* are not mechanisms: they are not, because they have irrelevant parts (irrelevant to some mechanisms). In view of the fact that machines are objects as wholes, this sounds like a formulation of the *W-M* distinction thesis—but it is the wrong formulation. For, as indicated in the discussion above, the real distinction is the one formulated by the CD Thesis, that is, the distinction drawn according to the behaviors associated with wholes and mechanisms. Let us now see the efficacy of this thesis.

I begin by re-quoting from Craver:

Machines contain many parts that are not in any mechanism. The hubcaps, mud-flaps, 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. (2007, 140 [original emphasis])

It is definitely far-fetched to argue that there are “*many* parts [of a machine] that are not in *any* mechanism”. Hubcaps may be an example in support of this claim, but it is not difficult to think of mechanisms which involve the mud-flaps or the windshield. Considering the next part of the second sentence, though, I speculate that Craver wants to think the machines in terms of the *major* function(s) associated with them (e.g., the moving of an automobile).³⁴ I can grant him this much: It is true that some functions (behaviors) of a machine are more important than others, especially if

³⁴ “Function” and “behavior” (of a whole) can be used interchangeably in this context.

the latter are associated rather with the parts but not the whole. For instance, the moving of an automobile is more important than (its) shielding the passengers from the wind, or than (its) reducing the drag. Both of the two minor behaviors are actually behaviors of one part of the whole (i.e., the windshield), or, at least, it seems to make more sense to regard them as such (instead of taking them as the behaviors of the whole). Then, most parts of machines (with the exception being the decorative parts) can be found to be parts in some mechanisms; yet Craver is not concerned with most of *these* mechanisms, he is concerned with the *major* mechanism(s) responsible for the *major* behavior(s) of the machine. It is with this rationale that he appears to distinguish between an object as a whole and a mechanism.

But does this amount to the *categorical* distinction I elaborated above? No, it does not. For this method of distinguishing a whole and a mechanism is still constituents-talk: it is a comparison drawn between *sets* of objects that are relevant and irrelevant to the exhibition of a certain behavior. It excludes the *causal organization* of components (which are not objects but occurrences) inherent in the definition of the mechanism that is not graspable by constituents-talk. An explanatory mechanism, on the other hand, is neither the set of component entities nor the set of components (the set *C* in the discussion above), because neither of the sets explains the explanandum phenomenon by itself.

Let us recall the *semantic ambivalence* I have pointed out several times so far: There are *causal mechanisms* (the referents of the Causal-M sense) and there are *machine mechanisms* (the referents of the Machine-M sense); “mechanism” refers to either of them, depending on the context. Following Nicholson (2012) I have indicated that the mechanistic philosophers' use of the notion ignores this distinction. The result of this is their treating causal mechanisms as though they are machine mechanisms. Craver is a case in point, so I shall substantiate these claims by the help of concepts emerged in discussing his account. This will also work as the explanation for Craver's mistake of not categorically distinguishing *W* and *M*.

Roughly put, the Machine-M sense is the one that can be captured by *constituents*-talk, and, as can be recalled from above, the Causal-M sense is the one that is usually meant in the scientific *mechanism*-talk. I claim that *when artifacts are considered* distinguishing these two senses may not have much philosophical importance. But since natural phenomena are unlike artifacts, it matters a lot to keep the Machine-M and the Causal-M senses apart when we (the philosophers) want to analyze the mechanism-talk in *science*.

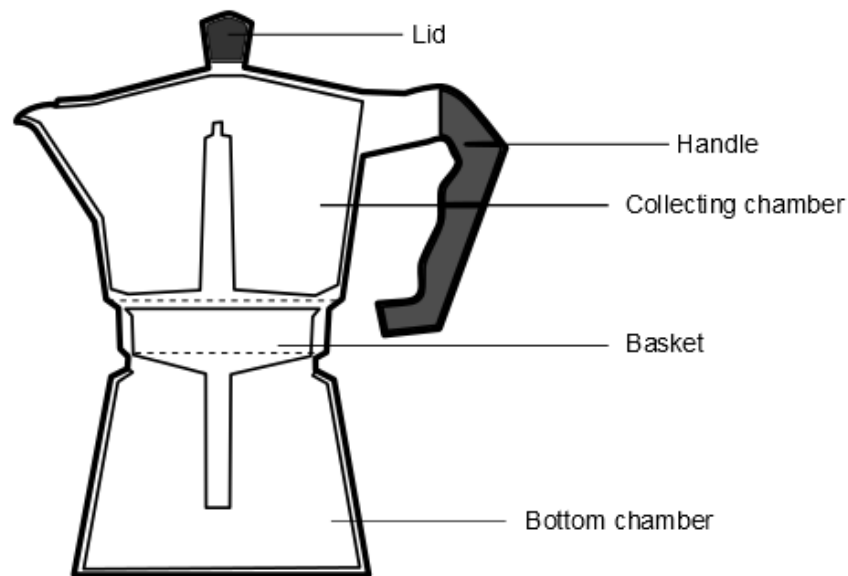


Figure 2-4. Moka pot

Adapted from the image uploaded to Wikimedia Commons (by the user platonides, on 11 June 2006). Licensed under Creative Commons Attribution – Share Alike 2.5 Generic License.

Consider a moka pot, which is a simple artifact designed to make strong coffee (Fig. 2-4). The pot is made up of five parts; three of them are directly involved in the making of coffee, so I shall label them as *relevant parts*. These are the bottom chamber, the basket and the collecting chamber. The two chambers are tightly screwed and the basket is rigidly held between these. The water is put in the bottom chamber, but some space for air is also left above it. When the whole pot is placed on a heat source (such as a stove) steam pressure builds up in this chamber, above the water surface. (This is because of vaporization and increase in the temperature of the air left.) As sufficient pressure is attained, water is pushed upwards through the funnel-shaped basket. Since this basket contains finely-ground coffee which dissolves in hot water, the *coffee drink* is collected in the collecting chamber above the basket, until the bottom chamber is depleted of water.

The moka pot could be seen as exemplifying the Machine-M sense, but I would rather say it is a machine simpliciter and not a mechanism *in any sense*. This should not confuse anyone who is already familiar with Craver's distinction between a machine and a mechanism. The moka pot is a machine because it has two parts irrelevant to the major function/behavior: the lid and the handle have nothing to do with the making of coffee. A good question is due: Does this come to mean that, if we remove these irrelevant parts we get a mechanism instead of a machine? I answer: *Yes* in one sense (Machine-M) and *no* in another sense (Causal-M), and herein lies the heart of the matter.

A machine mechanism is the assembly of the parts (of the whole of concern) that are relevant to

the major function, *and nothing else*. Therefore it excludes the *active* organization and this is why I argue that the Machine-M sense can be captured by constituents-talk alone. Now, when Craver writes of the “mechanism” that “makes” the automobile run, he must have the Machine-M sense in mind. Let us not miss the connotation of “make” here (the wording is due to Craver): The machine mechanism does not *explain* but *enable* the automobile's motion, so it is *separate* from the latter—just as the moka pot's basket and the chambers are separate from the making of coffee. Assemble them together but do not put in water (say, you have just cleaned and dried the moka pot and about to put it back next to the other kitchenware): the machine mechanism is still there (organized but not active), but you do not have the function. Notice that regardless of heating the pot or not, the machine mechanism is still the same mechanism. Because the Machine-M sense has a *material* basis (i.e., the material part-whole relationship) that is not affected by set-up conditions other than how the parts are assembled (organized). Further, it is because of this basis that even when the irrelevant parts (such as the lid and the handle) are removed, we can still say we are holding a “mechanism” in hand: a machine mechanism. “Mechanism” in this sense is a *whole* that can behave in distinguished and known ways under specific conditions.

To avoid confusion, let me clarify this: We speak of wholes whenever there are assembled parts. In addition, we can further speak of a *machine mechanism* if and only if there is a whole *minimally sufficient* to enable sensible and analyzable behaviors to be exhibited in its entirety, when specific conditions are met.³⁵ In view of this, both the moka pot machine (including the lid and the handle) and the mechanism that enables the making of coffee are wholes. But since the complete machine is not minimally sufficient for any sensible behavior in its entirety (the lid and the handle are always excess), only the latter is a machine mechanism. Note that being minimally sufficient for one behavior is a necessary and sufficient condition for a whole to be counted as a machine mechanism, but there is no necessity for the whole to be minimally sufficient for the *other* behaviors it exhibits. It is quite possible that there may be a whole which is also a machine mechanism, and which is not minimally sufficient for *all* of the behaviors it exhibits. I will come to this important point in my re-assessment of the CD Thesis below.

In contrast to a machine mechanism, a causal mechanism is characterized by being the organization of *active* parts. Hence we *never* get *any* causal mechanism just by leaving out the irrelevant parts of a machine. For firstly, some set-up and background conditions must be met so as to ensure that parts making up *a whole* are *active*; and secondly, additional entities which are not

³⁵ Although I am writing with *artifacts* in mind, this also applies to natural objects. The DNA and the heart examples given at the beginning of the chapter can be recalled.

originally parts of the machine (or, generally speaking, the whole of concern) might also be needed for the whole to exhibit a behavior. The “activity” of parts means, their entering in new causal relations that produce a phenomenon (when the necessary conditions are met). The phenomenon is the behavior of the whole the parts compose; the whole can also be considered as active when it engages in this behavior. The behavior may or may not be an intended *function* of a *machine* mechanism, but it is a behavior (an occurrence, categorically speaking) anyway whose explanation consists in one *causal* mechanism. Examples are due.

Say, you place the moka pot on the stove without the water and the ground coffee. Then the heat is conducted from the bottom chamber to the collecting chamber. So a whole W is active due to being heated, its behavior being the conduction of heat (W 's ψ -ing). Or, say, the moka pot is heated with water in it, but again without the coffee. Then a whole $\overline{W}$ can be said to be $\overline{\psi}$ -ing (transferring hot water to the collecting chamber). It can be questioned whether W and $\overline{W}$ are the same or different, whether they are the whole pot or not, etc. They may or they may not, it is up to us to decide, as long as they are kept as *names* of wholes (taken categorically)—and this is nothing but constituents-talk.³⁶

There might also be a concern regarding the status of water in $\overline{\psi}$ -ing. That is, we might ask if water is a constituent (part) of $\overline{W}$ or not. But again it is a matter of preference with regards to how much of $\overline{W}$ we want to separate from set-up and background conditions while talking about (or explaining) its $\overline{\psi}$ -ing. On the other hand, it is *not* up to *us* to decide if water is a constituent (component entity) in the *causal* mechanism $\overline{CM}$ of $\overline{\psi}$ -ing, for it strictly is: we do not have (or, cannot explain) the $\overline{\psi}$ -ing without the waters being heated in the bottom chamber (a component activity) and its being pushed upwards through the basket into the collecting chamber (another component activity). Nor is it up to us to decide if the respective *machine* mechanism (call it $\overline{mM}$) involves water as a part: No, water is strictly not a part of this machine mechanism. For p is not a part of a particular machine mechanism if it is not appropriated for a behavior (from the designer's point of view). But the two chambers and the basket are both component entities in $\overline{CM}$ and parts in the said machine mechanism.

Similar considerations apply to the original example, the making of coffee in the moka pot. What

³⁶ If I am to be consistent with the discussions above, recalling the neuron membrane–depolarization example and the CD Thesis, I should take W , $\overline{W}$ and the pot as the same thing, allowing that a whole can have parts irrelevant to the particular behavior in question. Example: the basket is an irrelevant part to the pot's conducting heat, because heat is conducted through the chambers that are in direct contact. Alternatively, I could take the whole to be the organization of parts minimally sufficient for the behavior. Continuing on the example, this would mean, W is the assembled two chambers and $\overline{W}$ is W plus the basket. Then W and $\overline{W}$ would be the names for the two different *machine mechanisms*. (They cannot be the same machine mechanism, because $\overline{W}$ has an extra part.)

is stated for water's status in $\overline{W}$'s $\overline{\psi}$ -ing can be repeated for ground coffee—without forgetting that the behavior in question now is the making of coffee in the pot (say, ψ^*), which is associated with a different causal mechanism (say, CM^*).

This merits attention: With the presence of ground coffee, there is no more $\overline{CM}$ but there is CM^* . However the *machine* mechanism is the same in transferring hot water ($\overline{\psi}$) and making of coffee (ψ^*). Therefore we observe that *the same machine mechanism can enable different behaviors to be exhibited*. Remarking the necessity that different behaviors are associated with different *causal* mechanisms, I hold that this clearly delineates the two senses of “mechanism”. The table below shows if an entity is a constituent of a (minimally sufficient) machine or causal mechanism for the three behaviors I have been considering.

Table 2-1. Behaviors and relevancies of objects to mechanisms associated with them

	Conduction of heat (ψ)		Collecting hot water ($\overline{\psi}$)		Making of coffee (ψ^*)	
	Machine-mechanism $\underline{mM}$	Causal-mechanism $\underline{CM}$	Machine-mechanism $\underline{mM}$	Causal-mechanism $\underline{CM}$	Machine-mechanism $\underline{mM}$	Causal-mechanism $\underline{CM}^*$
Any of the chambers	✓	✓	✓	✓	✓	✓
Basket	-	-	✓	✓	✓	✓
Water	-	-	-	✓	-	✓
Coffee	-	-	-	-	-	✓

We have thus disambiguated “mechanism”, so let us see its implication on my CD Thesis, which was a formulation of the categorical distinction between a whole and a mechanism. It must now be obvious that the categorical distinction is between being a whole and being a *causal* mechanism. Let us read the thesis with this proviso inserted:

The Whole-Mechanism Categorical Distinction Thesis (CD Thesis): It is possible to think of a whole $\mathbf{w}$ with a constituent (one of its parts) that is *irrelevant* to a behavior associated with $\mathbf{w}$ (possibly, a behavior *exhibited by* $\mathbf{w}$), but it is impossible to think of a (*causal*) mechanism M with a constituent (one of its components) that is *irrelevant*

to *the* behavior associated with *M* (this behavior is of course the explanandum phenomenon).

Why *M* cannot be *machine* mechanism should not go without explanation, though, because there is a point that requires heavy underlining here. The now-obvious reason for *M*'s being a causal mechanism is, a machine mechanism's being a whole: there would not be a categorical distinction between *W* and *M* if "mechanism" was understood in the Machine-M sense. Yet it is the following *much less noticable reason* that concerns us more in the CD Thesis: it is quite wrong to associate a machine mechanism with a single behavior, while it is a must to do so regarding a causal mechanism. The italicized definite article in the thesis reflects just this, delegitimizing the machine mechanism reading of *M*. Accordingly we can make the bolder statement that *M* is not only associated but also *is identified* by the behavior of concern.

For someone who is prepossessed by the image of an artifact, it may be difficult to accept that a machine mechanism enables a whole to exhibit *many* behaviors. However, a behavior of a whole need not be an intended function (of a machine). Just think of the conduction of heat: this is a behavior of the moka pot, exhibited *alongside* the making of coffee (the main function of the machine). This is a fact disguised by the design perspective, which highlights the *appropriateness* of a machine mechanism for (an) intended function(s). It also follows from being a whole that a machine mechanism itself can exhibit multiple behaviors; what is more, the machine mechanism may have constituents *irrelevant* to some of its *own* behaviors. This probably sounds odd, especially to people who cannot resist the inclination to think machine mechanisms strictly from the design point of view. Someone could be bothered by this line of thought: If we take the *machine* as the whole, then no doubt the whole has parts irrelevant to some of its behaviors (we have seen examples to this). But a *machine mechanism* is by definition freed from the irrelevant parts. So if we take *it* as the whole exhibiting the behavior, how come we can speak of its parts being irrelevant to this behavior? Take the example of the two moka pot chambers conducting heat. The two chambers are the respective machine mechanism, for the whole they make up (without the basket, the lid, or the handle) is *minimally sufficient* for this behavior (granted are the set-up conditions). In short: What can be the irrelevant part, if the machine mechanism *is* the minimally sufficient whole?

Lest there are people who cannot avoid thinking this way, I must remind that I have *not* defined the machine mechanism as the whole minimally sufficient *for all behaviors it exhibits*. As I indicated above right after the introduction of the criterion, being minimally sufficient for *one* behavior is a necessary and sufficient condition for the whole to be a machine mechanism. This is

why the two chambers of a pot do constitute a machine mechanism (they are minimally sufficient for conducting heat), and this is why when we abstract away the lid and the handle we still have a machine mechanism (minimally sufficient for making coffee, for instance). On the other hand the criterion does not disallow the chambers *plus the basket* to be counted as a machine mechanism *for conducting heat* (besides being the machine mechanism for making coffee).

In sum, machine mechanisms are quite unlike causal mechanisms. For a machine mechanism can be associated with more than one behavior, and some of its parts can be irrelevant to some of its behaviors. The same cannot be said for a causal mechanism, which is associated only with (and thereby identified by) the behavior it is explaining. This is of course nothing but a restatement of the CD Thesis, emphasizing the categorical distinction between a machine and a causal mechanism.

I used the moka pot example primarily for illustrating how the two senses of “mechanism” are demarcated.³⁷ The same distinction applies to “mechanisms” found in natural objects, too, which are the ones considered in the mechanistic philosophy of science. We can say that the parts of a natural object, such as a neuron membrane, constitute a machine mechanism on satisfying the minimal sufficiency criterion for a particular behavior, such as depolarization.³⁸ (The machine mechanism in this case overlaps with the voltage-gated Na^+ channel.) They constitute a machine mechanism, because in the right set-up and background conditions (the presence of Na^+ ions being the most important of all) they *enable* this very behavior. The machine mechanism constituted by these parts do not *explain* the behavior; what explains is the *causal* mechanism. There may well be constituents which are not parts of the machine mechanism but are component entities in the causal mechanism (in the depolarization example, Na^+ ions belong to this group).

In life sciences, the explanatory worth of causal mechanisms possibly makes the Causal-M sense gain dominance over the Machine-M sense. Therefore when scientists are said to discover the mechanism for a behavior/phenomenon, there is a high chance that the *causal* mechanism is of concern. Still, discovering machine mechanisms also makes perfect sense, as an important step in the research by which the phenomenon is “located” in the whole (as examined by Bechtel and Richardson [1993]). This, I take, is the reason that the semantic ambivalence in “mechanism” goes

³⁷ Regarding the explanans-explanandum relationship, the working of a moka pot is subject to an etiological mechanistic explanation: If we are to explain how there is coffee in the collecting chamber, we refer to the causal process leading to this end result. But even if my example was something calling for a *constitutive* mechanistic explanation, the respective machine and causal mechanisms would be distinguished in a similar way. For that matter, the type of mechanistic explanation that the example demands has no bearing on how to distinguish machine and causal mechanisms.

³⁸ Alongside “mechanism”, scientists also appeal to the notion of “machinery”. I believe that by this notion they usually refer to machine mechanisms. Obviously this claim requires textual evidence for a concrete defense.

unnoticed, for *both* senses are invoked in the scientific mechanism-talk. When a mechanistic philosopher (such as Craver) overlooks this, it can have the dramatic philosophical effect of losing the categorical distinction between a whole and a causal mechanism (as formulated by the CD Thesis).

2.3 THE (CAUSAL) MECHANISM HARBORED IN THE SYSTEM

The CD Thesis establishes the categorical distinction in terms of a difference in the number of possible behaviors associated with W and M . In the discussions above, it ultimately served as a tool to differentiate machine and causal mechanisms. I now want to indicate why this difference matters.

It matters because of the implicit attribution of the “ontic surplus” I mentioned at the beginning of this chapter. We have seen that—according to the mechanistic framework—there are two “mechanisms” within a system wherein the investigated behavior is exhibited: One is the machine mechanism that enables the behavior, the other is the causal mechanism by which the production of the phenomenon is explained. Given that the latter is identified by the very phenomenon that it explains, it becomes important to consider its status when there is no instance of the phenomenon. As a starter, think about the assembled moka pot with the coffee and water that is ready to be put on the heat source. If there is no heat, the coffee is not made; there is no ψ^* -ing (making of coffee), and there is no causal mechanism CM^* to be identified by the behavior (yet to be instantiated). CM^* is not instantiated, so to say. Notice that this is a statement made at the token-level. It presupposes that there is a CM^* -type, whose instances are identified by instances of the phenomenon-type of concern (making of coffee). And to remark one more time, there is an inactive *machine* mechanism (labeled as $\overline{mM}$ in Table 2-1 above) even when CM^* is not instantiated. I conjecture that these may create the following false impression: that there is a *latent causal mechanism* harbored in the system, ready to be triggered, which can be grasped or characterized prior to establishing the context of investigation (which determines the explanandum phenomenon). It is because of the need to clear away this impression that distinguishing machine and causal mechanisms has some importance. It is also meant to serve the need to differentiate the ontological categories machine and causal mechanisms belong to: while the former concerns a set of objects, the latter concerns a set of (causally organized) occurrences.

I do not intend to be *challenging* the mechanistic philosopher on this point. For as I have indicated above, distinguishing the two senses of “mechanism” is—in essence—a *service* I provide *for* the mechanistic philosophy of science so as to be clear about the metaphysical framework that comes with it. Also, to note in passing, the distinction I draw between machine mechanisms and causal mechanisms does not intend to address the question of the applicability of mechanistic explanation to real systems of scientific interest. Dupré (2013), Woodward (2013) and Levy (2014) have written on the issue of what makes a system *machine-like* (see next chapter), and the discussion they contribute to elucidates some limitations regarding the truthfulness of mechanism models. But the machine mechanisms *vs.* causal mechanisms distinction drawn above is required to make an ontological-categorical distinction between wholes made up of objects and wholes made up of occurrences. It does not bear a relevance to the discussion about the machine-like and the non-machine-like aspects of the real systems.

This much being just a stage setting, I will be objecting to this framework on the ground that it takes (causal) mechanisms as something harbored within the investigated real system that are worthy of a scientific realist commitment. My aim in the next chapter will be to show that mechanisms have an ontological status (that is, the status of an abstract model) that sets them apart from the real systems they represent. Afterwards, how this affects the realist commitment to them is the question I will be dealing with in Chapter 4. (Stating that x is an abstract entity does *not* directly entail that someone cannot be a realist concerning x . Therefore I will provide an additional meta-ontological argument to substantiate my claim that mechanistic realism is a wrong ontological attitude.) This is going to be offering a line of attack that is different from the two to be briefly considered here: ones due to the scientific antirealist and the reductionist.

The individuation of a causal mechanism by the explanandum phenomenon opens up a space for a scientific antirealist to forward the point that there is nothing to objectively and non-contextually individuate the mechanism and settle its boundaries. An exposition of this can be found in Colombo *et. al.* 2015 (I provide a taster for this paper at the beginning of the next chapter, in connection with the ontic *vs.* epistemic conceptions of explanation); here I shall illustrate the point by considering (the notions of) the “failure” and “success” of a mechanism. As indicated at the beginning of the chapter, it is known that the neurotransmitter release mechanism “fails” more often than it “succeeds”. How are we to make sense of these? Failure and success are notions that are relative to our scientific interests; the indeterministic cellular system of concern is impartial to the behaviors that the wholes exhibit. Therefore it can be argued that it is us who privilege, label and individuate a phenomenon (*viz.*, neurotransmitter release), and accordingly, that the scientific mechanism-talk

does not philosophically lead one to commit to the existence of mechanisms (independently of our scientific-theoretical perspective). To this antirealist charge, I do not feel any need to defend the mechanistic realist—this would only be a digress on my part. I have my own reasons to be against *both* mechanistic realism and scientific antirealism. The present study is concerned only with being against the former from a *scientific realist* viewpoint; the scientific realism vs. antirealism debate, on the other hand, is too huge to be handled here. Thus I assume that this line of attack is irrelevant to my project (and will be concerned only with why and how a realist could be a mechanistic *non*-realist).

The second type of challenge to be considered is due to Soom 2012, which is a sophisticated defense of functional reductionism in the context of mechanistic explanation. Soom's main target is Craver's claim that the mechanism *M*'s components stand in real (ontic) relations with the explanandum phenomenon. He summarizes his thesis as follows:

[G]iven appropriate background conditions, mechanisms are sufficient conditions for the phenomena they are mechanisms for. This sufficiency condition is the cornerstone for the discussion of the metaphysical status of phenomena and mechanisms since it ensures that the phenomenon is causally nothing “over and above” the underlying mechanism and thus that the concept of strong supervenience can be used to describe the relation between phenomena and mechanisms. Thus, contra the new mechanist orthodoxy [mechanistic discourse], and in accordance to Kim's supervenience argument, different levels of mechanisms are not different ontic levels, but rather different levels of description referring to various degrees of organization. (2012, 655)

This thesis can be regarded as a continuation of the contention that causal mechanisms are identified by a single behavior: It could be thought, then, that there is a philosophically interesting dependence relation between mechanisms and the phenomena they explain. It is this that Soom attempts to flesh out, by first providing a *functional* reading of the explanandum phenomenon and the mechanism—that is, by arguing for their individuation according to their *causal roles*. He then goes on to show that for the mechanism to explain the phenomenon, these causal roles must be identical. It follows from this that the two stand in a strong supervenience relation (as conceptualized in Kim 1984).

Soom's is in fact a contribution to the discussion found in Chapter 6 of Craver 2007. There, Craver defends his position concerning the causal powers of high- and low-level occurrences separated by ontic levels of mechanism. According to him, the causal exclusion problem introduced by Jaegwon Kim (concerning mental causation, see Kim 1998) might be thought to be applicable to his account, challenging his contention that high-level occurrences bring in distinctive and genuine

(irreducible) causal powers to the world. Very briefly, Craver aims to demonstrate that the causal exclusion problem appeals to “levels of realization” (to properties separated by the realization relation), which cannot be identified with levels of mechanism. The upshot is, levels of mechanism is a notion that is immune to Kim's well-known charge against non-reductive physicalism. From a broader perspective, the whole discussion by Craver can be seen as a defense of characterizing mechanistic explanation with its multi-level scope, that is, the non-reductive sentiment found in the mechanistic discourse (the third tenet introduced in Chapter 1). The question Craver tackles with amounts to *whether* the particular reductive argument put forth in philosophy of mind is relevant to mechanistic philosophy. His own answer is “No.” And Soom thinks that Craver is quite mistaken here: that Craver makes an incorrect reading of Kim's conceptualization of supervenience, and that the true metaphysical nature of the mechanism-phenomenon relation eventually makes Craver's account vulnerable to the relevant charge. In the context of mechanistic explanation, this vulnerability leaves Craver (or any other mechanistic philosopher for that matter) only one option (which follows from Kim's argument): accepting that an explanandum phenomenon *P* and its explanans mechanism *M* are token-identical.

This is surely a bold metaphysical claim, the argument for which deserves a longer exposition. But I rather have to focus on implicating what will be different in the argument to come in the following chapter. The key difference will be my challenging the conviction that models of mechanisms give a truthful representation of what goes on in real systems. Despite the challenge he poses, even Soom shares this conviction, for talking about token mechanisms. My view, however, is more sympathetic to the idea that there are no token mechanisms (see Section 4.4). For if by mechanisms we are to understand the causal organization of locally and mechanically interacting component entities, then the respective token occurrences can be shown to be insufficient for explaining the token phenomenon. This is to say, mechanism components (factors found relevant for the phenomenon) are not instantiated as they are described in models. As a scientific realist I do think that these components stand for real individuals engaging in real causal relations; so that there are real and repeatable patterns of token occurrences giving rise to token phenomena. Furthermore, it might even be granted that these patterns are token-identical with the phenomena, exactly due to Soom's argument. However, these patterns are quite different than mechanisms of concern in mechanistic explanations.

I will attempt to justify this claim by showing that the concrete production of phenomena exceeds the totality of the mechanical interactions between token objects (or object-like entities) represented by mechanism components. The causal narrative employed in mechanistic explanations

involve abstractions and idealizations; especially due to the latter, mechanisms turn out to be fictional configurations of abstract and type-level components.



CHAPTER 3

CAUSAL MECHANISMS AS IDEAL MODELS

3.1 MECHANISMIC REALISM

In this chapter I provide the argument for the main thesis of the dissertation, which can be expressed shortly as follows: Mechanismic philosophers of science are wrong in their ontological commitment to mechanisms. I want to show what is wrong with *mechanismic realism*, the brand of scientific realism involved in the mechanistic discourse. In virtue of my own convictions regarding scientific realism, however, the argument to be developed will not be made from a scientific antirealist standpoint. Being a dispositional realist (drawing mostly on Chakravartty 2007), I hold that there is an objective ground that justifies commitment to the causal (or dispositional) properties in the world, whose discovery and analysis are essential for scientific progress. In the context of this dissertation, this stance functions as the background assumption, implying that I will not base my argument in scientific antirealism. There will be an argument in favor of dispositional realism in the next chapter, but it will serve for making a constructive point with regards to this question: “If not to mechanisms, to what might mechanistic philosophers commit to?” This is to say, counterarguments to mechanistic realism developed in in this chapter do not require one to be a dispositional realist; but, as I will claim in the next chapter, endorsing dispositional realism is a good option for advancing a mechanistic non-realist *yet* scientific realist perspective. Acknowledging these at the outset should make the following discussion more intelligible—or at least, this is what I expect for.

Let us begin with a divide among mechanistic philosophers set on a difference in their general treatment of mechanistic explanations. It is a standard in mechanistic philosophy to separate mechanisms themselves from their descriptions (or models). This is taken to correspond to a difference between the objective features of the causal structure of the world that are relevant to the production of the explanandum phenomenon and the scientific representations of them, respectively. Although all mechanistic philosophers contend that mechanistic explanations are about mechanisms (call this “the aboutness claim” for referential purposes), they disagree over whether it is the mechanism or its representation that does the explaining. On one side, we have

Craver for instance, who writes “The explanans is a mechanism,” in his (2007, 139). This is a reiteration of the point he implies in other works, too:

[E]ntities and activities are organized such that they carry out some process, perform some function, or produce some end product. Call this the phenomenon to be explained by the mechanism. (Craver quoted in Wright 2012, 378)

Constitutive explanations go beyond merely describing the phenomenon. They describe the mechanism responsible for the phenomenon, that is, the mechanism that explains its diverse features. (Craver 2006, 359)

Wright observes how entrenched this idea is among the mechanistic philosophers (citing Darden, Machamer, Andersen and Glennan as well—see his 2012, 377-379). On the other side, we have Bechtel and Abrahamsen writing: “But it is crucial to note that offering an explanation is still an epistemic activity and that the mechanism in nature does not directly perform the explanatory work.” (2005, 425). It should be noted that this is a less frequently voiced thought in the relevant literature.

The terminology accompanying this disagreement draws upon Salmon's labels for distinguishing two conceptions of scientific explanation, namely, the *ontic* and the *epistemic* conceptions (see e.g., his 1984a). We can say that Salmon's purpose in introducing these was to conceptualize the originality in his causal-mechanical account of explanation. The account is well-known for abandoning the idea of seeing explanations as arguments (epistemic conception—Salmon was addressing Hempel's model in particular), favoring instead the *ontic* demonstration of the fit of the explanandum in the (causal) structure of the world: “The ontic conception sees explanations as exhibitions of the ways in which what is to be explained fits into natural patterns and regularities.” (1984b, 293) On the use of ontic vs. epistemic in the mechanisms literature, Illari makes the noteworthy claim that the discussion (concerning mechanistic philosophers) is not about Salmon's original target (the deductive-nomological model of explanation) anymore (2013, 240). It is rather about whether the constraints on mechanistic explanation are determined by the ontic or the epistemic criteria. She examines and compares Craver's (ontic) and Bechtel's (epistemic) positions in order to establish that a common ground can be found in between (Craver 2014¹ and Bechtel 2011 are the main papers she cites).

Illari's argument aside, there really appears to be a common ground among the mechanistic philosophers, and in essence, it is the aboutness claim I mentioned above: “mechanistic

¹ Craver 2012 in Illari.

explanations are about mechanisms.” My noting this may look trivial: After all, what else mechanistic explanations could be about? But if the arguments in the rest of the dissertation are sound, I should be able to at least raise a doubt on this point. To begin with, let me substantiate the claim that all mechanistic philosophers agree on the objective existence of mechanisms, by quoting Craver and Bechtel and Abrahamsen (below). What merits our attention is, of course, the latter authors' position; because in spite of having a different opinion about what the explanans is, Bechtel and Abrahamsen join the others in thinking that there *are* mechanisms “operative in nature”.

There are mechanisms (the objective explanations) and there are their descriptions (explanatory texts). *Objective explanations are not texts; they are full-bodied things.* They are facts, not representations. They are the kinds of things that are discovered and described. There is no question of objective explanations being “right” or “wrong,” or “good” or “bad.” They just are.

Objective explanations, the causes and mechanisms in the world, are the correct starting point in thinking about the criteria for evaluating explanatory texts in neuroscience. . . . Good mechanistic explanatory texts . . . are good in part because they correctly represent objective explanations. Complete explanatory texts are complete because they represent all and only the relevant portions of the causal structure of the world. (Craver 2007, 27 [emphasis added])

Providing explanations, including mechanistic explanations, is essentially a cognitive activity. This is particularly obvious when one considers incorrect mechanistic explanations—in such a case one has still appealed to a mechanism, but not one operative in nature. Another way to appreciate the point is to note that in many instances *the mechanism in question was operative long before scientists discovered the mechanism* and invoked it to explain the phenomenon. (Bechtel and Abrahamsen 2005, 425 [emphasis added])

Illari and Williamson's “Mechanisms Are Real And Local” (2011) should also be cited next to these, for displaying a quite explicit defense of the ontological commitment to mechanisms—in other words, of mechanistic realism. The two authors further think that the epistemic conception is “parasitic upon” the ontic conception (they use the term “physical explanation” to refer to the latter), for according to them, “[i]f epistemic explanations are to explain, rather than merely being stories, there must be real mechanisms to describe.” (2011, 824) This very idea is countered by Colombo *et. al.* from a scientific antirealist perspective. For these critics of the mechanistic philosophy, a mechanism model is true in virtue of its coherence with other theories and models (2015, 199), and an epistemic conception of mechanistic explanation can do quite well without requiring the model to correspond to anything in the world (2015, 202).² Nevertheless, to my

² In fact, Colombo *et. al.* think that an antirealist construal of mechanistic explanation has certain advantages in its dispelling two problems (that a realist construal faces): Identification of the spatial and temporal boundaries of mechanisms independently of the epistemic and pragmatic goals, and ascertaining that different mechanism tokens instantiate the same mechanism type. See their (2015, Section 6) for the relevant arguments. I think that these are

knowledge, Illari and Williamson are the only mechanistic philosophers who attempt to stand against antirealist charges. This might be related with the fact that the ground for this particular discussion is yet to be explored. For although being a mechanistic realist is the etiquette of all mechanistic philosophers, the inherent ontological commitment is not explicated or problematized, save for Illari and Williamson's paper. This is also acknowledged by the two philosophers: “[T]he mechanisms literature is implicitly or explicitly committed to mechanisms being real. The claim is seldom made so baldly: perhaps it has seemed too obvious to comment on.” (2011, 824) I agree with their guess that the claim possibly have seemed “too obvious” to mechanistic philosophers, but then this makes their case even more noteworthy for the following reason: As is also suggested by Colombo *et. al.*, there is nothing in mechanistic explanation that impels one to mechanistic realism. Hence the rationale or motivation for this brand of scientific realism appears to be external to the mechanistic discourse. (But as already noted, Illari and Williamson disagree with this—cf., pp. 823-824.) As a consequence, the only thing I see as obvious is the mechanistic philosophers' being mechanistic realists. Yet this is not a tenable position as the majority of them seem to suppose, for reasons that will be elaborated in the subsequent sections.

Illari and Williamson consider two charges that can be brought against mechanistic realism. The first concerns “non-‘physical’ mechanisms” that psychological and social sciences appeal to, the challenge with which is that these mechanisms are *prima facie* mind-dependent, so one may argue against their objective existence (2012, 824). The example the authors give is the mechanism for economical recession, where the relevant beliefs and acts of people can be thought as its components (e.g., belief that economy is going into recession, acting in accordance with this belief, which ultimately lead to the actualization of the recession). At least some of these components are constituted by the mental states of people and depend much on how the world is represented (instead of what objectively goes on); therefore one might contend that the explanation for the recession is an epistemic exhibition—that is, not ontic. Illari and Williamson respond to this by remarking that the minds and the people belong to the world (p. 825). So what underlies the recession is there in the world, and accordingly the mechanism is not unreal—it has a concrete existence.

From where I stand, however, this does not seem as a satisfactory response. To give a hint of a point I will make later, contending that there is an objective basis for a given mechanistic explanation makes one at most a scientific realist. Mechanistic realists, on the other hand, advocate

forceful challenges that a mechanistic realist cannot easily fend off. Nonetheless I contend that the dispositional realist position I outline in the next chapter offers a viable perspective for someone to stand against these charges.

the more strict thesis that the objective bases of mechanistic explanations are mechanisms as the real features of the world. The natural question that should come to mind is, then: What is the metaphysical warrant for this? Or, to express it under a different guise: Granted that there are real features of the world being the objective truthmakers for models of mechanisms, why must it be the case that these features are the activities of entities whose causal organization is as described in the models? In other words, why should it be that the mechanism models correspond to nothing but mechanisms? Just because of being models of mechanisms by definition? If so, we are back at where we start: How could the definition of their models warrant an ontological commitment to mechanisms? In short, Illari and Williamson's response lack a crucial element to bridge the metaphysical gap between mind-dependent mechanisms and the purported “physical” mechanisms.

Let us move on with how they reflect on the second charge that is related with the individuation of mechanisms by their functions. By now, we should be familiar with the mechanisms' always being mechanisms for phenomena. Illari and Williamson think that this enables the scientists to individuate the mechanisms according to their role in a given system, the role being “characteristically produc[ing] that behavior [the explanandum]” (2012, 825). One of their examples is natural selection: Depending on the explanatory context (or, the system considered), natural selection can be taken as the mechanism for adaptation, or just a process running in the background (p. 826). Illari and Williamson note the scientists' consistent choice for the label here: They observe that scientists name natural selection as a “mechanism” only when it is foregrounded by its role of producing the investigated phenomenon (i.e., adaptation, in the present example); otherwise it is not seen as a mechanism, but a “process” that is not individuated by a phenomenon. (Processes are, of course, supersets of mechanisms.) Then the charge against realism is, whether there is a mechanism or not hinges on the context; there is nothing to objectively settle if a process is a mechanism. The mechanistic philosophers respond to this by recalling that entities (objects) play many roles and belong to more than one system. “[I]t is a worldly fact,” Illari and Williamson say, that entities do play roles in many systems, regardless of what their roles are described to be, or whether or not the processes they are involved in are functionally individuated (p. 827). The authors' conclusion is, the issue of description does not impair the reality of mechanisms, for their component entities belong to the objective world with the numerous causal relations they really contribute to.

To this, again, I could raise the point suggested above: That there really being entities contributing to causal relations could be the objective basis for mechanistic explanations. (I leave the task of questioning and challenging this to the antirealist.) Yet mechanistic realism demands an

additional argument to the end of showing that the real components and their causal organization are as described in the model of a mechanism. Mechanismic philosophers assert that mechanistic explanations identify “the portion of the world” that is “responsible” of the investigated phenomenon. (The quoted phrases are from Illari and Williamson [2012, 823], but the same claim is found in many of the works on mechanismic philosophy—recall from the last chapter Craver's differentiating the relevant and irrelevant parts of a whole, for instance.) Basically drawing on Salmon's account of explanation, for them, a mechanism described in a complete mechanistic model truthfully represents (corresponds to) the organization of all and only the occurrences relevant to the explanandum phenomenon which are harbored within the causal structure. I suppose accordingly that the visual impression they have must be as follows: The causal structure is something that can plausibly be depicted as a complex neuron diagram (I am referring to the type made famous by Lewis in causal reasoning), where the nodes represent token phenomena, the arrows connecting the nodes represent the causal relations between them, and most importantly, mechanisms are represented by a select group of nodes connected by the arrows. Selecting a certain portion of this (to stand for the mechanism for the investigated phenomenon) depends a lot on epistemic and practical concerns; but it is still the *real* relations between the phenomena that are selected and represented by models of mechanisms. A mechanismic philosopher would take such a graphical representation of the relevant portion of the causal structure to be quite *realistic*—in the sense of (more or less) a painting's being a realistic depiction of the objective world. I have not encountered any mechanismic philosopher asserting the reality of mechanisms in this fashion; despite this, I hold that they would all be content with this sort of illustration of their stance.

And my being against mechanismic realism should be read as being against the models' supposed realistic representation of the causal structure (or its relevant portions, if this matters). According to the stance I favor, mechanisms are found nowhere other than the ideal domain—that is, the domain of mental representations. Models of them do represent an objective causal structure (this is where I depart from the antirealist), but the representational elements of the model (mechanism components and their configuration) do not directly correspond to real features of the world, be them entities, activities, objects or occurrences. This is not say that there are not entities etc.; this is to say that whatever there is within the causal structure, the relations between them are not as models of mechanisms describe them to be.

Mechanistic explanations provide epistemic access to the causal structure by *misrepresenting* the complex causal relations. The characteristic form of misrepresentation in mechanistic explanations is the *idealization* of the causal organization of occurrences: Mechanism models describe the

interactions between real occurrences by likening them to those that obtain among the localized parts of *machines*. This is surely a legitimate way of explaining natural phenomena, especially those investigated in life sciences. But in virtue of its appealing to qualitative reasoning (that is, its focus on the qualitative instead of the quantitative aspects of the system) this practice has certain limitations—in effect, making it doubtful if mechanistic explanations are as central as the mechanistic philosophers take them to be. My specific concern in the next three sections is the mechanistic philosophers' overstatement with regards to the significance of providing complete models of mechanisms. Here I examine Hodgkin and Huxley's mathematical model of the action potential as a case in point. While devising their model, the two scientists were certainly ignorant of the interactions at the molecular level that give rise to the generation of the action potential wave on the neuron membrane. I delve into the discussion of how critical this was in providing a causal explanation for the phenomenon. Craver takes this to be very critical, because according to him, not knowing the molecular details amounts to not knowing the “mechanism underlying the phenomenon”. The claims to be put against his view will hopefully lead us to the position I favor, according to which mechanisms do not “underlie” phenomena—they are just a type of mental apparatus for the scientists to represent what really underlie them (i.e., genuinely occurrent causal properties).

3.2 THE HH MODEL OF ACTION POTENTIAL AND CRAVER'S EVALUATION

I spent some time on introducing the phenomenon of the generation of the action potential (in neurons) in the previous chapter. There, I also reviewed how this phenomenon is explained mechanistically—or, put more properly, how Craver thinks his account of mechanistic explanation conforms to the modern textbook explanation of this phenomenon. According to Craver (and the mechanistic philosophers he represents), the more recent (non-quantitative) models of the action potential are superior to the landmark work on this phenomenon, the Hodgkin and Huxley model (the HH model), for revealing the real molecular mechanism underlying it. Almost all we need to know about the related discussions are already covered, but the details of the HH model were left out. This section will be the place to go into them, to be succeeded first by Craver's then by Weber's and Levy's evaluations. (The latter two are critical of Craver's views on the issue.)³

³ Besides the sources cited within the main text, I referred to (Johnston and Wu 1995, Chp. 6) while writing the parts on the HH model in this section.

The action potential had been investigated in the earlier decades of 1900s, but the real progress was initiated by Hodgkin and Huxley's experiments in early 1950s. (Katz was also involved in one of their studies.) Hodgkin and Huxley knew at that time that depolarization and the action potential were related, and that both phenomena were mainly results of the movement of sodium and potassium ions. Their aim was to understand how the magnitude of these ionic currents, which underlay the generation of the action potential, changed with time. They also studied the changes in the *conductance* of the membrane for sodium and potassium (explained below). By using a newly developed technique, the *voltage clamp method*, they were able to keep the voltage across the membrane of the giant axon of squid constant at desired levels.⁴ The maintenance of constant voltage (membrane potential) was important for eliminating the influence of voltage on ion conductance and the *capacitative current* across the membrane, which is the transient current due to instant voltage changes. By isolating the movements of the ions across the membrane, Hodgkin and Huxley were able to see that the generation of currents owed to rapid sodium influx followed by slower potassium efflux.

They thought that the changes in the ionic currents were caused by changes in the *permeability* of the membrane. In accordance with modeling the membrane as an electric circuit (see their figure on page 501 of 1952), they represented this as a change in the *conductance*.⁵ The latter is the inverse of the electrical resistance, and so, it follows from Ohm's Law (which states that the multiplication of current I and resistance R gives the voltage; $V = IR$) that the conductance for a specific ion (g_{ion}) is related with the ionic current (I_{ion}) in the following way:

$$I_{ion} = g_{ion} (V - E_{ion})$$

Here, V is the membrane potential maintained by the method Hodgkin and Huxley used and E_{ion} is the equilibrium potential for the ion: that is, the potential that obtains when there is no net ionic movement across the membrane (when diffusive force is balanced by the electrical attraction).⁶ The values V , E_{Na} , E_K being known, and with the observations regarding the time course of the ionic

⁴ The reason for choosing this animal was the size of its nerve fibers, which allowed the scientists to use the voltage clamp method. The diameter of the nerve fiber of the squid can be as much as 1 millimeter.

⁵ For this reason, in what follows, “permeability” and “conductance” are used interchangeably.

⁶ Ions diffuse downstream across the membrane, from the medium with high concentration to the medium with low concentration. On the other hand, they are attracted back by the oppositely charged substances. The difference between the membrane potential and the equilibrium potential for the ion ($V - E_{ion}$ in the equation) should be read as the *driving force* for that ion to move across the membrane, generating a current.

currents (I_{Na} and I_K) as a function of voltage, Hodgkin and Huxley could calculate the conductances for the ions (g_{Na} and g_K).

As a matter of fact, they observed that the conductances were time and voltage-dependent. The peak sodium and potassium conductances were found to increase similarly and progressively with voltage. However, as for the time-dependencies, while sodium conductance reached its peak quite rapidly and then fell, potassium conductance was found to have a delayed and steady increase (in comparison to sodium conductance, it increased later and slower and did not decrease). In light of these, Hodgkin and Huxley understood and showed that as the neuron membrane was depolarized (that is, as the inside was made less negative than the outside, and then, more positively charged), *i* - sodium conductance was activated and then immediately inactivated, and, *ii* - potassium conductance was also activated, but this occurred later. They then devised the following mathematical model of the change in the total current (I) through the neuron membrane:

$$I = I_c + I_{Na} + I_K + I_{leak}$$

Here, I_{Na} and I_K are the ionic currents as described above, and I_{leak} is the current due to minor leaking ions. Just as the sodium and potassium currents, the latter can be also written as a product of the relevant conductance value (g_{leak}), which is a constant, and the difference between the membrane potential and the equilibrium potential, ($V - E_{leak}$). I_c , on the other hand, is the capacitative current and is expressed as the product of the membrane capacitance C_m (the capability of the membrane to keep the opposite charges on two sides) and the instant change in the voltage, dV/dt . The total current equation then becomes:

$$[TC] \quad I = C_m dV/dt + g_{Na}(V - E_{Na}) + g_K(V - E_K) + g_{leak}(V - E_{leak})$$

Hodgkin and Huxley accounted for the voltage and time-dependencies of g_{Na} and g_K by introducing three exponential functions h , m and n (the *gating variables*), each giving a unitless value between 0 and 1. These are functions of voltage and time and they are associated with the *inactivation* of the sodium conductance, *activation* of the *sodium* conductance and the *activation* of the *potassium* conductance, respectively. It is important to note that these functions are used for the

purpose of curve-fitting: They mathematically express the rise in ion conductances with time (and the fall, in the case of sodium conductance) actually observed. Where G_{Na} and G_K are the *maximal* values of conductances for sodium and potassium observed, respectively, the voltage and time-dependent ion conductances are

$$[c1] \quad g_{Na} = G_{Na} m^3 h$$

$$[c2] \quad g_K = G_K n^4$$

making the total current equation

$$[TC'] \quad I = C_m dV/dt + G_{Na} m^3 h (V - E_{Na}) + G_K n^4 (V - E_K) + g_{leak} (V - E_{leak})$$

Formulated as above, [TC'] is the quantitative description of the time-dependent change in the ionic current across the neuron's membrane. Although it is the totality of [TC'] that does the work here (i.e., the modeling of the action potential), the expressions for conductances ([c1] and [c2]) present the crux of the matter. Let us focus in on them for a moment for a clear grasp of the discussion to follow.

The change in membrane conductance for specific ions is an important part in the generation of the action potential. Hodgkin and Huxley's conjecture about the phenomenon was that there could be ion carrier molecules in the membrane that would become free (activated) upon depolarization and would thus become able to transfer the sodium ions. These carriers would then be inactivated as depolarization became less effectual (Huxley 1963, 59-60). On an alternative consideration of this activation-inactivation hypothesis, there could be gating particles whose position in the membrane would allow or block (or, determine the rate of) the passage of the ions (Hodgkin and Huxley 1952, 502-503).⁷ It would later turn out that the correct explanation does not involve carrier molecules but ion-specific channels that account for the gating (as I mentioned in the previous chapter). Nonetheless (and fortunately), the falsity of the conjecture did not impede or devalue their work. For the essential task for Hodgkin and Huxley was to devise “reasonably accurate” and “sufficiently simple” (Hodgkin and Huxley 1952, 506) mathematical equations whose solutions would fit the

⁷ They indicate that similar considerations apply in the case of changes in potassium conductance.

experimental data, and this did not necessitate them to present any hypothesis (although they did, with a remarkable disclaimer—see the quotation below). The measurements they made at different steps of V (the maintained membrane potential) provided the maximal values of ion conductances reached (the voltage-dependent, time-independent G_{Na} and G_K), and the time courses of g_{Na} and g_K . The mathematical labor that followed this was essentially curve-fitting: in other words, describing the functions whose graphs agree with the empirical data optimally. Preference for simplicity led Hodgkin and Huxley to approach this by first-order differential equations taken to low values of powers: Hence the “ m^3h ” and “ n^4 ” in [c1] and [c2].⁸ (Let me remark that m , h and n are functions. We do not need to elaborate on the relevant equations showing how these gating variables change with voltage and time.)

Showing how the membrane potential is related with the ion permeability, Hodgkin and Huxley's mathematical model of the action potential stands as a neat achievement in the history of electrochemistry and neuroscience. The model (including the experimental apparatus [voltage clamp method] and the assumption that the membrane behaves like an electric circuit) enhances the relevant calculations and predictions concerning many electrical properties of an axon.⁹ Arguably, though, its *explanatory worth* is a different matter. Hodgkin and Huxley's activation-inactivation hypothesis does not seem to have an effect on how the model is constructed. Likewise the model has no bearing on the actual electrochemical processes giving rise to the regularities of concern. Their own comment on the success of the model (its agreement with the empirical data) is important for acknowledging the possibility of other hypotheses and explanations:

The agreement must not be taken as evidence that our equations are anything more than an empirical description of the time-course of the changes in permeability to sodium and potassium. An equally satisfactory description of the voltage clamp data could no doubt have been achieved with equations of very different form, which would probably have been equally successful in predicting the electrical behaviour of the membrane. It was pointed out in Part II of this paper that certain features of our equations were capable of a physical interpretation, but the success of the equations is no evidence in favour of the mechanism of permeability change that we tentatively had in mind when formulating them. (1952, 541)

⁸ In their article they briefly consider the alternatives: using one variable which obeys a second-order equation instead of the “ m^3h ” and taking n to the power of 5 or 6. All of these alternatives result in better fit with the data, with a compromise in simplicity. (1952, 509 and 512)

⁹ “[T]he form, duration and amplitude of spike [the action potential wave], both 'membrane' and propagated; the conduction velocity; the impedance changes during the spike; the refractory period; ionic exchanges; subthreshold responses; and oscillations.” (Hodgkin and Huxley 1952, 540-541) See also (Craver 2007, 51). Of course, to recall, these are not claimed to be universal properties of *all* axons, but that of an isolated giant squid. There is a discussion of the range of applicability of the equations in (Hodgkin and Huxley 1952, 541-542).

We might agree with Hodgkin and Huxley here, on the contention that providing an “empirical description” does not complete the story. As I touched upon in the previous chapter, this is Craver's position, who asserts that the mathematical dependencies between the variables in [TC'] do not harbor a *causal content*. While it is true that Hodgkin and Huxley thought the existence of some ion carrier molecules could provide such content, they had “no evidence in favour of” this causal interpretation, making their model at best a “‘how possibly’ sketch of the action potential.” (Craver 2007, 57) The HH model, according to Craver, was a *sketch*, because it left a “black box” in the explanation concerning the permeability change; and it was a *how possibly model*, because Hodgkin and Huxley had no idea regarding the existence of the conjectured entities (the ion-carrier molecules) (2006, 360-361). Quoting from him:

The equations for the conductance changes . . . are merely phenomenological models. These equations characterize how specific ion conductances change with voltage and time, but *they do not explain why* they change as they do, when they do. Without an account of the underlying mechanisms of the conductance change, the buck of accounting for the temporal features of the action potential . . . is merely passed on to *some-conductance-changing-process-we-know-not-what*. (2007, 59 [emphases added])

Craver thus contrasts the HH model with the models enhanced by the subsequent researches on the neuron membrane (structural studies), which uncovered (to a great extent) the conformational changes in the voltage-gated ion channels. He thinks, at the present state of science, a plausible mechanism of the action potential can be described, thanks to these developments. And it is these descriptions but not the “merely phenomenological models” that are the explanatory ones: “Models are explanatory when they describe mechanisms. Perhaps not all explanations are mechanistic. In many cases, however, the distinction between explanatory and non-explanatory models is that the latter, and not the former, describe mechanisms.” (2006, 367)

For Craver, describing mechanisms is important not only for its own sake, that is, for widening and deepening the knowledge about the objective, empirical world (mechanismic realism). It is also important as it enables scientists to knowingly control and manipulate the phenomena. Besides providing an account of why the HH model is an accurate mathematical tool, the description of the mechanism for the action potential answers many “what-if-things-had-been-different” questions—much more of these than the HH model itself can address. (Here, Craver relies on Woodward's [2003] terminology and his manipulationist account; 2006, 358.) To see the connection between describing and controlling-manipulating, let us first recall from the last chapter: Concerning the depolarization phase, we talk about, for instance, the helix-shaped segment of the sodium channel, how it is kept in fixed position during the resting state of the neuron membrane, and by which

activity it contributes to the opening of the channel. On Craver's account, the explanation involving the component entities and activities as such exemplifies a *pragmatically* complete (if not *ideally* complete) and *good* mechanistic explanation. More specifically, its superiority follows from the discoveries of the intricate features of the (spatio-temporally organized) parts of the neuron membrane, which paved the way to an understanding of the causal relations that they figure in:¹⁰ Many experiments made against multiple theoretical settings confirmed the ion channels' existence (in philosophical jargon, the robustness criterion for them has been satisfied) and revealed their numerous properties. By knowing the behaviors of the channels in different conditions and their interactions with different substances, scientists can intervene in the causal processes. They can thus see how manipulating a part (e.g., a channel itself, or a part of a channel) results in changes in the other parts, and model the causal organization of the mechanism accordingly. Likewise, they can understand how changes in the initial conditions (e.g., extracellular sodium concentration, density of the ion channels in a given membrane area) can lead to variations in the manifestations of the phenomenon (e.g., the rate of rise of the action potential spike, its peak magnitude, etc.), or, possibly, to its complete inhibition. In short, in the mechanistic discourse, all of these amount to a progress from providing a mechanism sketch to the complete mechanistic explanation for the phenomenon. Arguably, this is quite as expected from a scientific and mechanistic realist standpoint: By each step, scientist *discover more* of what is objectively there—that is, the mechanism. The models describing the mechanism are getting better by giving a more truthful picture of how the generation of the action potential fits into the causal nexus. They become more truthful because they tell us more about the qualitative properties of the component entities, component activities and how these are causally organized.

Or so a mechanistic realist thinks.

3.3 APPRECIATING THE CAUSAL CONTENT IN THE HH MODEL

Seen from Craver's perspective, Hodgkin and Huxley's own evaluation of the total current equation supports the argument for the explanatory deficiency in the HH model. Nonetheless,

¹⁰ The following is a very brief overview of (Craver 2006, 367-373), where his main objective is to lay down a tentative checklist for assessing mechanistic explanations in general. Craver treats each of the four characterizing features of a mechanism (the explanandum phenomenon, component entities, component activities and organization) separately here. This suggests, each of these poses normative requirements, and the goodness of a mechanistic explanation depends on the success in accounting for all, giving each its own due. For a more extensive discussion of the requirements, see (Craver 2007, Sections 4.4 – 4.7)

Weber (2008) and Levy (2013b) offer an alternative reading of the case, which is persuasive in elucidating the causal content that Craver thinks the model lacks. Their compatible arguments take cue from the following paragraph in Hodgkin and Huxley's article which comes immediately after the one I quoted above:

The point that we do consider to be established is that fairly simple permeability changes, in response to alterations in membrane potential, of the kind deduced from the voltage clamp results, *are a sufficient explanation* of the wide range of phenomena that have been fitted by solutions of the equations. (1952, 541 [emphasis added])

In Levy's words, Craver focuses on the “gap” in the model that all sides of the debate acknowledge; but what merits attention is “how it was bridged” by Hodgkin and Huxley (2013b, 8). Levy and Weber's accounts for the causal/explanatory part of the HH model (the “bridge”) differ in details, but they are in agreement on keeping the *causal content* apart from the *molecular basis* of the model (Weber 2008, 1000; Levy 2013b, 14). They share the thought that what Hodgkin and Huxley—legitimately—counted as “a sufficient explanation of the wide range of phenomena” had nothing to do with the latter. (This is where they both object to Craver's interpretation.)

Weber uses textual evidence in order to show that Hodgkin and Huxley had an unmistakable causal picture in mind, concerning the dependence of the temporal and directional (inwards/outwards) aspects of particular ionic movements on the changes in the membrane potential (Weber quoting from Hodgkin and Huxley 1952, 500-501; cf., Craver 2007, 56). He states that this picture corresponds to the “action model” part of the HH model, while the mathematical expressions given for the changes in conductances ([c1] and [c2] above, plus the accompanying formulations given for the gating variables [not reproduced here]) constitute the “conductance model” (2008, 999). The fact Weber wants us to note is, despite the falsity in Hodgkin and Huxley's speculation (the ion-carrier molecules and activation-inactivation hypothesis), the *conductance* model had been “part of a successful empirical test of another model [the *action* model] that is essentially correct.” (2008, 1001) Their action model was to be confirmed by many other studies on excitable cells to follow, but the fit of the solutions of [TC'] (the total current equation) with the phenomena at hand had already been a good evidence for Hodgkin and Huxley. They acknowledge this achievement as follows:

The results presented here show that the equations . . . predict with fair accuracy many of the electrical properties of the squid giant axon. . . . This is a satisfactory degree of agreement, because the equations and constants were derived entirely from 'voltage clamp' records, without any adjustments to make them fit the phenomena to which they were subsequently applied. (1952, 540-541)

The properties they refer to are those represented by the actual action potential wave (see my footnote 9)—that is to say, the ion conductances are *not* included among them. Curve-fitting had been the issue in deriving “the equations and constants” needed for the formulation of [c1] and [c2], the data for which was obtained *independently* from actual tracings of action potentials recorded (at different depolarization values). Then, on Weber's construal, these were put to work alongside the causal assumptions inherent in the action model, in order to be expressed as the total current equation. And it is *this* equation whose remarkable agreement with the tracings of action potentials suggests strong evidence for the *action model*. To remark, the conductance model “was good enough to help showing that the *causal content of the action model* was able to generate action potentials (because it contained the right dynamics for these variables, even if the physical model of how this dynamics came about was false).” (2008, 1001 [emphasis added])

On top of this, Weber resorts to Woodward's account to argue that the HH model is a “a beautiful case of an experimentally determined *causal regularity*.” (p. 1002 [original emphasis]) On Woodward's account, generalizations provide causal explanations if two requirements are met: invariance under certain interventions (stability), and modularity. Weber contends that the HH model satisfies both. By the voltage clamp method, Hodgkin and Huxley were able to manipulate the transmembrane voltage and could observe how the value of the total current changed accordingly. The total current equation they developed thus described this relation between V and I , which held under many interventions on the independent variable (the voltage), satisfying the invariance requirement. With respect to the modularity requirement, Woodward writes: “[A] system of equations will be modular if it is possible to disrupt or replace (the relationships represented by) any one of the equations in the system by means of an intervention on (the magnitude corresponding to) the dependent variable in that equation, without disrupting any of the other equations.” (2003, 48; see also his 2013, 51) Weber observes that this is also met by the HH model, for it allows interventions on a maximal conductance for a particular ion (recall [c1] and [c2]—Weber gives the example of increasing the potassium conductance by a potassium ionophore) that do not affect the other equations. He thereby concludes that Hodgkin and Huxley's equations do not fall short of representing the causal structure: “There is simply no conflict between a generalization being causal and explanatory and its being an experimental generalization or being fitted to data. Experimentation is known to be a powerful way for finding out about causal relations, and this is what Hodgkin and Huxley did.” (p. 1003)

There is also a final argument in Weber 2008 that I want to overview, before passing on to Levy's take on the HH model. This argument is put against the supposition that unless the underlying

mechanisms are unraveled, a generalization cannot be causally explanatory.¹¹ Weber considers the physical model of the motion of a block on an inclined plane, and uses this to draw a similarity between Hodgkin and Huxley's conductance equations and the well-known physical generalization concerning the frictional force. The physical model describes the motion of the block under the influence of the gravitational and frictional forces. The resultant force accelerates the block down the plane with a magnitude of

$$a = g \sin (\Phi) - (\mu_k) g \cos (\Phi)$$

where g is the gravitational constant and Φ is the angle of inclination. The second summand in the above equation involves μ_k , the coefficient of friction, a constant that accounts for the physical properties of the surface. Weber underlines that this is an empirically determined magnitude, and *why* it is quantitatively so for that surface (i.e., its explanation at the level of quantum mechanics) “is not necessary to give a causal explanation for why a block slides down an inclined plane at a certain acceleration.” (p. 1005) There is an experimental generalization used in determining this magnitude; the magnitude is used in providing an invariant generalization regarding the value of a (the equation above), where we can see how the dependent variable (acceleration) would change under certain interventions. This amounts to a causal explanation (in Woodward's sense) of a macroscopic phenomenon that abstracts away from the lower-level (quantum-)mechanistic details.

Weber's point is that, similar considerations should apply to the status of the gating variables h , m and n in Hodgkin and Huxley's conductance equations: Just as the regularity represented by the coefficient of friction is taken (at this level of analysis) as a “brute fact”, so are the regularities concerning the sodium and potassium conductances (on the HH model); the three exponential functions h , m , n share the same status with μ_k in virtue of “summarizing the experimentally determined behavior of a physical object in a causal regularity.” (2008, 1005) Taking their values as brute facts simplifies the mathematical work involved, whereas overlooking the underlying mechanisms does not sever them from the causal structure.

I see Weber's interpretation of the HH model as one of the persuasive objections to a core idea found in the works of Craver and many other mechanistic philosophers: that explanations in higher-level sciences have the goal of providing the complete mechanistic details of the phenomena. This mechanistic thesis involves the risk of overemphasizing the “search” for mechanisms as if they are the only candidates for good causal explanations (of high-level phenomena). We should

¹¹ Actually the main purpose of the argument is to show how the HH model is grounded in fundamental principles (laws). But here I am concerned only with its first part.

read the thesis in connection with mechanistic realism, the stance I outlined at the beginning of the chapter, which champions an ontological commitment to mechanisms as being real features of the objective causal structure. If the mechanistic philosophers were not realists in this sense, then they would not be so much motivated to disregard “how possibly sketches” (the HH model being an exemplar) as non-explanatory. Yet, if Weber's points are right, then there are alternative ways for such “sketches” to count as explanatory. As a matter of fact, Weber does not consider the HH model to provide a mechanistic explanation—that is, not even by being a how possibly sketch. His contention is rather that the model provides a causal explanation without being a mechanistic model (see his concluding remarks on 2008, 1006). As we will see shortly, this is where Levy's interpretation departs from his. As will be quoted later, too, Levy thinks that “[a]bstracting from molecular mechanisms is a way of achieving inter-level integration.” (2013b, 20) And given that inter-level integration is a trademark of constitutive mechanistic explanation, there is a ground for one to assert that the latter is exemplified by the HH model as well—in a certain sense, at least, to be explored below.

Let me stop at this point, right before moving on with Levy, so as to make some remarks to place the present discussion in a context. I am interested in the question regarding the nature of the *mechanistic philosophers'* interpretation of mechanistic explanation (“*mechanistic modeling*”, to be elaborated). We *may* explain a phenomenon by (mentally) decomposing the system producing it into its parts, and weaving a causal story into the system by reflecting on the relations between these parts. When we do this, we provide a mechanistic explanation for the phenomenon, and we can label the configuration¹² of these parts as a *mechanism*. To be more precise, this configuration is a *causal mechanism* as characterized in the previous chapter: Its parts are the components we label as $X_1\Phi_1$, $X_2\Phi_2$, etc., which are picked because they are thought as relevant to the production of the phenomenon. This “picking” is the practice of *abstraction* that I will shed light on in the following discussion. According to Levy, the HH model embodies a kind of abstraction that mechanistic philosophers fail to acknowledge; and this failure is significant for showing a limitation of mechanistic philosophy. I do not think that this has a bearing on mechanistic explanation construed as above (I would say it has a bearing on *mechanistic modeling*). Furthermore, I do not see a need to take a stance on *whether* to interpret the HH model as a special case of mechanistic explanation (as Levy seems to think) *or* as a causally explanatory generalization which has nothing to do with unraveling the mechanism underlying a phenomenon (Weber's position). The reason being that,

¹² Hereafter I use “configuration” in contexts where I want to avoid some *mechanistic* implications. This preference will become more clear later.

explaining phenomena mechanistically is just *one* possible way to *represent* the causal structure; it does not entail the unraveling of a real mechanism in the empirical world, waiting to be discovered.

Now, how does Levy contend that there is indeed a mechanistic explanation (in one special sense) inherent in the HH model? The answer concerns the macroscopic electrical behavior. We shall recall the individual current equations, which I gave as summands in [TC'] above:

$$I_{Na} = G_{Na} m^3 h (V - E_{Na})$$

$$I_K = G_K n^4 (V - E_K)$$

These express the time-courses of the changes in transmembrane ionic currents in terms of interactions between conductance and the driving force (which is the quantity within brackets, calculated as the difference between the maintained transmembrane potential and the equilibrium potential for the specific ion). Levy then calls attention to how these summands bear the explanatory import of the HH model: They account for the contribution of each ionic current realized at the level of the whole neuron membrane; so we have the decomposition of the total current into its parts and we are provided a portrayal of how these parts come together to make up the whole phenomenon (2013b, 14).

On Levy's interpretation, it is crucial to clarify the theoretical underpinning of this portrayal (Section 5.2): It is not simply that each summand can represent a part in a mechanism for the phenomenon. This is due to the known reason that the mathematical relations do not pair with actual causal dependencies. Yet with their model, Hodgkin and Huxley were still able to relate the molecular level occurrences with the electrical behavior at the higher level. How did they do that, given their lack of knowledge regarding the membrane structure? Levy's answer is, by “aggregative abstraction” (p. 20) (to be explained shortly) which did not require the knowledge of the ion-specific channels or confirmation of the hypothesized mechanism for permeability change. All Hodgkin and Huxley needed was a perspective that let them treat (unwittingly) conductance as a property of the membrane enabled by “discrete, selective, independently acting gates: each one can be either open, in which case ions of a particular type may diffuse through it, or else closed.” (p. 14) Levy labels this perspective as the “discrete-gating picture” and he notes that although the two electrophysicists did not give an explicit statement for it, they embraced it tentatively. What he counts as evidence for this is Hodgkin and Huxley's presuppositions in the mathematical technique they used (i.e., independently behaving gating particles can be in either of the two states) and how they disqualify the hypothesis alternative to discrete-gating (i.e., the fact that they could not discern an electrical signature that would be left had there been carrier molecules) (pp. 15-16).

What is important in the discrete-gating picture is its accordance with more recent models of the action potential. Hodgkin and Huxley thought that the voltage-dependent permeability of the ions through neuron membrane was a continuous and deterministic phenomenon. Yet, as the single sodium channel recordings made in 1970's showed, each channel is a stochastic device: it randomly fluctuates between its open and closed states independently of the voltage.¹³ Still, at the macroscopic (whole-cell) level, when a *population* of channels is considered, the *probabilities* of the openings and closings are voltage-dependent (this is a generalization to be refined below—Leuchtag 2008, 5; Levy 2013b, 16), calling scientists to statistically model the behavior in question. For more than four decades, the practice (“discrete channel modeling”) has been to predict the membrane behavior by simulations of Markov kinetics of ion channel populations. The notable fact is, the averaged neural responses gained from these simulations converge to the predictions of the continuous-deterministic HH model (Clay and DeFelice 1983; Strassberg and DeFelice 1993; Meunier and Segev 2002). This is neither a coincidence nor was a foreseeable achievement of the HH model; but is interpreted as “the power of macroscopic approaches [which illustrate] the general fact that equations always contain much more than it first seems.” (Meunier and Segev 2002, 560-561).¹⁴

According to Levy, what underlies this is the discrete-gating picture that motivated Hodgkin and Huxley—more precisely, its validity in the larger scale attained by the aggregation of the behaviors of the individual channels. Hodgkin and Huxley's ignorance of the molecular details of the action potential, so to say, raised their model to a level of abstraction that is wittingly addressed by more recent approaches as well (2013b, 20).

As mentioned above, Levy's contention is that the HH model explains the macroscopic, whole-cell behavior (the total current) in terms of its constituents (the summands I_{Na} and I_K), and that—in some sense—this counts as a mechanistic explanation. His justification for this is based on the abstraction involved, where the higher- and lower-levels are related via aggregation:

[T]here is a sense in which the HH model falls short of a full explanation of the process underlying action potentials. However, in another sense, the model is a satisfactory

¹³ The randomness affects the number of open ion channels at a given instant, resulting in fluctuations in the firing time of the action potential spike. This is known as the “noise problem”.

¹⁴ Right after this statement, though, Meunier and Segev acknowledges an important limitation of the HH model, which reads as follows: “However, gating current measurements and single-channel recordings also revealed important discrepancies between the H–H model and the actual behavior of neurons: most notably, the inactivation kinetics of the Na^+ channel turned out to be voltage-independent. This, in turn, illustrates the difficulty in deriving a microscopic model from a macroscopic description, and shows that the original H–H model should be modified to fit the microscopic reality.” (2002, 561)

explanation. It answers a well-defined question concerning the whole-cell currents involved and how they arise from underlying molecular events. . . .

Here we see that, at least in some respects, an explanation that ignores underlying molecules isn't mechanistically deficient. Indeed it might be thought truer to the mechanist [mechanismic] ideal, because it explains the relationship between lower-level mechanisms and higher-level ones. Abstracting from molecular mechanisms is a way of achieving inter-level integration. (pp. 19-20)

In the context of the discussion he engages in, Levy's intention here is to expose a crucial misjudgment found in Craver's account of mechanistic explanation. We shall recall (from the previous section and the preceding chapter) Craver's evaluation of the HH model, that it at best provides a how possibly sketch of the action potential: It lacks information about the real component entities and activities involved and it leaves gaps in the explanation due to non-understood processes. Levy's argument is set directly against this verdict. He writes: "[T]he judgment that the HH model is a sketch stems, I think, from a gap in the mechanistic outlook itself, in which room has not been made for the explanatory fruits of abstracting away from structural detail." (p. 20) He further notes the widespread use of such abstraction in many biological sciences (populational sciences and evolutionary biology), where the practice is to make do on simplified descriptions of lower-level entities (e.g., genes or organisms) and to focus on (i.e., to model) the overall behavior the populations of these. This essentially requires mathematical reasoning, and such an approach (found in quantitatively-oriented enterprises) cannot be explicated justly within the scope of qualitative causal reasoning the mechanistic philosophers embrace (more on this below) (p. 22).

In Chapter 1, I have noted capturing the actual scientific practice (in high-level sciences) as one of the main tenets of the mechanistic discourse. *Prima facie*, the point just made seems to pose a challenge to this: The practice of abstraction exemplified by the HH model can yield explanatory insight about phenomena in spite of not meeting the mechanistic standards of a good (complete) explanation. A mechanistic philosopher's natural response to this could be to remind us that mechanistic explanation has not been meant to apply across the board, to cover all modeling practices in high-level sciences. So, he could accept that the causal phenomena underlain by aggregative processes may not be studied mechanistically. (As quoted on page 85 above, we have seen Craver accepting this: "Perhaps not all explanations are mechanistic.") According to his view, then, no real challenge to the mechanistic standards arises in this context. But with this legitimate remark the following becomes an immediate question to put forward: What is it that delineates the scope of the applicability of mechanistic explanation (and the mechanistic standards)?

The guiding assumption of mechanistic philosophy of science is that “mechanism” (the notion) is worthy of philosophical attention. It is an attempt to put a flesh on the scientific mechanism-talk. As a consequence, it is expected of the mechanistic philosophers to stay within the spread of the latter, and herein lies the tentative answer to the question: What delineates the scope of the applicability of mechanistic explanation is nothing other than the scientific mechanism-talk. If an explanation of a phenomenon appeals to “mechanism”, mechanistic philosophy regards it as possible that a complete mechanistic explanation of the phenomenon can be provided one day (that is, the explanation approaches its complete state as science progresses), if it has not been already. Now, biologists' talking of the *mechanism* of the action potential being a correct observation, the mechanistic philosophers' account of explanation must be valid here—and they do think that it is: This is why the explanation for the action potential is an example they refer to in their important works (e.g., MDC 2000 and Craver 2007), to show that their account of explanation reflects the practice. Does this suggest that they are wrong, given what we have seen in the above discussion about explanation by way of aggregative abstraction? In the final analysis, yes, I should say that they are wrong. For undeniably, there is an actual scientific practice associated with mechanism-talk which involves explaining phenomena with mechanisms. But these mechanisms are *unlike* the sort that mechanistic philosophers theorize about.

3.4 MACHINE-LIKENESS AND EXPLANATION AT THE TYPE-LEVEL

Let us label the *mechanistic philosophers' interpretation* of mechanistic explanation as “*mechanistic modeling*”. Mechanistic modeling asserts the mechanism to be an organized whole, composed of the parts of a system which are relevant to the production of the explanandum phenomenon. It comes in the same package with mechanistic realism in virtue of holding these parts, the mechanism components, as belonging to the real causal structure. Components are (object-involving) occurrences, such as the rushing of ions through the open ion channels in the neuron membrane, and the subsequent opening of more ion channels due to the change in voltage across the membrane. This way of representing the causal structure supposes that the causal relations between the components are localized mechanical interactions, of the sort we would observe between the parts of machines. Yet in most of the real systems that concern mechanistic philosophers (such as cellular environments, to begin with, and many other settings which involve a complex interplay of myriads of objects), global dynamics are as relevant to the outcome as the

local states of affairs. In a discussion to come, this fact will be conceptualized as “causation by concentration” (Nathan 2012). With this notion I will point out the *idealization* involved in representing the causal systems by the configurations of mechanistic components. Before doing this, though, I need to canvass the practice of *abstraction* utilized in constructing *mechanistic* models, by which mechanism components are distinguished from the irrelevant features of the investigated system. (“Idealization” and “abstraction” are debatable notions. Discussions surrounding their characterizations will be a topic in the next section. Until then, it suffices to take “abstraction” to refer to the model-building practice by which some details of the investigated system are omitted.)

3.4.1 Machine-Likeness of a System

As Levy indicates, there is a strong tendency among the mechanistic philosophers to implicitly treat an investigated system as a *machine*-like structure: “a system with localized parts, in which operations often occur in sequence, and where geometrical-mechanical properties and processes play a key role.” (2013b, 21)¹⁵ The “machine image” (Levy’s phrase) is surely an efficient cognitive tool for us to think and convey information about systems. In using the moka pot example (recall the previous chapter), for instance, I was implicating a parallelism between explanations (of behaviors) of biological and mechanical systems. This would not be as straightforward were we not accustomed to think of biological systems with the machine image in mind. That said, a quite important point is due: One must never discount the fact that this image basically functions as a mental representation. *The empirical reality that it represents might be different.* The following remark by Levy is to the point: Aggregative processes in general cannot be modeled justly as phenomena produced by machine-like systems, “because their constituents are dispersed in both time and space,” (*ibid.*) meaning that the instances of the mechanism components are not localized occurrences with a determinate sequential order. Continuing on with Levy:

Geometrical properties such as shape and size and mechanical interactions such as pushing, blocking, and conformation changing are typically of lesser importance. There is a spectrum here: some systems are more machine-like than others. The action potential, which we have looked at in depth, is a mixed case. It has machine-like aspects such as the location and role of the membrane, and, of course, the molecular structure of ion channels. It also has aggregative aspects such as ionic fluxes and overall conductance changes. (*ibid.*)

¹⁵ He mentions the emphases (in mechanistic discourse) on properties such shape, size, orientation and characterizations that bear on spatial and temporal contiguity.

I would like to enlarge upon this now, so as to contextualize the discussion to come. Dupré (2013), Woodward (2013) and Levy (2014) all note that the real systems investigated in non-fundamental sciences (especially life sciences) impose some limitations on the applicability of mechanistic explanation. There is a shared conviction that machine-like (mechanical) systems (e.g., the moka pot) are the ones that can be represented by mechanism models the most successfully; as an investigated system's character diverge from that of a machine, its mechanistic representation becomes less truthful. The above-cited authors rightfully point that this is not appreciated sufficiently in the mechanistic literature. This has implications on mechanistic realism (especially on Craver's position) which I elaborate below and in the next chapter. But a more urgent question is the following: What is it that makes a real system machine-like or non-machine-like?

Dupré addresses this question as a part of a general inquiry concerning causal explanation in life sciences. According to him, explaining biological systems by mechanisms is mostly misled for likening the former to paradigmatic machines, where the parts' causal roles can be identified “very precisely” by manipulations and removals, because “their interactions with other entities are generally specific and limited.” (2013, 29) This is, however, not the case in biological systems, where the parts are known to have multiple causal roles and enter into multiple (types of) causal interactions. (Dupré gives the example of protein moonlighting, that is, the phenomenon by which proteins—such as enzymes or ion channels—serve different functions according to their location and their level of organization.) For him, this is against the mechanistic conviction that component entities of a mechanism are robust and functionally stable; what is actually stable and robust are the processes, not the objects themselves (2013, 30). In this respect Dupré shows an inclination towards process theories of causation as more appropriate to accommodate the causal relations in biological systems.

I conjecture that a mechanistic philosopher such as Craver could respond to Dupré's criticism on the grounds that even if an object in a system is not functionally stable *tout court*, what makes it a component entity of a particular mechanism is its particular causal role in the production of the explanandum phenomenon. Whether or not this would save the *applicability* of mechanistic explanation is *not* the primary issue here, though. For even when the investigated system is not actually machine-like, it might be possible to represent it mechanistically. We have seen this by Levy's interpretation of the HH model, and I will give further examples to show how a certain mode of causal narrative, that is, mechanism models, is constructed despite the non-machine-like aspects of a system. To that extent, I am sympathetic to the wider applicability of mechanistic explanation. Yet there is a ground for raising objections to *the extent of truthfulness* of mechanism models, and it

is this ground that I am surveying.

We see that Dupré's characterization of “paradigmatic machines” is echoed in Woodward's and Levy's papers—to be more precise, the latter two authors provide more refined characterizations of machine-likeness of a system. Woodward's contribution is less direct than Levy's: He attends to the “contrast between mechanical and non-mechanical *forms of explanation*” (2013, 40 [emphasis added]), whereas Levy is more interested in the question regarding the character of real systems (not how they are represented): “The discussion so far has been directed at the question: what makes a system machine-like? This is a question about the world, so to speak.” (2014, 9) Nevertheless, even though Woodward is not directly considering the real systems themselves, he makes it clear that the form the explanation takes (mechanical or non-mechanical) depends on one's empirical presuppositions regarding the system at hand which are to be evaluated against how much machine-like the latter actually is (2013, 64). Therefore we can contend that in Woodward, too, it is “a question about the world” at the end of the day. A last thing to note is, machine-likeness is understood as a graded notion by both (recall the quotation from Levy above): it is not that a system is either machine-like or not, it is just that some systems are more machine-like than others (and accordingly, in Woodward, “some explanations are more 'mechanical' than others (in various respects).” [2013, 41]).

What is the criterion here, how should one evaluate how much a system is machine-like? To this, Woodward responds by using the notions of *stability*, *modularity* and *fine-tunedness* (or *organization sensitivity*) (2013, Section IV), and Levy suggests us to look at the *orderliness* of the system with respect to the parts' functional (qualitative) distinctness from each other and with respect to how much the parts play their roles in virtue of the *local* relations they enter (2014, 4). Ignoring some details,¹⁶ I can judge that there is not much difference between the two approaches on offer here, at least with regards to what concerns us. While examining Weber's take on the HH model above, we came across Woodward's notions of stability and modularity (Section 3.3, page 88). In that context, they appeared as the two requirements a generalization must satisfy so as to provide causal explanation. Now the same notions are invoked, but this time it is for evaluating the success of a mechanistic explanation in representing the system. Shortly put, Woodward asserts that (*i-*) the stability of causal relations among the components of a mechanism (i.e., the intervening links) with respect to that of the overall input-output relationship, and (*ii-*) the modularity exhibited

¹⁶ For instance Levy does not agree with Woodward on stability being a sign of machine-likeness (2014, 7). I also have doubts at this point, mostly due to the discussion in the next chapter, in Section 4.1. But I should admit that I have not thought on this extensively.

by a representation (which concerns the stability of a set of causal relations when other sets of causal relations are changed) count towards the success of a mechanistic explanation (2013, 50-51). Here, it is not the notions (of stability and modularity) themselves but what they account for that is significant (Dupré's, Woodward's and Levy's accounts all converge on the same ground, when seen from this angle), and it is nothing other than the precise differentiation of the (sets of) parts from each other: That is, machine-like systems are the ones that can be mentally decomposed according to the local interactions between the qualitatively/functionally different but spatiotemporally proximate parts. But there is also a sense in machine-likeness that lets one consider the components as parts of a whole, or in other words, things whose separate functions constitute an overall behavior of the system. It is this sense that Woodward intends to capture by the third notion of fine-tunedness/organization sensitivity (cf., Levy 2014, Section 3.4):

[T]his is the idea that it is characteristic of many machines that the specific details of the causal properties possessed by the components, and the way in which those components are connected to each other, including their spatio-temporal relations, matter a great deal to the outcome produced, in the sense that if these features or organization were different, the outcome would be different. (2013, 56)

Let us look at a paradigm case of a non-machine-like system: at the example of diffusion in cellular contexts (taken from Levy 2014, 4). It is non-machine-like because the parts of the system (particles moving in a medium with a concentration gradient) are not functionally different from each other, and the organization is not sensitive to what the parts are or to the spatio-temporal relations between them. Accordingly, modeling the phenomenon of diffusion rests on considering the aggregative behavior of many particles. It starts with how a typical single particle behaves in a medium with a concentration gradient. It then generalizes this to the population of particles, without attending much to the structural, qualitative or spatio-temporal details about them. Therefore it should not come as a surprise when we realize that there is no mechanistic explanation offered for diffusion: The phenomenon undebatably concerns non-machine-like systems and we rely on versions of *quantitative* modeling in explaining them. (See Section V in Woodward 2013 for more examples to non-machine-like systems.)

It is the *mixed cases*, systems having both machine-like and non-machine-like aspects, that are of interest for discussions regarding mechanistic philosophy. Starting with the phenomenon of the action potential, with all the examples I consider below, I intend to make the following point about such cases: The relevance of the components of mechanisms (to the explanantia) is, by definition, construed in terms of the repeatable *mechanical* contributions they make in the system—by likening

them to the contributions of the parts of a machine-like system. The first step in this construal is, then, distinguishing the actually relevant factors in the real systems from the irrelevant ones. (Relevance is, of course, context-laden, and depends on how the explanandum is identified.) The relevant factors can be picked by the criterion of difference-making, and here, Woodward's manipulationist theory of causation can be used (as Craver advocates). This picking constitutes the practice of *abstraction* where the irrelevant factors are said to be abstracted away. I will state that claims about the relevant factors must be made, not at the token-level but at the type-level. This is how the repeatability of the explanans is assured: The token, concrete occurrences may not be repeatable, therefore they are not explanatory in themselves, but they are explanatory for being instances of the components that must be set at type-level. Then, the second step for determining the components is excluding some of the *relevant* factors for their being parts of the non-machine-like aspects of the system. They are excluded, because in a mechanistic model of the system, the components can only account for the machine-like aspects: that is, by definition, the components can only represent the actual difference-makers that make local, qualitative causal contributions. Thus I will be saying that there are actual difference-makers that are *not* represented in a mechanistic model: they are *idealized away*, so to say, for the sake of the attaining simplicity in the causal narrative about the system. (To be clear: The assumption here is that mechanistic models simplify the systems by focusing only on their machine-like aspects.) In the final analysis, mechanistic models—just like other scientific models—involve distortions in their narrative about the real systems: Distortions due to *abstracting away* some occurrences for not being found relevant, and distortions due to *idealizing away* some other occurrences which are actually relevant but cannot be represented in the mechanistic causal narrative. Because of these distortions, how much the models *truthfully* represent the reality becomes questionable.

3.4.2 Type-Level Analysis

Let us think about the *non-machine-like* aspects of the action potential. As the discussion above has hinted, in the depolarization phase the opening of an individual sodium channel is *not* the causally relevant component activity in the production of the explanandum phenomenon.¹⁷ The conformational change of the channel is a random event and unless enough number of channels

¹⁷ Similar remarks can be made regarding the repolarization phase of the action potential and the role of the potassium channels.

open, the threshold potential (the critical value beyond which the membrane potential immediately changes as more and more channels open) cannot be attained; therefore no depolarization occurs (the action potential is not initiated). Thus the actual difference-maker is the opening of the *population* of channels. In explanations of the type the HH model exemplifies, what aggregative abstraction achieves is picking difference-makers of this sort.

It might be reasonable to assume this population as something behaving like a component entity in a mechanism. In fact this is the assumption made when we give the mechanistic explanation of the action potential: We (the scientists and their informed readers) say that sodium *channels'* opening cause the ion influx.¹⁸ By this we refer to their aggregative behavior. We ignore whether particular channels open or not, we abstract away whatever happens below the population level. Put more properly, in considering the population level, we are concerned with the actualization of the enough number of channels' opening. This can be seen as an occurrence that is preceded by others where there are indeed many open channels, but the number of them is not enough for enabling ion movements across the membrane with a considerably higher rate. In other words, asserting the opening of the *population* of channels as the actual difference-maker entails that there is not a difference (of significance in the context of the investigation) unless a *number* N of channels are opened. And this being so, we can speak of a contrast class here which is made up of numerous prior occurrences of n channels' opening, where n can be any number smaller than N . Were we to take the difference-maker simply as the ion channels' opening (without any mentioning of the population), the contrast would look as though it was set between their opening and not opening (in which case, the contrast class would have only one element, that is, 0 channels' opening, or the state of all channels being closed).

These are worth to highlight due to the stochastic nature of the activation-inactivation (opening and closing) of the ion channels, and the promotion of the inactivation by activation. If the channels worked deterministically and their inactivation was a direct response to the rise of the voltage inside the neuron (as Hodgkin and Huxley supposed),¹⁹ then the case at hand would be similar to one of the causal steps in the mechanism for the making of coffee in a moka pot. We shall see in which respect the ion channel dynamics differ from this step in the moka pot example (that is, in which

¹⁸ Or, we say in the similar vein that the *potassium* channels' opening causes the ion *efflux*. In what follows, I will not be concerned with channels specific to particular ions, but ion channels in general (and ion *movement* to stand for both the influx and the efflux). See also the subsequent footnote.

¹⁹ As indicated in an earlier footnote, Hodgkin and Huxley were wrong in thinking that the sodium channels' inactivation was directly related with the voltage inside (more on this below). Although *potassium* channels inactivate with voltage and so their dynamics is different than that of sodium channels, here I am extending the claim to their case, too. This implies an incorrect generalization, but it will simplify the text of the argument for us.

respect the former has non-machine-like aspects).

Recalling our example from the previous chapter, the water in the bottom chamber of the pot is pushed through the ground coffee in the basket due to the steam pressure built up above its surface. Accordingly, the difference-making component in this case is the attaining of a quantitatively determinate amount of pressure P . Values of pressure lower than P are insufficient to push the water, and their attainment make up a contrast class for different values of p (p is the pressure-related variable numerically less than P). If we assume that an occurrence corresponds to each p (i.e., the attainment of p), then as the pot is heated, each one of these occurrences come right after another. The thing to note is, the occurrences follow a continuously increasing function of time of heating: There is no fluctuation in values of p , which deterministically increase as the pot is heated continuously. When the threshold of P is attained in the chamber, we are able to talk about *the steam's pushing the water* (upwards through the basket) as an actualized component (an event, or an occurrence) in the relevant causal mechanism, for the making of coffee.

Now, back to the ion channels of a neuron: The number of open channels is related with the rise of the voltage across the membrane, but the change in the voltage *indeterministically* causes the channels to open and furthermore, the channels may close *independently* of the rise in voltage they contribute to. I have already mentioned the stochasticity involved in the former. Concerning the independence claim, on the other hand, the following quotation will be helpful:

In this model [the HH model], the inactivation of [sodium] channels is governed by a separate voltage-dependent process than that governing the activation of channels. In fact, various experiments have shown that inactivation derives its voltage-dependence from the same voltage-dependent steps as activation. In modern models for sodium channel gating, inactivation is viewed as an inherently non-voltage-dependent process that is promoted by activation because activated channels can more readily undergo inactivation than non-activated channels. (Bean 2007, 462, Box 3)

These points indicate that the number of open ion channels randomly fluctuates: A particular channel may open or not (although more of them tend to open when the voltage rises), and the particular openings can cause particular closings. In virtue of this, the fluctuation in the number corresponds to the actualization of elements of the contrast class (occurrences). This is to say, there are innumerable occurrence sequence *tokens* that can instantiate the occurrence sequence *type*,²⁰

²⁰ I could also use “process” instead of “occurrence sequence” (then I would be speaking about process tokens and process type). But the loose dependence of one occurrence on the previous one in the sequence (say, the opening of the 500th channel after the opening of the 499th channel) made me prefer “occurrence sequence”, which is a less restricted notion when compared with “process”. I will be more comfortable with using the latter in considering the moka pot example below, because in that case, the attaining of each value of p have a certain dependence on the previous occurrences.

which begins with the first stimulus on the ion channels and ends with the sufficient number of channels' opening (i.e., the actualization of a component of the depolarization phase in the case of sodium channels, for instance). And this is the point where this example becomes dissimilar to the process happening in the collecting chamber of a moka pot. In the latter example, each of the occurrences making up the contrast class (of the actual difference-making occurrence, the attaining of *P*) is actualized *once and for all*. Hence, the type of process is instantiated in the same way, anytime a moka pot (with similar properties) is heated: A token of the process is similar to another, with respect to the history behind the attaining of *P*. In contrast to this, the opening of *N* ion channels is a multiply realizable occurrence sequence—and the multiple realizability here is in two senses: Both the final configuration of the open channels and the sequence of occurrences that leads to any particular configuration are multiply realizable. A particular sequence may not be repeatable (or actualized) in another occasion, so to say.

Several important observations are due. To begin with, the relevant mechanistic explanation provides a type-level analysis of the causal relation between ion channels' opening and the subsequent rapid change in the voltage across the membrane. What is involved in this is the abstract description of occurrence sequences leading to the opening of enough number of ion channels. It is of utmost importance to distinguish this abstract description from these instantiations. If mechanisms are to explain phenomena, they must have multiple occurrences. Otherwise any chain of causally connected occurrences would have to be counted as a mechanism—obviously neither the scientific mechanism-talk nor the mechanistic philosophy built on this is this much liberal. (This point will be considered in detail in the next chapter.) From these, we see that a token sequence of occurrences, which instantiates the type-level abstract descriptions, cannot be a part of mechanisms that explain phenomena. For a token sequence itself may not be repeatable, as the example of the opening-closing of the ion-specific channels illustrate.

I join the mechanistic philosophers in thinking that these sequences are located within the causal structure, and that in virtue of this they provide the real bases for the mechanistic explanations to be about the objective world. However, explanation by mechanisms is made possible by abstracting away from the details of token sequences, to the end of handling them with respect to their repeatable features, the common outcomes. Considering the same example, a particular sequence of the opening of ion channels (assuming that it is a “successful” one in the sense of meeting the threshold number *N*) is similar to another for causing a rapid change in the voltage. It is this causal role itself *and not how this is realized* that is of concern in the mechanism for the action potential, because it is the causal role that is repeated in every actualization of the

relevant component—the component being the *type-level, abstract description* of the occurrence of the opening of population of channels.

That said, I want to emphasize how this view on offer is different from the mechanistic philosophers'. On their perspective (mechanistic modeling), what scientists do by engaging in mechanism-talk amounts to singling out the features of the system that are found relevant to the production of the explanandum phenomenon (i.e., *abstracting* away the irrelevant parts). They view the relevant parts of the system as the mechanism components that are causally organized. A token components-plus-organization, for them, is a token (causal) mechanism explaining a token phenomenon. Instances of the same mechanism type explains (obviously) instances of the same phenomenon type. But this does not imply that the phenomenon type is to be explained by one specific mechanism type. This is because of the fact that high-level phenomena (that can be subjects of mechanism-talk) are known to be *multiply realizable*. Yet this does not pose a problem for the mechanistic philosopher, for whom mechanisms explain how a phenomenon is produced *when* it is produced, and “[a]ll that matters is that the phenomenon is realized by some underlying mechanism.” (Craver 2007, 110)

It is not required that all instances of ψ -ing [the explanandum phenomenon] be explained by the same underlying mechanisms. What matters is that each instance of ψ -ing is explained by a set of components that are relevant to ψ in that particular mechanistic context. There are no doubt epistemic difficulties of determining when two mechanistic contexts are equal, but there is no conceptual difficulty seeing how the same type of phenomenon could be explained by different components in different contexts. (Craver 2007, 160)

This is the mechanistic philosopher's take on the issue of multiple realizability. However, the comparison I draw above between a biological system and a machine-like artifact demonstrates that, concerning mechanisms there is another aspect to multiple realizability. Besides the investigated phenomenon (type), the components of a mechanism type are multiply realizable as well. On first reading, this should not be a news for the mechanistic philosopher at all. Because, after all, he holds mechanisms to be in a nested hierarchy; one component in a mechanism can be an explanandum phenomenon explained by the lower-level mechanism; and given that he himself asserts the investigated phenomena to be multiply realizable, there is nothing surprising in a component's being multiply realizable, too. But here is the crucial point: Multiple realizability concerns how an occurrence is realized which has no bearing on *if* this occurrence is to be counted as a mechanism component or not. Occurrences are individuated as mechanism components on the condition that they are thought relevant to the explanandum phenomenon (as mentioned above); leaving out the irrelevant occurrences constitutes a known and uncontroversial example for the

abstraction practice in model-building. There is yet another facet of abstraction involved in modeling systems mechanistically, and that one is about omitting how individual components are brought about. Multiple realizability of occurrences is relevant to this latter sort, but even when multiple realizability is not an issue, mechanistic explanations appeal to it. The abstract description *vs.* instantiations distinction underlined above is meant to bring this sort of abstraction under light.

The claim I make in connection with this point can be summed up as follows. Mechanism models involve abstract descriptions of occurrences relevant to the explanandum phenomenon. The abstraction I have just brought under light enables the modeling of the *repeated* features of the investigated systems actualized in every instantiation of the explanandum phenomenon.²¹ This is done by making type-level claims concerning the difference-makers (the token occurrences do not count as difference-makers, especially when the system has non-machine-like aspects). According to the mechanistic realist, the repeated features are the token components construed within a causal organization, yet in what follows I will voice my objection to this.

I began the section by focusing on one component in the mechanism for the action potential and comparing it with a component in the mechanism for making of a coffee in a moka pot. The comparison was important for seeing the respect in which the instantiations of component types differ in the two cases. But what is more crucial is to distinguish component types from the concrete occurrences they are about. It is the former which are explanatory, because it is at the type-level that the repeatable patterns of the world are handled. (This applies to the moka pot example as well, but there, the abstraction involved at the type-level is not as striking.) The point made here for one component will be generalized for the whole mechanism shortly: Following Overton (2011) I will claim that mechanisms are abstract configurations of types. (Accordingly I will point out that the supposed dichotomy between mechanisms and models of mechanisms is a false one.) This will be one step towards “detaching” the mechanisms from the causal structure they represent. The next step will be showing that the causal relations between mechanism components *are not* instantiated the way they are described by the models. For this, I will draw on Nathan's work on *causation by concentration* (2012). Nathan asserts that concentrations of substances in a system have causal roles

²¹ This should be distinguished from the abstraction by which only the relevant occurrences are picked and the irrelevant occurrences are left out of the model, without attending to type *vs.* token distinction. For instance, not all of the heat given to a moka pot can be used for boiling the water in the bottom chamber; there is conduction loss, so some heat is given away. In explaining the making of coffee, we stay silent with regards to this loss: Although it is an occurrence within the system, it is not relevant to the phenomenon explained. It does not count as a component, *not* because it is not a repeated feature of the system (in fact it *is* repeated anytime a moka pot is put on a stove), but because the loss is usually negligible when compared to the heat absorbed by the water in the chamber (which is enough for it to pass through the ground coffee in the basket).

which are irreducible to actual interactions between concrete objects. This entails that not all of the actual difference-makers modeled by mechanisms (especially those provided in biological sciences) are mechanistic entities engaged in mechanistic activities, and relatedly, mechanism models are highly idealized (cf., Love and Nathan *forthcoming*). My argument against mechanistic realism should be justified on these grounds.

3.5 ABSTRACTION, IDEALIZATION, AND MECHANISMS AS MODELS

The main thesis of Overton 2011 is one that I will endorse with a disclaimer. Quoting from the abstract of this paper, the thesis in brief is: “Mechanisms have an ontic status distinct from the causal chains of token entities and token activities that instantiate them.” (2011, 941) I am not much sympathetic to the idea that mechanisms are instantiated by token entities and token activities, strictly speaking—but let this wait for the time being, until the last section in Chapter 4. Apart from this, I agree with Overton that mechanisms are abstract while the particulars that instantiate the mechanisms (i.e., causally organized occurrences) belong to the furniture of the world. Two notes must be made straightaway:

- Overton does not make a claim regarding whether mechanisms as abstractions (or the types they appeal to in this context) are real or unreal; his only claim is that there is an ontological divide to note between mechanisms (the component types involved) and their instantiations: “Whatever ontic status one wishes to ascribe to abstractions, mechanisms are distinct from causal chains because they abstract away from properties such as the length of the chain.” (p. 943; see also p. 948—the content of the abstraction is elaborated below.) Later (in Section 4.4), I will formulate this ontological divide in terms of an existent being fundamental or derivative.

- We have seen that in mechanistic philosophy and in the literature built around it, mechanisms and models of mechanisms are held separately. On the approach that I advocate in accordance with Overton's claims, though, there is no such dichotomy between mechanisms and models. For mechanisms are nothing other than the abstract and idealized models of concrete occurrences. This, in turn, is because the label “mechanism” must be reserved for what explains the investigated phenomenon type. Thus I take (causal) mechanisms to be a subset of scientific models that aid causal reasoning. The causal narrative that these models provide can be conveyed visually by diagrams (e.g., Figure 2-3 in the previous chapter, Figure 3-1 below). This may give the

impression that the diagrams themselves are the mechanisms, but this is not what I am saying. As any mechanistic philosopher would say, too, such diagrams should be taken only as representations of mechanisms. But in opposition to a mechanistic realist, I hold that the causal narrative is not truthful in the sense envisaged by him: While it is *about* the objective causal structure, it involves significant deviations from the actual state of affairs.

Mechanistic philosophers suppose that given the explanatory context, similar token organization of components are found to be responsible whenever the explanandum phenomenon is realized. It is this supposition that leads them to label the token organization as “mechanism” (that is, the explanans). Yet as it is suggested both in the previous discussion of the opening-closing of the ion channels and in the ones to come, the token organization may not be repeatable—and still this is no hindrance for the explanation, because mechanistic explanation is set at type-level and which token organization instantiates the mechanism type (the model) in a particular occasion does not have explanatory worth. Hence it is not the repeatability of the token organization that matters; rather, a sufficient deviation from the facts ensures that the epistemic and manipulation-related demands of the context are met.

There are two known and studied ways by which scientific models deviate from the real systems they are about: omission of details and misdescription of the real states of affairs. It is typical for models to involve both, while it is also true that (at least from a certain perspective) an omission is a misdescription anyway. Probably because of these, we observe different philosophers describing and treating them according to different criteria, and sometimes under different labels. In what follows, I adopt the framework found in Jones 2005 and Godfrey-Smith 2009, where omissions are *contrasted* with misrepresentations, and the labels associated with these are *abstraction* and *idealization*, respectively. To me, this seems to offer a both more general and fine-grained analysis than those found in Weisberg 2007 or Strevens 2008, for example. Let me first consider Weisberg's and Strevens' use of these labels, to acknowledge the multiplicity of views concerning the issue.

In Strevens' “kairitic” account of causal explanation, “abstraction” stands for *all* procedures by which the details of causal processes are stripped away (2008, 97). “Idealization” is one such procedure (alongside “preidealization” and “mathematical explanation of physical phenomena”, it is a *type of abstraction*) (p. 297), and it is defined as “*any* misrepresentation of the causal process that improves the explanatory power of a model by comparison with its veridical counterpart.” (p. 300 [emphasis added]) So defined, idealization involves *both* the *omissions* and the *distortions* (of the natures) of some elements (p. 297) for the sake of developing causal models which appeal only to

the difference-makers. If a cannonball is thrown to break a window, its exact mass is not to be included in the causal model that explains the breaking: For it is not the ball's being, say, 10 kilograms that made a difference, but its mass being *between a and b kilograms* ($a \leq 10 \leq b$; pp. 96-97). Strevens uses this as an example for abstraction, and from the later discussions in his book, we can infer that it provides a case for idealization, too. Here, we observe a clear *omission* of a detail. A case for an idealization due to *distortion*, on the other hand, can be found in several assumptions of the ideal gas model that Strevens examines closely in relation with Boyle's Law²²—e.g., the gas molecules (in a closed container) do not collide with each other, when molecules hit the container wall the collisions are perfectly elastic, etc. (pp. 308-309) These are known to be false assumptions, but they certainly simplify the investigated system. Much more importantly (for Strevens' account), idealizations point at the *non*-difference-makers: “the role of an idealization is to assert the explanatory irrelevance, that is, the failure to make a difference of a salient causal factor.” (p. 318) Note that Strevens states this for all idealizations, be them omissions or distortions. In fact, from his point of view, distortions of the nature of the investigated system might perfectly be seen as omissions, too: Assuming the gas molecules to be not colliding with each other is *omitting* the collisions between them, after all; or saying that the collisions with the container wall is elastic amounts to *omitting* the loss of energy, etc. Therefore there is basically no difference in omissions and distortions with respect to *their role* in Strevens' account of explanation built on causal difference-makers.

This is idealization conceptualized in the service of a particular account of explanation, and it is examined in Weisberg's paper as an exemplar for one kind of idealization (2007). The aim of this paper (entitled “Three Kinds of Idealization”) is to explicate the different *practices* that go under the name “idealization”, which for Weisberg means “the intentional introduction of distortion into scientific theories.” (2007, 639) Being a pluralist on the nature of idealization (p. 644, fn. 14), Weisberg is more interested in the epistemic goals associated with idealizations than a fine-grained systematization of how models deviate from truth. Accordingly he is not much concerned with the terminology itself; a “distortion” is taken as any deviation, it is not set apart from “omission” because distinguishing these two (or idealization from abstraction—see *ibid.*) has no implication on Weisberg's project. On his approach, Strevens' account exemplifies “minimalist idealization” for having the end result as a minimal causal model where only difference-makers count. It is not expected of such model to change as science progresses, because accuracy is not sought after in this

²² This is the law which states that for a fixed amount of gas in a closed container kept at a fixed temperature, the volume and pressure properties are inversely proportional.

practice of idealization (p. 645). Not so with “Galilean idealization”, which is another kind of idealization where rational tractability of the system encourages a stepwise approximation to truth, by the removal of the distortions introduced previously. This experimental practice is discussed in depth in McMullin 1985 (the label Weisberg adopts is due to this paper), with the example of Galileo envisioning a medium devoid of air resistance on free-falling bodies. The scientist acknowledges that he “lack[s] such a space”, so he first devises a setup that best resembles the idealized system, observes what happens in this setup, and then compares this “with what happens in others less thin and more resistant.” (quoted in Weisberg 2007, 641) Weisberg notes that this experimental technique is practiced by, for instance, today's computational chemists who study molecular properties by approaching gradually to the exact solution of the Schrödinger equation. So Galilean idealization is typified by a progress in the idealization procedure: Unlike those in minimalist idealization, the constructed models are expected to get better in representing the real systems.

While minimalist and Galilean idealizations differ with respect to this feature regarding progress, they share the goal of constructing the single best model. This is what separates the two from the last kind of idealization Weisberg reviews, “multiple-models idealization”, or MMI in short (2007, 646). Weisberg describes MMI as “the practice of building multiple related but incompatible models, each of which makes distinct claims about the nature [of] and causal structure giving rise to a phenomenon.” (p. 645) Models generated through this approach involve different sorts of deviations from the investigated complex systems; analyses and predictions are based on the vision formed by their joint utilization compensating for the tradeoffs in each. One of the examples Weisberg gives for MMI is the method the United States National Weather Service adopts in predicting the global weather: Having seen that no single model is successful enough, the Service brings together the outputs of three primary forecasting models.

The approach²³ found in Jones' (2005) and Godfrey-Smith's (2009) papers differ from the just reviewed ones especially due to their aiming for a fine-grained analysis. Both authors state their intention to capture the distinguishable senses of the practices of abstraction and idealization in scientific contexts, and what is achieved by their respective utilization. They characterize them as follows:

An *abstract* description of a system leaves a lot out. But it is not intended to say things that are literally false. An *idealized* description of a system is a description that *fictionalizes in the*

²³ Both authors make the same points, this is why I treat their papers together.

service of simplification, in the way described above [i.e., by providing simpler analogues of real-world systems]. (Godfrey-Smith 2009, 48 [original emphasis])

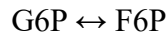
The term 'idealization' applies, first and foremost, to specific respects in which a given representation misrepresents, whereas the term 'abstraction' applies to mere omissions.

. . . [A]bstractions involve omission *without* misrepresentation. Omission in this restricted sense is, so to speak, a matter of complete silence. (Jones 2005, 174-175 [original emphasis])

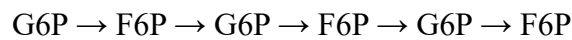
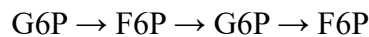
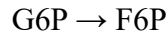
It must be added that on this construal, in a given model, each of these concerns different features of the modeled real system. This is to say, *no* single feature of the system can be subject to *both* idealization and abstraction (Jones 2005, 175). Recall the ideal gas model Strevens reflects on: We said that according to Strevens' kairetic account, the *idealization* that states that the gas molecules do not collide with each other can also be put in terms of the *omission* of the collisions (their being *abstracted away*). This, however, would not be a legitimate use of the notions on the regimentation offered by Jones and Godfrey-Smith. On their conceptualization, the no-collision assumption is definitely an idealization because it misrepresents (fictionalizes) the system for the sake of simplification. The omission involved is not “a matter of complete silence”; it rather comes with the assertion that no collisions take place in the model (true in the model, false in the world). This is to be contrasted with the omission of the irrelevant occurrences in a system in building a mechanistic model. The *inclusion* of some occurrences in the model as mechanism components implies nothing about the excluded occurrences (except that they are found irrelevant to the explanation). Here we have a case of abstraction.

As indicated before, in the present study I adopt this approach to idealization and abstraction. The approach will be helpful in conceptualizing what is wrong in mechanistic modeling, because it offers a more clear understanding of in which respects the models differ from the systems they represent. I can now return back to Overton, to see the example he gives for abstraction in mechanism models.

Overton considers the phosphoglucose isomerism step of glycolysis (2011, Section 2.3). This is a case of *reversible isomerization* where molecules change their structure back and forth, and for many times, instantiating *isomer types*. (Isomers are molecules with the same chemical formula but with different structure.) The isomers of concern in glycolysis are glucose 6-phosphate (G6P) and fructose 6-phosphate (F6P). The same enzyme (glucose-6-phosphate isomerase) catalyzes the isomerization of G6P to F6P, and vice versa. This reaction is shown by a double-sided arrow as below:



What is important for us is, this is an expression of a type-level reversible reaction: G6P and F6P stand for molecule types and the arrow stand for numerous different causal chains that can instantiate the phosphoglucose isomerism step. Overton notes this and puts emphasis on the length of the respective causal chain which can vary from one instantiation to another. Below are the possible ways he mentions where a G6P token is ultimately converted into a F6P token (2011, 947-948):



He proposes “that each chain is a token instance of the same phosphoglucose isomerism mechanism.” (p. 947) He questions if it would make sense to see in each of these chains instantiations of different mechanisms, such as one concerning the conversion of G6P to F6P and another concerning the conversion of F6P to G6P. (In such case the latter mechanism is not instantiated in the first chain.) But as he remarks, the phosphoglucose isomerism step is a “submechanism” in a larger mechanism, namely, glycolysis. What matters in glycolysis is the conversion of G6P to F6P; the reverse reaction does not bear an explanatory worth (it is not relevant), so it cannot be considered as a different mechanism. To aid visualization, here is a schema of the metabolic pathway (glucose is converted into pyruvate and there is a gain of energy in the form of ATP; some steps of this are omitted):



Due to the points just made, the metabolic pathway is different from the causal chain it represents: It leaves out the possible token reverse reactions at the phosphoglucose isomerism step (as they do not make a difference to the production of pyruvate in the end, as long as enough number of F6P's are present), making it an abstract model. Now the question to consider is, whether to identify the metabolic pathway (with the addition of the steps I omitted above) as the glycolysis *mechanism*. The inquiry amounts to asking where to draw the line between mechanisms, mechanism schemata and mechanism sketches. For the latter two also involve omissions, be it deliberate or due to epistemic limitations. MDC defines mechanism schema as “a truncated abstract description of a mechanism that can be filled with descriptions of known component parts and activities”, and a sketch as “an abstraction for which bottom out entities and activities cannot (yet)

be supplied or which contains gaps in its stages.” (MDC 2000, 15 and 18) These definitions are supposed to serve the need for making sense of a scientific progress, witnessed in improving the mechanistic explanation for particular phenomenon. (As indicated before, Craver's work is specifically designed to explicate the normative aspect in this progress.) Nonetheless, Overton finds it a mistaken idea that there is a linear progress (according to which, schemata are improvements on sketches, and as the former get better they approach complete descriptions of mechanisms),

since the three cannot all lie on a spectrum: [according to MDC's definitions] a mechanism is an arrangement of entities and activities, while a mechanism schema is a description, and a mechanism sketch is an abstraction. For instance, glycolysis [*sic*] would be a mechanism in molecular biology, a mechanism schema in quantum chemistry, and a mechanism sketch in quantum field theory. We would like “glycolysis” to refer to the same thing in all three cases, but surely we cannot change the ontic status of glycolysis simply by changing the context. (2011, 946-947)

This remark is meant to display a confusion on the part of mechanistic philosophers, but unfortunately, it seems to me that here Overton himself leads his reader to a confusion.²⁴ Resolving this should make us see his point more clearly—a point I agree on with Overton, despite the confusion created. The confusion concerns the explanans and the explanandum of glycolysis. In his paper Overton appears to make the mistake of using “glycolysis” for both of these, but the term should be reserved for the explanandum phenomenon investigated at the level (of organization) molecular biologists are engaged with. When we restrict the use of the term to a certain level, though, it becomes natural to ask *why* “we would like 'glycolysis' to refer to the same thing in all three cases” (the three different levels of scientific inquiry Overton mentions). This is the question a mechanistic philosopher might want to ask to Overton—in particular I have Craver in mind, who thinks levels of mechanisms can provide a layered model of the empirical world (2007, Chp. 5). Accordingly, for the mechanistic philosopher, it cannot be the same mechanism that is located at all these levels—so why should anyone think that “glycolysis” refers to the same thing?

My answer to this is, it is *not* that Overton thinks the same *mechanism* is located at all three levels. (The following is *my* reconstruction of Overton's line of thought expressed in the quotation above.) Given the main claim in his paper (i.e., the ontic statuses of mechanisms and the causal chains instantiating them are different), he must be saying that it is the same *causal chain* that is investigated by a molecular biologist, a quantum chemist and a theoretical physicist. But what differs in their studies is the degree of zooming in on this chain. That said, there may arise yet another question regarding whether *the whole phenomenon of glycolysis* can be explained at levels

²⁴ I thank Maria Kronfeldner for making me notice this.

lower than that concerning molecular biology, with more and more fine-grained models providing a more detailed picture of the causal chain. A mechanistic realist philosopher must give a positive answer to this, for thinking that models of mechanisms get better and better as scientists discover more about the organization of occurrences. Yet this entails that the explanans at the level of molecular biology is more abstract (and possibly more distorted) than an explanans at the level of quantum chemistry, and looks like a sketch made up of black boxes and filler terms standing for “processes-we-know-not-what” (term borrowed from Craver) at the level concerning theoretical physics. Obviously this must be bad news for the mechanistic realist, because the implication is that the explanans of glycolysis in the hand of the molecular biologist is not the mechanism for glycolysis, it is a schema in one context and a sketch in another.

I believe this exposition makes Overton's point more clear, the point being that there is no non-arbitrary way to distinguish mechanisms from schema and sketches. For the same system can be modeled in different contexts (different levels of inquiry) at different degrees of abstraction, but it is not plausible to think that scientists from different domains all aim at reaching a single complete description (of mechanism). Instead, different domains impose different norms to mechanistically explain the phenomena, and most importantly, criteria for the incompleteness of a model is inherent to the context of investigation—it is not domain-independent. This is also to say, achieving the goal of providing a complete *mechanistic* model does not capture the scientific practice well, where sketches and schemata may well be quite sufficient for epistemic and pragmatic purposes. To return to the question of regimenting mechanisms and the other abstractions, Overton's proposal is to see all as organizations of abstract types, differing with respect to the epistemic/pragmatic aspects of the context. Mechanisms set idealized start and termination conditions for the phenomena they explain (MDC 2000, 11-12; Overton 2011, 945), they are abstract organizations of types; whereas mechanism schemata are second-degree abstractions of mechanisms by leaving out much more of the real causal story underlying the investigated phenomena:

It seems that the main difference between a mechanism and a mechanism schema is the bottom-out conditions, which depend on context. A mechanism sketch cannot (yet) be filled out by any bottom-out conditions accepted in the context, but we can include it as another abstraction alongside mechanisms and mechanism schemata. (Overton 2011, 948-950)

In sum, with Overton's evaluation of glycolysis, we place token causal chains on one side and mechanisms and the related abstractions to another side, with respect to their ontic statutes. (To be fair let me remind that Overton only distinguishes them according to type *vs.* token distinction. After considering the next example, I will assert that the divide is between real causal organizations

of tokens and idealized configurations of types.) It would be wrong to think that this point can be made only with cases like glycolysis, that is, phenomena which involve reversible reactions. The mechanism - causal chain divide can be made visible by other examples, too. Let us take the example in Love and Nathan (*forthcoming*): the RNA transcription mechanism explaining eukaryotic gene expression. With this, we will also proceed swiftly to Nathan's articulation of causation by concentration.

Expression of certain genes is regulated by the proteins binding to certain nucleotide sequences on the DNA. When a protein called the “transcription factor” binds to the promoter sequence (called the “operator”), it mediates the binding of the RNA polymerase to a further sequence (the gene) on the DNA. The latter is used as a template by this enzyme for the synthesis of the RNA chain. The following is an illustration of these steps (which is a diagram showing the mechanism for gene expression). As usual the arrows represent the causal relations.

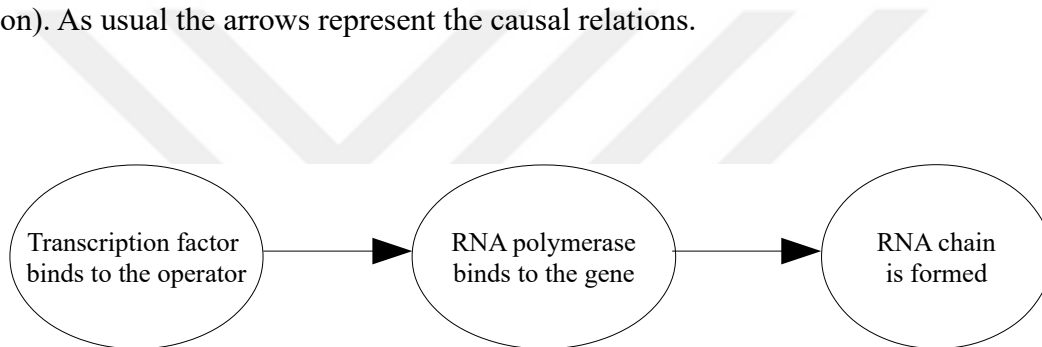


Figure 3-1. RNA transcription mechanism

Love and Nathan present two pictorial depictions of the same mechanism, which differ with respect to the amount of detail included (*forthcoming*, 5-6). The more detailed one (their Figure 3) specifically emphasizes that the transcription factor is not a single molecule, but a complex functional unit made up of many molecules instead. Their other figure (Figure 2) lacks this very detail. Nevertheless, according to the authors, the compromise in the more sketchy figure does not imply an explanatory deficit. This is the same point I have made within the present chapter, that the sketchiness of a model is not a defect in the causal/mechanistic explanation involved. On the other hand, Love and Nathan further indicate the following, which is especially relevant to the type-token distinction: The token occurrence of the binding of the transcription factor is not sufficient or conclusive for the process to proceed, because

[t]hese functional units are constantly dislodged and other molecules or functional unit tokens

of the same type take their place. Hence, the causal initiation of gene expression cannot be represented accurately as an individual binding event; a diachronic sequence of binding events involving many different molecules is required. (2015, 6-7)

Therefore we have to agree with the authors that the occurrences at the token-level are not difference-makers themselves, “but represent, at best, *necessary* conditions (if interpreted at the type level).” (p. 7) And this means, even if we give a further detailed description of the first component in my diagram, the causal organization of the refined components (the binding activities of each entity making up the functional unit) should have to be described at the type-level. The claim can be generalized for all components of the mechanism, too: They stand as abstract type-level individuals, and it is not their token instances *per se* that are the difference-makers. This is one aspect of the dissimilarity between mechanisms as models and the causal structure (the real system) they are about.

There is also another aspect addressed by Love and Nathan that is based on the latter author's articulation of causation by concentration (Nathan 2012): Some of the actual and known difference-makers are not represented in the mechanism for gene expression. (In discussing this aspect, I will be examining the practice of idealization.) The RNA transcription phenomenon does not only depend on the binding of the transcription factor to the operator; for the RNA chain to be formed stably there must be adequate supply of molecules to substitute for the detached ones. On top of this, this is not an issue about the *number* of available molecules; it is rather about their quantity *relative to* that of other molecules present in the system—that is, their concentration. This is because of the following reasons (only given in brief here, to be elaborated while discussing the subsequent example):

(1) The available molecules must be able to traverse distances in the cell, so as to reach the operator sites on the DNA. For this (i.e., diffusion) to take place, a quantitative threshold set with respect to how much of other molecules are present must be attained, to drive the molecules to the neighborhood of the DNA (cf., Nathan 2012, 10); and

(2) There are also repressor molecules in the cell which can bind to the operators and thereby hinder gene expression. The relative amount of the transcription factor to the repressor molecules is a significant difference-maker in the stable regulation of gene expression.

The upshot is, the binding of the transcription factor plays a *partial* explanatory role in gene expression. If the mechanism model is expected to involve all of the occurrences relevant to the phenomenon, then it must include the role of the unattached molecules (both the initiators and the

repressors) in the cellular environment as well. The relative concentration of the molecules captures this aspect of the system. However—and here is the crux of the matter—*this aspect is not amenable to be truthfully represented in models of mechanisms*. For on a mechanistic explanation, the configuration of the components (what mechanistic philosophers label as the causal organization) represent the system as if the relations between them are mechanical interactions.

Biologists are perfectly aware of the difference-making role of concentrations in cellular contexts . . . but they deliberately misrepresent them. The causal relations that produce the explanandum are idealized in mechanistic diagrams; their representation intentionally ignores known variations in properties and other components that make an actual difference. (Love and Nathan *forthcoming*, 7)

We have a case of *idealization* here, as the mechanism does not truthfully represent all of the actual difference-makers relevant to the investigated phenomenon. It misrepresents the binding of the transcription factor as if it is an occurrence that takes place in an environment without the resisting factors, and without the contributions of the relative quantities of different compounds on the binding. No actual binding can be as such, though. (If this looks like a case of abstraction rather than idealization, the elaboration of causation by concentration in the next section should correct that impression.)

To recall a point previously made, this is a shortcoming of mechanistic explanation in general; it is neither the failure in the particular explanation given for the gene expression phenomenon nor due to the explanation's being incomplete. For the most part, it is due to the epistemic motivation to represent the causal relations by construing the components as though they are the activities of localized parts of a machine—or in other words, the motivation to simplify the representation of the system by focusing only on its machine-like aspects. But as the case at hand (and Love and Nathan's discussion) displays, such components are fictions that are epistemically and pragmatically useful.

Bestowing the components this ontic status should not mean that there are no entities engaged in such activities as described by mechanisms. I accept that there are molecules, genes, operators on DNA, bindings, initiations and blockings of gene expression, etc. Nevertheless when the corresponding types are identified as the component entities and activities of the mechanism for gene expression, their relevance to the phenomenon is often fictionalized. They stand for myriads of interactions and their relata that form the causal structure, but the actualization of the phenomenon is not due to the mechanistically described organization of the corresponding tokens. In fact, strictly speaking, the objective causal structure does not harbor instances of components. For the

components are individuated by their relevance to the phenomenon and their concrete instantiations in themselves do not bear this very relevance. Components only denote that there are occurrences within the causal structure that play recurring roles in the occurrence of the investigated phenomenon.

These are quite justified strategies in modeling, but they do not provide the metaphysical warrant for counting tokens as relevant to the phenomenon. It could be objected that the tokens are really tokens of components in virtue of their definition. According to this objection, an F6P molecule can isomerize to G6P or a transcription factor can be detached, but what is relevant to the explanandum phenomenon are still the token occurrences where G6P isomerizes to F6P or the transcription factor is bound—so there must be nothing wrong in taking *these* as the actual instantiations of components. Nevertheless this would not save mechanistic realism, because it is not true that mechanisms are composed of *all* of the relevant occurrences relevant to the phenomenon. The reason has already been implicated: In most of the real systems that concerns mechanistic philosophy, there is causation by concentration, by which we should understand that not all of the actual difference-makers can be modeled mechanistically. In other words there are difference-makers quite relevant to the phenomenon but which cannot be tokens of components. This follows from the construal of components, by which only the machine-like aspects of a system is represented. Leaving out the non-machine-like aspects amounts to idealization, because it is distortion in the service of providing a simplified causal narrative about the system. I shall now enlarge upon Nathan's account to provide a more persuasive argument.

3.6 NATHAN'S ACCOUNT OF CAUSATION BY CONCENTRATION

In order to gain a better insight into the notion of causation by concentration, we need to consider the gene expression phenomenon with some more details involved. Nathan's (2012) own example is the viral infection of the bacteria *E. coli* where the virus reproduces by making use of the bacterium DNA machinery. There are two modes for the virus (the *bacteriophage*) to replicate: In *lysogeny*, it integrates its nucleic acid into the host's genome. It is said to be *passively replicated* in this case, as the process is carried out by the normal reproduction of the bacterium. In *lytic growth* (the second mode) the viral genetic material is used for taking the active and full control of the host's machinery, to the end of replicating rapidly and extensively. This *active replication* results in lysis, the destruction of the host cell and its membrane. Whether lytic or lysogenic growth takes

place is determined primarily by which gene of the bacterium is expressed. The two genes of concern are labeled shortly as *cI* and *cro*; when these are transcribed, cI and Cro proteins are synthesized in the cell (respectively). These are also the proteins which bind to the operator on the bacterium DNA, thereby regulating the respective gene's transcription.

Briefly put, the regulation is based on the antagonistic functioning of these proteins. When cI binds to the operator, *cI* is transcribed, *cro* is silenced and the bacteriophage is kept in its lysogenic state (passive replication). Likewise, when the Cro protein binds (to the same operator), the gene *cI* is silenced, *cro* is transcribed and lytic growth (active replication) of the bacteriophage is induced. The norm is lysogenic growth; this is to say, unless there is a perturbation on the cell (e.g., UV radiation) passive replication is maintained and there are more cI than Cro. Otherwise, there is a flipping of a switch: for instance, due to UV radiation, the chemical structure of the cI molecules change and it becomes much less probable for them to bind to the operator and much more probable for the Cro molecules to replace them, leading to active replication that ends with the lysis of the cell.²⁵

It was mentioned above that the relative concentrations of molecules have a causal impact on phenomena like this one, the gene expression in viral reproduction. Let us see it more concretely now. Just like the case with the transcription factor in eukaryotic gene expression, the binding of cI proteins is not sufficient for the stable transcription of the *cI* gene. For a token cI binding does not last long enough; the protein detaches, and it must be replaced by the other cI molecules. This requires not only enough supply of cI in the cellular medium but also the cI molecules' being close to the vicinity of the operator on the bacterium DNA. Otherwise the Cro molecules may bind and repress the expression of *cI* to initiate lytic growth. This has a low probability to occur, though (unless there is an external perturbation as explained above), due to the cI:Cro relative concentration. The concentration of cI relative to that of the repressors is high enough to ensure that even Cro binds to the operator, there are enough cI in the neighborhood to take the seat (Nathan 2012, 9 and 11).

Thus the relative concentration in favor of cI guarantees the binding—but this is not its sole causal contribution to the process; it also enables positive and negative feedback loops. If the magnitude of cI:Cro is high enough, until a certain threshold, it keeps triggering the synthesis of more and more cI molecules. Yet the rise in the number of cI also has the consequence of cI's

²⁵ These are explained in more detail in Section 3 of Nathan 2012.

binding to the O_{R3} part²⁶ of the operator, to which they actually have a low affinity (so if their concentration is not that high, usually no binding takes place). The binding results in the inhibition of the transcription of *cI* and consequently, to a decrease in cI:Cro (i.e., in favor of the Cro relative concentration). Hence, the relative concentration causally determines the gene expression, and it is due to this that Nathan sees it as an actual difference-maker (2012, 15).

From these Nathan infers that the regulation of gene expression is a result of two distinct kinds of causal processes: “a mechanical one, instantiated by protein binding to operators, and a statistical one, instantiated by the relative quantity of molecules satisfying a concentration threshold.” (2012, 13) External perturbations like UV radiation affects the mechanical process by cleaving the structure of the cI proteins, and this causes the flipping of the switch for the initiation of the synthesis of Cro proteins. But in normal circumstances, lysogenic growth of the bacteriophage is secured, and to a great extent this is due to the statistical causal contribution of relative concentrations. According to Nathan, what we have here is a case of redundant causation (preemption), in which the available unbound proteins act as the backups—potential causes, so to say: the binding (the preempting cause) brings about the transcription of the particular gene, but if it fails there are other proteins that ensures the same effect takes place (the preempted causes). But unlike most of the cases of redundant causation we are familiar from the literature, the redundancy here emerges not only because *there are* backups. It is also (and more critically) because of their quantitative status. Nathan illustrates this by a comparison of the following scenarios (which are, arguably, more true to the spirit of canonical preemption examples where murders are planned with a great deliberation):

- To make sure that a person is killed, an assassin is hired to shoot a victim and, unbeknownst to the killer, a deadly poison is also administered, in the eventuality that the killer fails. The assassin successfully performs her job; the poison turns out to be redundant.
- To make sure that a person is killed, a mobster hires a number of assassins who roam the city in search of the designated victim. One of the assassins finds her and kills her; all other killers turn out to be redundant.

Nathan writes, “[I]n the second scenario, but not in the first, the number (concentration) of killers plays an important role in explaining why the victim was guaranteed (or almost guaranteed) to die.”

²⁶ More properly put, the operator that I have been citing is a *complex of operators* consisting of three seventeen-base-pair individual operators (Nathan 2012, 7). O_{R3} is one of these, the one that is closest to the *cI* gene. O_{RI} is situated at the other end of the complex, next to the *cro* gene; and O_{R2} is the one in the middle.

(2012, 15) He asks us to consider the size of the city: The mobster must determine the number of the backup assassins according to the size, because the same number can become insufficient (to guarantee the killing) if a larger city is of concern. The analogy with the concentration of the molecules in the cellular environment is obvious.

Nathan identifies the causal role that concentrations play with three characteristics (2012, 17). First, concentrations are multiply realizable: Different distribution tokens can realize the same concentration. Second, the concentrations make probabilistic differences: Even when the Cro molecules are high in concentration, there is still some probability that lytic growth may not prevail. Third, causation is redundant in the sense given above: The redundant character is due to the quantitative status of the potential, backup causes (which make actual difference by ensuring that the expected effect occurs). He concludes the discussion as follows:

I suggest that we call *causation by concentration* . . . a causal relation that is multiply-realizable, probabilistic, and ‘quantitatively’ redundant, since the effect is triggered by changes in the concentration of actual or potential causes. When all three conditions are met, the effect is irreducible . . . to actual causal interactions. (*ibid.*)

This account is not put forward as an alternative theory of causation to differentiate between what is and is not a cause, but to expose that there is a plurality in the kinds of causal relations (2012, 14 and 16). Moreover it is meant to apply beyond the explanatory contexts where concentrations are appealed to. Nathan gives an example from simple economics concerning the efficiency of profit-making, for which not only the actual transactions but also the availability of potential customers should be taken into account (pp. 19-20). Pursuing on this line, I think we can view the opening-closing of the ion channels (in the generation of the action potential) in terms of causation by concentration, too. With respect to its causal character, channel opening is analogous to the protein molecules' binding to the operators: It is probabilistic, and a token occurrence is not sufficient for the effect (channels can close and proteins can be detached). Likewise, what guarantees the occurrence of the expected phenomenon (significantly rapid change in the voltage across the membrane due to the movement of large amounts of ions) is the backup causes, that is, the availability of the ion channels that can open. We should contend that the ratio of the number of available channels to that of the unavailable ones (the inactivated channels) makes a significant and actual difference on the outcome. For it is directly this ratio that determines (in the case of sodium channels) if sodium influx will trigger the firing of the spike, if the membrane will be repolarized, or (in the case of potassium channels) if the membrane will enter into the refractory period.

Here, then, we have a multiply realizable, probabilistic and quantitatively redundant causal relation, obtaining between this ratio and an expected phenomenon. This is clearly a case of causation by concentration, which illustrates that cellular environments and chemical compounds are not necessary for this kind of causation to obtain. In this respect “concentration” can be understood more broadly, extending the scope of applicability of the notion introduced by Nathan. Wherever mechanistic explanations (in high-level sciences) involve the modeling of systems with causation by concentration, we can be sure that there is idealization.

3.7 MECHANISMIC NON-REALISM

A recap is due, after six sections on the metaphysical nature of mechanistic explanation and its implications for mechanistic philosophy of science.

I take mechanisms to be quite useful representations of the causal structure. One can be said to be explaining a phenomenon *mechanistically* when she likens the system wherein the phenomenon is produced to a machine-like whole. This modeling practice involves both abstractions and idealizations—that is, it involves the omission/exclusion of many features of the system and also assertions of falsehoods regarding the causal interactions being modeled. The practice rests on qualitative causal reasoning, which is made possible by weaving a causal narrative into the configuration of mechanism components. The narrative is about objects entering into mechanical interactions (such as binding, turning, opening) with each other. Usually, though, this amounts to a fictionalization: We know that in most of the real systems that are subjected to scientific mechanism-talk, what is actually relevant to the phenomena (the actual difference-makers) exceeds the totality of the token (local) mechanical interactions holding between real objects. Global quantitative properties such as relative concentrations of compounds/objects (“concentration” is construed broadly, to apply to numbers of objects, too [numbers of open and closed ion channels, for instance]) and the aggregative effects of particular occurrences (e.g., stabilization of gene expression enabled by the bindings and detachments of many transcription factors, the total behavior of a population of stochastically operating ion channels) make critical contributions to the production of the explanandum phenomenon. Even so, scientists can explain types of phenomena mechanistically. They are able to do so by introducing idealizations regarding the non-machine-like aspects of the system, and abstracting away from the token-level that enables them to focus on the repeatable features of the system (whenever the explanandum phenomenon is produced).

Let me highlight that, seen from a scientific realist perspective, it is the practice of idealization—not the practice of abstraction—that renders the truthfulness of the model questionable. For, given that idealization is the practice by which the scientist makes false assertions (usually deliberately) regarding the investigated system, the truthful representation of the system is compromised for the sake of simplicity. This is the case wherever we have an idealized scientific model; it is true both for a model of, say, a Carnot engine (Psillos 2011), or for models which make use of a frictionless surface, point masses, infinite population sizes or infinite amount of supply. I assert that mechanisms are idealized models in this sense, where the concern for simplicity is fulfilled by the assumption that the system is machine-like. Abstraction, on the other hand, is not a threat for a scientific realist, because it only involves remaining silent regarding some details of the system: Adding the omitted information to the model (such as the information about the causal relations in a system that are irrelevant to the investigated behavior, or the token-level information concerning concrete, particular occurrences) would not make it more truthful—it would only make it more detailed.

My argument in this chapter, which put specific emphasis on the type *vs.* token distinction and the role of abstractions and idealizations in modeling, is meant to serve the purpose of distinguishing mechanisms from their instantiations. We take the former as the explanantia in scientific mechanism-talk concerning actual regular patterns. This implies that mechanisms are the abstract and idealized models themselves (or, models of ideal, machine-like systems), set at type-level. This further implies that, especially due to the fictionalization involved in constructing these models, what mechanisms are models of cannot be mechanisms. For according to the way they are described, mechanisms are configurations of component entities entering into mechanical interactions; but then, the instantiations of these are not the only relevant occurrences (relevant to the phenomenon they actually contribute to)—this is to say, the configuration of the components is not instantiated as it is described in the model. Mechanismic realists are mistaken at this very point, for holding that mechanism models are about the objectively existing mechanisms. As a consequence, they are mistaken in their ontological commitment to mechanisms.

That said, for being a scientific realist, I hold that mechanisms are *about* some real features of the world: Nothing in my argument is an objection to the premise that there are real objects and occurrences causally involved in the production of phenomena. Yet, when the real causal organizations of these is considered, they do not look like they are described by mechanisms. So my position concerns a *non*-realism with respect to mechanisms. This is not an antirealist position (although it could be made compatible with scientific antirealism), for I do hold that these models

approximate an objective reality. It is a scientific realist position differing from that of the mechanistic realist, with regards to what is being modeled and what really underlies actual regularities. The next chapter is an attempt to put flesh on this position, by suggesting an answer to how mechanisms can explain regularities without being there in the empirical world, that is, without being harbored within real systems of interest (as causally organized components making up a machine-like whole).

A very plausible mechanistic realist objection to this position can be voiced by drawing on Craver 2014. In this paper, Craver emphasizes that it is the ontic structures that explain a phenomenon, and what makes a model good/bad and complete/incomplete is to be evaluated according to how much the model reveals these structures. He does not overlook the fact that explanations (not just mechanistic explanations, but scientific explanations in general) bear communicative and cognitive virtues (i.e., epistemic virtues) as well, but his stress is on the ontic mode of explanation where what is really “in the world” dictates the normative requirements, not how well the world is represented or how well the information about the ontic structures is conveyed (2014, 50). In this regard, there being false assumptions about the world inherent in an idealized model of a real system do not delegitimize a commitment to the real features of the world. In Craver's words:

The very idea of an idealized model of an explanation commits one, at least implicitly, to the existence of an ontic explanation against which the model can be evaluated. It is more sensible to say that idealized models convey explanatory information in virtue of making false assumptions that bring certain truths about the ontic explanation to light. (*ibid.*)

To a certain extent, this just rephrases my thesis that mechanisms being idealized models are still about the world. Nonetheless, our agreement on this can be exploited by a mechanistic realist to question my non-realism concerning mechanisms. Plausibly, due to the chief connotation of the label “mechanistic non-realism”, he could ask me what is wrong in committing to mechanisms being the very features of the world represented by the not-so-truthful idealized models. This is to say, after the discussion in this chapter, a mechanistic realist could be persuaded to abandon his implicit conviction that mechanisms represent real machine-like systems; but he could still retain his commitment to *whatever it is* that is being mechanistically modeled and insist to view these as mechanisms. He would then be offering a new formulation of mechanistic realism and it would not seem very different than the *non*-realist position I am defending, except for the *prima facie terminological* disagreement on whether or not to call the modeled real existent as mechanism.

I respond to this objection (and a related one regarding the realism about abstract objects, such as scientific models) at the final section of the next chapter. I will assert that the disagreement is more severe: Contrary to the first impression, it is not terminological but meta-ontological; it depends a lot on what we understand from being a realist (or non-realist) about anything. I will state a preference for the question of “what grounds what” over the question of “what there is”. The consequence of the preference is that for a realist, whether to commit to this or that (as being existents) is a concern *less important* than evaluating the worth of the commitment itself.



CHAPTER 4

A FAVORABLE METAPHYSICS

FOR THE MECHANISMIC NON-REALIST

How can one make sense of the claim that mechanisms explain actual regularities obtaining in the world, while arguing that mechanisms are not real features in the world? This will be the question that will concern me in this chapter. I will attempt to forge a metaphysics on which the real features of the world, that is, causal properties, underlying the regular patterns of concern (repeatable manifestations of dispositions) can be shown as appropriate to be modeled by mechanisms.

At the beginning of Chapter 2, I mentioned Bogen's critique of the regularity requirement on the characterization of mechanisms (found in his 2005). To recall, Bogen addresses MDC's definition of mechanisms which supposes that mechanisms are productive of regular changes. His critique is successful, so long as the notion of regularity is taken to imply that the phenomena associated with mechanisms occur *frequently*. It is a fact that many of the phenomena investigated in high-level sciences (especially those in life sciences) do not occur frequently; what is more frequent is their non-occurrence. The gene expression is a case in point: Most of the times, the gene is silenced rather than expressed. Or if we remember Bogen's example, the failure of neurotransmitter release upon the arrival of the action potential wave to the pre-synaptic axon is much more frequent than its success (the failure rate is as high as 90%). Nonetheless we know that scientists provide mechanistic explanations for the successfully yet infrequently occurring phenomena. This is, of course, not to say that they are not interested in explaining the failures; but if they do that, they explain different phenomena: They are not explaining gene expression when they explain its failure, and they are not explaining signal transmission between the neurons when they explain the failure of neurotransmitter release. (Actually a good explanation for why neurotransmitters are not released has not been given yet. So my point concerns the hypothetical case where it is provided.) This underlines the same verdict once more: Mechanistic explanations are independent of observations concerning the frequency of the phenomena; a mechanism is characterized by the phenomenon it explains, and this is the end of one side of the story.

On the flipside the story continues as follows: Characterization of mechanisms must still be

made by appealing to some form of regularity, because otherwise *any* causal chain can be counted as a mechanism. (I am writing from the side of mechanistic philosophers, assuming—for the sake of the argument—that mechanisms are within the causal structure.) This is a point made in DesAutels 2011 and Andersen 2012. DesAutels gives the example of the assassination of Archduke Franz Ferdinand in 1914, followed by many countries' declaring war on each other, which in the end led to World War I.

This is a clear example of a singular causal chain; it only happened once. However, it apparently qualifies as a mechanism. . . . we have entities (Franz Ferdinand and the forgoing countries) and activities (assassination and declarations of war) that are productive of change (the start of World War I). What we have, on an irregularist characterization, is apparently a mechanism. (DesAutels 2011, 922)

As one can guess, the “irregularist characterization” that DesAutels writes about is one by which mechanisms need not be productive of regular changes. Glennan has laid down the connections between historical explanations and mechanistic explanations in terms of their narrative structures (2010). On his approach, the agents and their actions in historical narratives bear similar explanatory roles to those of the entities and activities figuring in mechanistic explanations that mechanistic philosophy is concerned with. This opens up the possibility that there can be “ephemeral mechanisms” which produce changes in the world only once. We could take these mechanisms to exemplify the irregularist characterization.

But as I have argued in the opening chapter, Glennan is concerned more with analyzing causal relations in terms of mechanisms. This is the reason why I never considered him as a mechanistic philosopher, because mechanistic philosophy of science is first and foremost a position taken with respect to a particular type of scientific explanation. For the mechanistic philosophers, the task is not analyzing causal relations or statements in general—they leave that to the metaphysicians. Mechanistic discourse is rather shaped for making sense of the *regular patterns* by the use of mechanisms. Hence, whereas Glennan can be content with identifying singular causal chains with mechanisms, a mechanistic philosopher should ask for a more strict use for the notion of “mechanism”. DesAutels and Andersen are two such mechanistic philosophers who are explicit at this point. They both stress the explanatory worth of mechanisms for the *scientifically* investigated phenomena which occur recurrently, and the need to distinguish mechanisms from other causal chains, in order to reserve the notion for organized occurrences which are the (objective) bases for scientific generalizations (DesAutels 2011, 922; Andersen 2012, Section 4). Andersen expresses the idea succinctly:

Retaining regularity in the definition of a mechanism involves retaining the ability to differentiate a stronger and more explanatorily useful subset of causal explanations from other kinds of causal explanations. . . . The regularly recurring chains of entities and activities that constitute a mechanism make them more useful in explanations of other chains of causal-activity-connected entities because regularity grounds generalization from one instance to further instances. The more regular a mechanism is, the more potential instances to which the same explanation applies. (2012, 428)

Andersen's paper offers a broadened characterization of regularity so as not to abandon the requirement completely. I will enlarge on her account below. But to close the present discussion, I shall state that for a mechanistic philosopher, there can be no such thing as an ephemeral mechanism. So far this point is made in connection with the role mechanisms are supposed to play in explanations. But there is an additional argument for this that concerns the identification of the components of a mechanism. If mechanisms were allowed to be chains that run only once, we would not be able to ascertain which causal factors actually contributed to the production of the phenomena. On this line of thought (see Andersen 2012, 426-427) phenomena having multiple instances allow scientists to draw comparisons and let them find out what is common in all. It goes without saying that for a mechanistic realist, that which is common is the mechanism for the phenomenon under consideration. For instance Andersen puts this realist contention in words as follows: "The more often a mechanism occurs, and the more regularly it occurs, the clearer it becomes which entities and activities are part of the mechanism (i.e. what the mechanism is). It also becomes clearer that *there is a mechanism there at all*, something that either could explain a phenomenon or which could require further explanation." (*ibid.* [emphasis added])

If one is to adopt the non-realist interpretation of mechanisms I put forward in the previous chapter, she cannot endorse the view expressed by Andersen as is. For once, according to the view I favor, mechanisms are not there. One only has a mechanism as an abstract and idealized model of type-level components, and the configuration of the latter is not instantiated as it is described in the model (strictly speaking). I indicated before that the non-realist interpretation is compatible with a scientific antirealist stance regarding mechanisms, and that this is not the line I want to take. I rather want to maintain that there are objective truthmakers for these models, but they are not mechanisms themselves. They are real, "genuinely existing" causal properties that confer dispositions to their bearers (as Chakravartty holds, 2007 and *forthcoming*).

This is what I am going to argue for, after spending some time on Andersen's characterization of regularity and Nathan's "simulacrum account of dispositions" (2013). With the help of the former, I

will make the point that *regularity construed as frequency* is one thing and *regularity construed as repeatability* is another thing; and it is the latter that is of concern in the present discussion. Then, Nathan's account will serve as a path that one could take, if he wants to bring mechanistic non-realism together with dispositions-talk. In the section following this, I will state my reason to avoid this path, while still holding the position that dispositional properties (causal properties that confer dispositions on their bearers) are the key to adopt a consistent scientific realist and mechanistic non-realist stance. Finally in the last section I will discuss a line objection that can be put against my stance, and I will advocate a meta-ontological stance to argue for the following:

- (1) A mechanistic realist might want to view the cohering set of dispositional properties itself that is represented in a mechanistic explanation as a real mechanism. But this would not amount to adding a new item to our ontology. Understanding mechanisms being real features of the world in this vein results in appreciating their *derivative* ontic status.
- (2) Alternatively, one might want to commit to mechanisms as abstract objects. But this metaphysical realist attitude does not have much bearing on our scientific realist ontology.

4.1 REGULARITY CONSTRUED AS REPEATABILITY

In her paper Andersen works out a taxonomy that is intended to accommodate different forms of regularity mechanisms exhibit. For her, mechanisms can fail to produce the phenomena they are individuated by. But since mechanisms overlap with and/or are embedded within each other, the failures can serve functions in some other mechanisms. A failure at a particular causal step in a token chain does not imply a failure on the part of a larger (i.e., higher-level, overlapping) token mechanism depending on this step.

Andersen's taxonomy is based on three parameters along which the regularity of a mechanism can vary: organizational location, strength of the causal connection, and the statistical pattern for the exhibition of the regularity.¹ The first parameter concerns the four *stages within the spatiotemporal organization* of occurrences at which the causal relations might be regular or irregular (2012, 419). These are: (1) the start-up conditions for the mechanism, (2) the triggering of the first component, (3) particular causal relations internal to the mechanism (obtaining between the components), and

¹ I shall follow Andersen's enumeration (Arabic numerals for the first parameter, small letters for the second parameter and small Roman numerals for the third parameter). While talking about the last parameter, I will reverse the order of introduction; that is why her "ii" will be introduced before "i".

(4) the termination conditions. As it is apparent, it is necessary that the states of affairs located at each stage must obtain in order to progress into the next stage.

The second parameter, *strength*, concerns how often the effect follows the cause (at the organizational location selected). Andersen addresses four degrees of strength (2012, 420): (a) deterministic, (b) “reliable but not exceptionless”, (c) sporadic (the cause *often* fails to result in the effect), and (d) infrequent (the effect does not follow in most cases). This offers a more refined classification than that found in Barros 2008, where stochastic mechanisms are divided into two as *unbiased* and *biased*, according to the probabilities of the expected outcomes: unbiased if the outcome has a probability of 50% or less (e.g., fair coin flipping mechanism, neurotransmitter release mechanism), and biased otherwise (e.g., natural selection of a beneficial trait). A mechanism exhibiting either of the latter two degrees of strength addressed by Andersen could be grouped as an unbiased stochastic mechanism by Barros' classification; likewise, the other two degrees of strength overlap with the cases of biased stochastic mechanisms (p. 421).

Finally the third parameter concerns the *patterns of deterministic failure* (*ibid.*): It could be that (ii) there are interferences to the mechanism that scientists have some knowledge of, but whose probabilistic impact cannot be estimated. Andersen's example for such case is ecosystem succession mechanisms being countered by various chancy disturbances (such as adverse weather, human interference, geological events). Alternatively, but not necessarily excluding the previous pattern, (i) it is possible for the indeterminacy to be expressed in a statistical distribution, which shows a consistency when multiple occurrences are considered. The failure in the neurotransmitter release mechanism displays this pattern (more to be said below). Andersen argues that deterministic failure does not make a mechanism straightforwardly irregular. If the mechanism conforms to one of these patterns, it can still be considered to exhibit a kind of regularity. The following example is meant to clarify both her taxonomy and the argument for the assertion just mentioned (*ibid.*).

Let us see how Andersen puts this taxonomy into work. The example she considers is the failure in the neurotransmitter release mechanism (that is supposed to show that the mechanism cannot abide by a regularity requirement). Andersen puts particular emphasis on the embeddedness of mechanisms and each mechanism's roles in others (i.e., higher-level, lower-level and/or overlapping mechanisms). Her main contention is, different (embedded) mechanisms display different forms of regularity, and in spite of the indeterminacies concerning a single occurrence, it can be shown that a “meta-regularity” is exhibited “that justifies calling the process a regularly occurring mechanism.” (2012, 424) She analyzes the mechanism for the release of neurotransmitters (NT) into the synaptic

cleft separately from the broader mechanism for neuron-to-neuron communication. Let M_L and M_H be these mechanisms, respectively (“L” standing for lower, “H” standing for higher). The former phenomenon explained by M_L should be seen as a component in M_H , the mechanism for the triggering of a post-synaptic firing due to pre-synaptic action potential (AP). A schematic illustration of these is as follows.

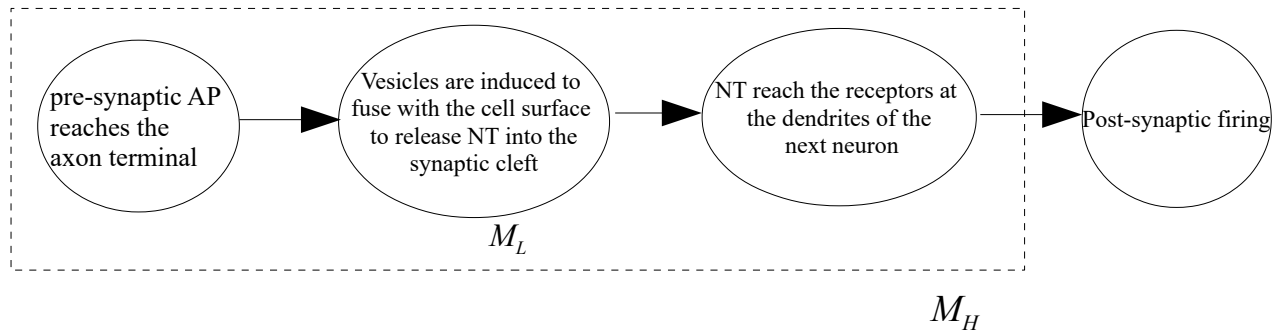


Figure 4-1. Diagram showing the triggering of post-synaptic firing. M_L is the mechanism explaining the release of NT to the synaptic cleft. M_H is the mechanism explaining neuron-to-neuron communication.

Consider M_L first. When a pre-synaptic AP reaches the axon terminal, it is expected to trigger the release of the neurotransmitters, but as has been mentioned before, the failure rate is as high as 90%. Andersen views the strength of the cause-effect connection as sporadic (second parameter, [c]): “it is not an exception or unusual when a vesicle fails to dock and release transmitter.” (2012, 423) (I would rather see it as infrequent [d], as the 90% failure rate means the effect does not follow in most cases. But this is inconsequential for the actual point to be made.) The organizational location of the sporadic causal relation in question (first parameter) is (2): Although the start-up conditions for M_L are present, they only sporadically *trigger* the first component. Lastly, as yet, scientists do not know the factors interfering to the release of neurotransmitters; they suspect that calcium ions might have some role in this, but even if it is so, this role is yet to be discovered (*ibid.*). However, the statistical magnitude of the failure can be provided, which makes it a case of (i) concerning the third parameter. Thus, according to Andersen's taxonomy, the pattern of regularity in the embedded mechanism M_L is labeled as (2-c-i).

What about the regularity that the embedding mechanism M_H (an etiological mechanism that causally gives rise to post-synaptic firing) exhibits, given that one of its components, M_L , is a sporadically triggered mechanism? Andersen wants us to observe that the picture changes in this

case. First of all, the organizational location is now (3)—that is, the step we are concerned with is now located among the causal relations internal to the mechanism in question. At this step of the causal chain, only 10% of the vesicles release neurotransmitters into the synaptic cleft. The important thing is, although this is a low percentage, the number of the successful release of neurotransmitters is sufficient for the chain to proceed. This is equal to saying that the failures of M_L does not directly translate as the failures of M_H . Because the embedding mechanism M_H involves a large number of attempts to trigger the embedded mechanism M_L . For this reason, when considered within M_H , the cause-effect connection at the relevant step has a heightened strength: It is a *reliable* phenomenon (second parameter, [b]) that neurotransmitters are released. This said, the patterns of deterministic failure (third parameter) does not change with the shift in the mechanisms, because we are concerned with the same locus in both mechanisms and so, there is no difference with regards to the nature of the indeterminacy involved (2012, 424). Hence, Andersen concludes that the label for the regularity observed in M_H should be (3-b-i).

With this broadened characterization of regularity, Andersen can pinpoint how one sort of *irregularity* can make a contribution to the overall process: “By only responding to some fraction of potential triggers, pre- and post-synaptic cells effectively dampen many signal propagations, thereby limiting the amount of noise in the overall system while still ensuring that a sufficient number of cells overall respond to propagate signals.” (2012, 425) What this means is, if the phenomenon the embedded mechanism M_L generates occurred much more frequently, the regulation of signal propagation in the brain would not be as finely tuned (it would probably result in a signal overload). In effect, this emphasizes the interconnectedness of mechanisms, by urging one to explain a regularity pattern *qua* its role in larger, higher-level mechanisms (e.g., particular functions of the brain). Therefore failures in regularity should not be taken as a proof that the regularity requirement on mechanisms must be abandoned. Different patterns of regularities obtain depending on which phenomenon is investigated; finding out about these amounts to how one mechanism is situated within the others.

In my opinion, Andersen's characterization of regularity offers a workable *mechanistic*² account of infrequent yet repeatable (recurring) causal patterns in the world, by shifting the emphasis away from whatever happens (or not happens) at a particular locus on the causal nexus. *Repeatability* is an important notion here. It differs from frequency for requiring only the pattern to have multiple

² Why not “mechanismic”? The reason is, while Andersen's proposal itself is surely mechanistic, I will adopt its *mechanistic* aspect only. Here I am referring back to the distinction made in Chapter 3, between mechanistic explanation and mechanistic modeling. To recall, the latter is the presentation of the former under the mechanistic realist guise.

instances—that is to say, *without* requiring that the pattern is instantiated more (often) than it is not instantiated. In saying that mechanisms explain regularities, one must be presupposing that there are repeatable patterns that give rise to these regularities, and that these patterns are formed by particulars whose occurrences are *necessary* for every occurrence of the recurring phenomena (whose regularity is of concern).³ Both a mechanistic realist and a mechanistic non-realist hold that these patterns are repeatable for being instances of the same type-level mechanism. But they differ on the issue of whether to consider these instances as token mechanisms.

The mechanistic realist affirms that they are, but this is not a tenable position. It is not difficult to see this: Take M_L as a token mechanism, that is, as one of the recurring instances of the mechanism type $\mathbf{M}_L$ that explains how NT are released into the synaptic cleft. We have already seen that M_L is neither sufficient nor necessary for the instantiation of the token embedding mechanism M_H : Because M_H depends on many instantiations of $\mathbf{M}_L$ (not one), and the failure of one instantiation is not a hindrance at all for a token post-synaptic firing (i.e., the phenomenon explained by M_H)—in fact, post-synaptic firing is still possible despite the 90% failure rate. Yet if M_L is neither sufficient nor necessary, there is no way for it to be relevant for M_H , either.⁴ This contradicts M_L 's being a *component* in M_H , which is a premise we do not have the luxury to do away with.

The mechanistic realist faces this difficulty, because she does not acknowledge that mechanism components can only belong to the abstract and idealized model (the mechanism)—not to the real causal structure. She holds that mechanism models represent token causal chains, by abstracting away the irrelevant occurrences, and leaving only the ones that are relevant to the explanandum phenomenon which she dubs as the components. On this interpretation, the instances of the model are the actualization of the components as described at the type-level: The really existing particulars which underlie the repeatable patterns, in other words, must be the concrete, token components, and the patterns of concern must be the causal organization of these tokens. This perspective urges the mechanistic realist to regard components as necessary occurrences at the token-level, but they are not, as the neuron-to-neuron communication example demonstrates. The correct way to grasp the components requires detaching them from the token-level, which means, regarding them only as types.

³ Why “necessary”? Because by the realist premise that there are real patterns giving rise to actual regularities, the obtaining of an actual regularity can be seen as a *sufficient* condition for there to be patterns that are repeated. Then the particulars forming the patterns are necessary for there to be repeatable phenomena.

⁴ Here I am only invoking an intuition about causal necessity/sufficiency being related with causal relevance. The causal relevance of a factor to a phenomenon can be expressed in terms of difference-making, and the manipulationist theory (Woodward) can be of use here (as mentioned before in Section 3.4.1). The latter can also be of use to provide a theoretical underpinning for how causal relevance is related with causal necessity/sufficiency. In particular, Woodward's notion of invariance should be addressed in this context.

On *any* scientific realist reading, saying that a type-level component is necessary for the phenomenon-type amounts to the following existence claim: That whenever the phenomenon-type is instantiated, there are concrete particulars which instantiate the component. On the position I am advancing (that is, mechanistic non-realism), these particulars are *modeled* by the mechanism components; the latter are not instantiated as they are described in the idealized model (this is where the non-realist asserts a different claim than the realist), but stand for the real repeatable features of the world. What exactly are these features, then?

Causal properties, I should say as a mechanistic non-realist *and* a dispositional realist. It is due to bearing these properties that objects enter into causal relations, that is, repeatedly manifest their dispositions in the presence of other dispositions. This position requires elaboration, which I will do in the next two sections, but I shall present its core tenet here, that which concerns the identification of causal properties (or, the identity criterion for properties). It is a simple thesis that can be found in Shoemaker 1980 and Swoyer 1982 (cf., Chakravartty 2007, 113-114): Two causal properties are identical if they confer the same dispositions to the objects that have them. Take fragility of glass, for instance (a paradigm example). This property of one glass is identical to that of another glass if both glasses are disposed to break under similar conditions. Or consider the set of properties of the neuron membrane, by which it is able to generate action potentials (e.g., its selective permeability to ions, its being composed of voltage-gated ion channels, etc.). This set can be said to constitute a causal property of the membrane, one that disposes it to generate the action potential wave when it is stimulated (in the presence of other causal properties). Two membranes having the same causal property must generate the same wave (same in terms of peak magnitude, shape, etc.); or the same membrane generates similar waves in similar conditions, due to having the relevant causal property. This is the neat feature of the *causal theory of properties* that is associated essentially with Shoemaker: The repeatability of the relations that the objects enter is inscribed on the causal properties they bear. I will use this as the metaphysical warrant for there being actual and regular patterns, by which the repeatability of the phenomena can be explained.

I propose to view the components of mechanisms as manifestations of dispositions obtaining in idealized settings. This should offer a much more fine-grained metaphysical analysis that is suitable for elucidating the investigation of causal chains and systems through mechanisms. The translation of its basic premise to the mechanistic discourse would be this: Entities engage in activities due to the causal properties they possess. These properties are identified by their causal profiles, which means that they confer the same (genuine and occurrent) dispositions on their bearers to manifest the same behavior under similar conditions. This is the metaphysical basis for the actually recurring

patterns in the objective world investigated by sciences. Mechanistic explanations are about these causal properties, by mechanisms serving as models for how these properties *cohere* recurrently, and consistently. On this view, an actual occurrence involves the manifestations of *many* dispositions (within a system) conferred by the cohering causal properties. The configuration of mechanism components is an *explanatory misrepresentation* of the real causal organization of these properties.

Before stating my position more clearly, I shall review Nathan's account of dispositions first. As it stands, this account is not directly related with mechanistic explanation; Nathan's proposal rather concerns the connection between dispositional *ascriptions* (*not* the causal properties or the occurrent dispositions) and regularities. On the other hand, as we will see, it would not take much for one to bring this account together with the mechanistic *non*-realist position I have outlined. I will show how this could be done, and then leave it at that point. This is to say, I will not adopt Nathan's account—I cannot, because concerning dispositions it is not a realist view. Hence, by reviewing Nathan's account, I hope to point at a path that a mechanistic non-realist can be inclined to take, especially if he is interested in dispositions-talk like I do. But I will avoid that path, for reasons to be elaborated in Section 4.3.

4.2 AN AGNOSTIC VIEW: NATHAN'S SIMULACRUM ACCOUNT

Genes are transcribed into RNA, neurotransmitters are released into the synaptic cleft, and voltage-gated ion channels open to allow ion influx and efflux across the neuron membrane. All of these occur when suitable conditions are present and when some things trigger the respective individuals to behave so: Transcription factors' binding to the operator site on the DNA, an action potential wave's reaching the axon terminal, and an increase of positive charges inside the neuron (within the neighborhood of the membrane) are the triggers of these, respectively. This being so, dispositional properties can be ascribed to the objects of concern in these occurrences, by which we can say that genes are disposed to be transcribed, neurotransmitters are disposed to be released, and ion channels are disposed to open (under this and that condition). This can allow us to provide grounds for the regular patterns we observe (with a suitable analysis of dispositions). But there is this well-known problem with these ascriptions: They are meant to capture the regularities, but we know that even when the necessary conditions are met the dispositions may not be manifested.

I shall refer to Nathan's elaboration of and solution to this problem (2013), because it fits nicely with the theses made in regards to mechanisms. (This should come as natural, given that concerning mechanistic explanation I have drawn a lot upon Love and Nathan's views. But there is a point where my approach strays away from the position he favors.) That “dispositions capture regularities” is one of the three claims Nathan takes to be widely accepted in the analyses of dispositions (p. 2). The other two claims are the following:

- Dispositions are properties of entities.
- Dispositions are connected with conditionals.

Nathan's attempt is to examine dispositional properties in such a way that the simple conditional analysis (SCA, given below) can be applicable, even though the object can fail to manifest the disposition it bears.

(SCA) x is disposed to D when C iff x would D if it were the case that C .

(x : the object bearing the disposition; C : the necessary conditions for the manifestation of the disposition)

He inquires into the truthmaker of SCA and first considers the following two options, taking the genes' disposition to be transcribed as his example (pp. 4-5).

(i) Genes' disposition to be transcribed is grounded on the fact that most genes are reliably transcribed as RNA.

(ii) When all the necessary molecules are present in the cell, genes are reliably transcribed as RNA.

We can guess his evaluation of these, given our familiarity with the example: Neither of them would work, because genes are not reliably transcribed even when all the necessary molecules are present. ((ii) is the refined statement of (i), so its falsity implies the falsity of (i).) For there are repression mechanisms that can prevent the genes' transcription. This urges Nathan to resort to a refinement of (ii):

(iii) When all necessary molecules are present in the cell and there are no repression mechanisms, genes are reliably transcribed as RNA.

Nathan acknowledges that (iii) captures an actual regularity, but he voices two worries: It is a statement of a generalization which obtains under quite particular conditions. This is not what one

wants from the truthmaker of the claim concerning the disposition of the genes. On top of this, (iii) reads as an analytic claim, as it boils down to saying “genes are transcribed when there are no mechanisms to prevent transcription.” (2013, 5) What should be the solution to offer, then? In the passage quoted below, Nathan hints at the direction to take:

The disposition ascription makes a statement about *all* genes—namely that, under appropriate circumstances, they display a tendency to be transcribed. What grounds this statement is a regularity: genes are regularly transcribed as RNA. Now, as noted, this regularity does not hold in the world since, as a matter of fact, most genes are actually silenced. Nevertheless, in *idealized settings*, when the transcription apparatus is in place, genes are always transcribed as RNA. (2013, 7 [original emphases])

By idealized settings, he means those that are held to obtain only in *models* of causally produced phenomena. He gives an example for such a model, a colorful depiction of gene expression in the cell (his Figure 1 on p. 8). It illustrates how the double helix unwinds as the enzyme RNA polymerase proceeds along the DNA sequence, to allow the binding of the RNA nucleotides to the DNA template strand. This is a typical illustration of the phenomenon that can be found in many relevant textbooks. On the other hand, what it depicts is “seldom or never instantiated” in real-life cells (*ibid.*). Nathan underlines the four respects in which the model asserts a falsehood⁵: It does not depict the repressors at all, it shows the base pairs as completely visible and accessible, it involves the assumption that the enzymes and transcription factors are present in the right quantities, and it represents the cellular environment as quite spacious where molecules can easily travel without bumping into any impediments.

Nathan raises the following question: Given that such models (and their depictions) are known to be inaccurate, how can we hold them to describe the phenomena investigated by sciences? His answer is ready: With these models, our aim is not to *describe* the phenomena but to *explain* them. (*ibid.*) He relies on Cartwright's (1983) contention that there is a necessary trade-off between descriptive accuracy and explanatory power; aiming at accuracy comes with the cost of losing the explanatory insight. Models provide explanations by exhibiting fictional states of affairs which assert that the investigated phenomena occur when the real-life situations “suitably resemble” them: “In this respect, the model of gene expression is analogous to the ideal gas law ($PV = nRT$) or the claim that water is H_2O , which also represent idealized conditions that are seldom or never instantiated (no real gas satisfies the law, and pure H_2O is extremely rare outside the lab), but are highly explanatory.” (Nathan 2013, 8)

⁵ As can be recalled from the last chapter, “assertion of falsehood” is a characteristic of idealizations, according to Jones (2005, 175).

Nathan's proposal is to take dispositional ascriptions to apply only to such models. This is to say, to abandon the idea that dispositions are properties of *real* objects, given that they are manifested in idealized settings. Put in his words: “[D]isposition ascriptions should not be taken at face value; dispositional properties are not satisfied by real entities. Dispositions are properties that pick out classes of models in which a certain behavior occurs without exceptions.” (2013, 9) For once, this has the benefit of preserving SCA. The conditional is made true in the idea model, and it can be safely put to work for the analysis of dispositional ascriptions (p. 14). Furthermore, this “simulacrum account of dispositions”⁶ has the consequence of capturing regularity, though at the abstract level. This comes to mean that regularities captured by the dispositions do not need to correspond to actual, real-world regularities (p. 16). Nathan welcomes this outcome, because according to him, dispositions are “placeholders” for properties that explain why the real objects behave the way they do (pp. 11-12).

This is perhaps the most outstanding claim he puts forth (and one that I do not endorse—see below). His justification for this is that, when—for instance—a gene is said to have the disposition to be transcribed, the dispositional ascription is utilized only to the end of capturing a pattern that itself requires an explanation: The explanans for why the gene is transcribed is not the disposition, it is the idealized model where the ascription holds. This entails the elucidation of how the analysis of dispositions is related with the explanatory context they are situated in, by “provid[ing] a unified framework for analyzing and explaining dispositions.” (p. 16) The unification is due to the truthmaker of the subjunctive conditional in SCA being the model that explains how generalizations such as “genes are transcribed” are made true.

This is also the point where the model-theoretic approach to mechanisms, the one that we find in Nathan's work with Love (as reviewed in the previous chapter), can enter the stage. It seems quite intuitive and straightforward to form a coherent account that brings the simulacrum account of dispositional properties together with earlier theses made with regards to mechanisms.⁷ There is only one philosophical “trick” to perform, actually: We should regard the explanandum phenomenon as a manifestation of a disposition. So where we observe, for example, instances of neuron membranes' generating action potentials, we say that the neuron membrane has the relevant disposition and it is this property that is manifested in these instances. On Nathan's simulacrum account, the possession of the disposition does not explain the phenomenon (or its repeated

⁶ Nathan gives credit to Cartwright's “simulacrum account of explanation” as the basis for this label (2013, 9).

⁷ To my knowledge, Nathan and Love's analysis of mechanistic explanation and Nathan's own work on dispositions have not been officially brought together. Here I attempt to do the work on behalf of them.

instances); the dispositional ascription itself awaits explanation. Hence, enter the mechanism for this phenomenon as its explanans: It is nothing other than the model we postulate the idealized settings with, by which the dispositional ascription is made true. This would also work towards justifying the repeatability of phenomena. Earlier I have claimed that repeatability follows from the multiple instantiation of types (of mechanisms). Now one could propose to view these types as the “classes of models” that Nathan mentions (“in which a certain behavior occurs without exceptions”—to re-quote from above).

4.3 THE NEED FOR A METAPHYSICAL WARRANT:

DISPOSITIONAL REALISM

I admit that the framework just presented is neat enough as it stands. It covers sufficient ground for a mechanistic non-realist to base his thesis that mechanisms do not underlie actual regularities. But it remains silent on the issue of what there is to commit to. This is the crucial question for a scientific realist. If one cannot provide an answer to this, then he cannot put forward a concrete alternative to mechanistic realism. Apparently Nathan is a philosopher who does not intend to undertake this task. Instead, he is explicitly promoting agnosticism with regards to the real basis of regularities, and “with respect to the nature of the most fundamental denizens of reality.”(2013, 13-14). He seems right in contending that the simulacrum account is compatible with different options (including one on which fundamental entities may turn out to be irreducibly dispositional at the end of the day⁸). Nevertheless, as a scientific realist, I feel obliged to make a more exact statement, rather than saying that there may be several options.

As indicated earlier, the position I adopt is not compatible with the simulacrum account of dispositions. This is because, I contend that dispositions are real, occurrent properties of objects, *regardless of whether they are manifested or not*. The claim (to be elaborated below) is a direct consequence of the need to show the *metaphysical* warrant for why there are actual regularities—that is, why there are repeatable (but not necessarily frequent) patterns in the empirical world. The argument to come rests on the causal theory of properties (Shoemaker) mentioned above, and Chakravartty's “dispositional identity thesis” (DIT) which is just a different reading of the former

⁸ Nathan: “However [in that case] such primitive dispositions would be different from dispositions of ordinary macroscopic objects (salt, glasses), and of those scientific posits (genes) to which the simulacrum approach can be applied.” (2013, 13)

for putting an emphasis on dispositions (2007, 123): “It is a consequence of DIT that objects with the same causal properties have the very same dispositions for behavior, simply because of what makes these properties the properties they are.” (p. 129) Here is another quote from Chakravartty, to see the identity thesis in a simple example:

If entities [objects] having negative charge are disposed to repel other entities with negative charge, and to attract entities with positive charge, then these dispositions are part of the nature of charge—they constitute (in part, since negative charge is also associated with other dispositions) the identity of charge. (*forthcoming*, 23)

Bearing a causal property does not necessitate that the object will manifest its disposition, because manifestations require certain causal settings to obtain. A dispositional realist understands the obtaining of a causal setting in terms of a multiplicity of manifestations occurring altogether, in a causally organized manner. Therefore a single instance of manifestation cannot be an isolated occurrence; it is an instance located within a concrete causal pattern formed by the causal properties conferring dispositions on the objects. By DIT, it is then true that another instance of a manifestation that is suitably similar to this one entails the involvement of a suitably similar pattern of causal properties (and *vice versa*). The repeatability of the instances of this pattern is the metaphysical warrant we are looking for.

This is a crude and hasty version of the argument a dispositional realist can propose, so as to assert that it is the patterns of causal properties (and therefore, the patterns of dispositions they confer) that underlie actual regularities. If this is a sound argument, then we have a scientific realist alternative to mechanistic realism—an alternative by which we can put forward some positive claims with regards to mechanistic explanation, that is, about what is being represented by mechanism models (if not mechanisms themselves). Here, then, is a slower and thorough construction of this argument which is based on Chakravartty's work on dispositional and scientific realism (2007 and *forthcoming*).

We begin by specifying actual regularities as the recurring (repeatable) behaviors of objects. On dispositions-talk, this is read as the manifestations of the same dispositions in the same causal settings. Consider a pumpkin pie with a mass of 350 grams.⁹ We can say that this mass is a causal property the pie possesses (causal properties are quantitative and determinate properties¹⁰), conferring numerous dispositions on the pie, such as the disposition to fall to the ground with a determinate acceleration (*if we let it fall*), or the disposition to make the screen of a digital weighing

⁹ This example is based on (Chakravartty 2007, 62-63).

¹⁰ See (Chakravartty 2007, Section 2.3).

scale display “350g” (*in the presence of a working scale, which rests plainly on a table, etc.*). The manifestations of these dispositions require the presence of other objects, and the obtaining of many conditions, altogether making up a causal setting (that is summed up by the italicized *ceteris paribus* clauses). If the falling of the pie with that acceleration, or the displaying of “350g” on the scale are the recurring behaviors we are concerned with, then we can say that such regularities obtain thanks to the manifestations of the same dispositions and the obtaining of the similar (or suitably similar) causal settings. The warrant for these manifestations is the dispositions that are genuinely occurrent, and the warrant for the *repeatability* of these manifestations is nothing other than DIT. Such is the way a dispositional realist could think.

Before applying this framework to the scientific context, let me consider two objections that the opponents of dispositional realism could make. The first is the familiar one that is due to Nathan, according to which, saying that the object is disposed to behave in a certain way *when* the favorable causal setting obtains (i.e., when the *ceteris paribus* clause is made true) would not be saying anything new or philosophically interesting. This objection makes a good point, indeed: It is true that this cannot provide an analysis of the disposition (or the dispositional ascription) in question. In other words, referring back to our example, it is true that the pie's disposition to make the digital scale display “350g” cannot be analyzed in terms of the fact that this is the case whenever the pie is put on the scale, and the scale is turned on and tared, etc. But who wants to *analyze this disposition*? Nathan, for sure, but not the dispositional realist. It is certainly a concern for the *empiricist* to account for dispositional ascriptions in terms of the actual states of affairs. For the dispositional realist, though, the explanation works the other way round: He posits dispositions in order to be able to account for the actual regular patterns.

I would say that we should see this as an instance of *metaphysical* explanation, which is different than the type of (empiricist) analysis of dispositions that Nathan is looking for. So this type of objection, which is a motivation for Nathan to advance a simulacrum account of dispositions, misses the target: The dispositional realist is not the one to be bothered by the triviality of the statement that there is manifestation whenever the setting is favorable. We (the dispositional realists) do not appeal to dispositions in order to *scientifically* explain the behaviors of the entities. We appeal to them in order to *metaphysically* explain *that* which is common in the same behavior in its repeated instances. Scientific explanation concerns what happens when a disposition is manifested—that is, say, when the scale screen displays “350g” or when the gene is transcribed to RNA. On the other hand metaphysical explanation concerns the repeatability of these behaviors. There is nothing in the empirical world that warrants the latter; repeatability is not a subject of

empirical investigation. Hence our motivation to resort to dispositional realism as a way of rationalization of the empirical world. Nathan says: “In many cases, a disposition is posited for lack of a better explanation [of a regular behavior].” (2013, 11) But if by “better explanation” we mean the *scientific* explanation of how the behavior is exhibited, all we can hope for is gaining an understanding about the *manifestation* of the disposition—which is different than both the disposition and the *repeatability* of the manifestation.

Let us now move on with another objection to the dispositional realist framework I have been promoting. This one concerns the very commitment to dispositions: I say such commitment is grounded in the motivation to metaphysically explain the basis for actual regularities. To a metaphysical antirealist or a deflationist (about dispositions), though, this may look as a *strained* metaphysical position that is too much *tailor-made* for a realist ontology. Both of these opponents of the dispositional realist can question the ground for realism, that is, both of them might make this assertion: “Your commitment boils down to a commitment for the sake of commitment, because your criterion of property identity (DIT) boils down to saying that what explains regularity is what is regular (which, you suppose, are the behaviors objects exhibit and the causal relations they enter into). Why should anyone *buy* this?”

Such a questioning is an issue that is taken up in Chakravartty *forthcoming*, and Chakravartty's response to this appeals to the need for a coherent “scientistic” account. The debate is on the accommodation of non-empirical notions in one's ontology, and the opponents of the realist seem to have the upper hand for being free of the baggage that comes with the commitment to dispositions. But in fact they do not, because—as we are about to see—their own criterion for property identity are disadvantageous for not falling within the reach of science.

Let us first take a look at the larger context that concerns Chakravartty, to see the advantage of adding dispositions in one's ontology. What is at stake for the author is forging “a metaphysics for scientific realism” (the title of his 2007). The opening discussion of the book is on entity realism and brands of structural realism. Chakravartty's thesis that follows this discussion can be put very briefly as this: Scientific theories incorporate judgments that are based on detected properties and auxiliary posits for heuristic reasons (2007, 47-50). Theory changes affect primarily the latter while preserving the *concrete* structures that are scientifically discovered. Scientific realists should be comfortable in making ontological commitments regarding concrete structures, because they are identified with relations between first-order properties *which are nothing other than the causal properties scientists detect* (pp. 39-43). A detection involves the obtaining of many interrelated

causal processes underlain by quantitatively determinate causal properties. Causal processes involve incessant manifestations of dispositions in the presence of other dispositions (pp. 108-111), therefore detecting a causal property means detecting the behaviors associated with (the manifestations of) a network of dispositions. This is a ticket to knowing the nature of a causal property, and this knowledge is packed together with the knowledge of the other properties repeatably cohering with it. (This is because, by DIT, which other properties a causal property coheres with is “encoded” on that property.) Hence, given the scientific interest in the repeatable features of the world, it makes great sense for the philosopher of science to suppose that scientists make claims regarding the identity of properties of objects; and given the scientists' causal contact with the network of properties, DIT does not seem so strained as an ontological thesis of property identity. Accepting DIT also opens up the following line of thought for the scientific realist: The knowledge of concrete structures are scientifically conveyed in generalizations presenting the second-order properties, which are the formal properties of relations between causal properties. Mathematical formulations (such as law statements) are good examples to these, which yield the knowledge in *abstract* structures (Chapter 6). So do the numerous other scientific models, which involve abstractions and idealizations and therefore deviate from truth—even so, they give approximately true information about the concrete structures they describe (in the case of abstractions) or make *existential claims* about (in the case of idealizations) (Chapter 8). In short, the commitment to dispositions opens up a philosophically robust theoretical space for the scientific realist that accords well with the scientific practice.

Now, let us assess how the opponents of dispositional realism fare with respect to the metaphysics of science they offer. For the antirealist (say, a Humean) what identifies a property of scientific interest is not its causal profile (its dispositional character), but its primitive whatness (*quiddity*) that separates it from the other properties, and nothing else. Take negative charge, for instance. From the viewpoint of the antirealist, the causal relations the bearer of the negative charge can enter has nothing to do with this property. A particle can happen to possess negative charge, but it is just a contingent fact that it *attracts* a particle with the opposite charge—there can be a possible world where the two rather *repel* each other. Following this line of thought, however, creates a “pragmatic incoherence on the part of the dispositional antirealist” (Chakravartty *forthcoming*, 24). For at the end of the day, it is not compatible with the empiricist sensibilities of the antirealist: Construed as such, *quidditism* implies either that not much can be known or said about the natures of properties—but given that science does aspire to investigate these, this hardly offers a productive (and empiricist) metaphysics of science. The deflationist is not in a better position. The above

discussion of Nathan's account can be read along this line. The simulacrum account has been introduced as an *agnostic* view about dispositions, but in essence it builds on deflationism—that is, for treating dispositions (the ascriptions) as linguistic functional units that are elliptical for something else. Nathan's agnosticism implies a neutrality towards the latter, and it is this feature of deflationism that is subjected to Chakravartty's criticism:

[Deflationism] is no good either, because what would dispositional ascription here be elliptical *for*? If it is elliptical for quiddities, then one is back where one started. On the other hand, if dispositional ascription is merely an inference ticket, this would make the natures of properties in the world a function of human inferential practices, which surely sounds too much like idealism for any contemporary philosopher of science to accept. (*forthcoming*, 25 [original emphasis])

Hence, in spite of the metaphysical baggage that comes with it, committing to dispositions can be shown to bring a payback: a robust and productive metaphysics of science that accords much better with the scientific practice in comparison with the empiricist, non-realist positions concerning dispositions.

Time for a recapitulation: The central discussion in this section has been justifying the commitment to dispositions. Relying on Chakravartty's account, I have provided a justification on the basis that a coherent metaphysics for scientific realism follows neatly from dispositional realism. Given dispositional realism, one can contend that science gives us a true (or, approximately true) picture of reality—namely, the causal structure harboring concrete structures underlain by naturally cohering (quantitative and determinate) causal properties. These properties confer dispositions on their bearers to behave in a certain way under similar conditions (in the presence of other dispositions manifesting themselves). Dispositions being occurrent properties of objects is the warrant for the repeatability of behaviors.

Now, to couple this with mechanistic explanation: Recalling the motivation for mechanistic non-realism, we define mechanisms as the configurations of components—that is, as *mentally formed abstract structures of idealized manifestations*. We can take these abstract structures to be the idealized models Nathan is talking about: Settings where behaviors occur or dispositional ascriptions are made true without exceptions. The talk of such models constitute the point where our dispositional realist interpretation of mechanistic explanation is in accordance with Nathan's simulacrum account. We take mechanisms to explain phenomena by cashing out the latter in terms of an object (e.g., a neuron membrane) manifesting its dispositions (e.g., the disposition to generate action potential) in idealized settings. Nathan take these settings to be “idealized” due to there being no interference in objects manifesting their dispositions, such as there being no factors to repress the

gene's transcription as RNA. Considering mechanisms under the lens of dispositions, we could make the additional remark that the settings are idealized for being likened to the workings of machines (i.e., machine-like systems). This is to say, on this analysis, it is not only that we take mechanistic explanations to *idealize away* the factors that repress transcription in real systems, but also we hold that mechanisms model the coherence of the causal properties as if they are reducible to local states of mechanical affairs between objects: For instance, the binding of the transcription factor to the operator on the DNA is cashed out as the bilateral manifestations of the relevant dispositions of the transcription factor, of the operator and perhaps of the objects in the vicinity of these—but in no case we consider the manifestations of the global properties of the system (e.g., the relative concentrations of different compounds). So far, this is not much different than the simple approach by which I coupled the simulacrum account with my mechanistic non-realism, at the end of the last section.

And here is the point where I (the dispositional realist) start to speak differently from the deflationist: I say that, despite their abstract status, mechanisms are still about the concrete structures wherein complex causal relations obtain between objects manifesting their dispositions. As a term-of-art that I take from Chakravartty 2007, a “concrete structure” is formed by the complex interplay of causal properties at both local and global level. The need to handle this complexity firstly leads to abstracting away some occurrences as irrelevant to the behaviors we investigate, and secondly leads to simplifying the relations obtaining between the *relevant* causal properties. I take mechanistic explanations as causal narratives with the salient feature of *fictionalizing* a configuration of relations (between components) by likening the system to a machine. The fictional configuration is the mechanism, which is explanatory by being not *depictive* but *denotative* of the relevant portion of the concrete structure wherein the explanandum (simpliciter) is harbored. Mechanistic realists are wrong for supposing that mechanisms are concrete causal organizations of occurrences as described in mechanistic explanations; this is why I say that the explanatory fictional configuration is not depictive. I say it is rather denotative, for it makes an *existential claim* with reference to the repeatable features of the causal structure. These features are the manifestations of the occurrent dispositions, and I want to underline that the repeatability involved in these has two aspects. First, a particular manifestation is always the manifestation of the same disposition (there is no multiple realizability [or something close to that] of manifestations according to the theory of dispositions favored here), and second, a manifestation of a particular disposition always requires the same causal setting (i.e., the presence of the same dispositions). But separating these aspects in our rhetoric should not give the impression that they are separable. After

all, both the causal theory of properties and DIT rests on their inseparability.

If these make sense for the mechanistic non-realist, then I believe he has the metaphysical resources required for answering the question at the beginning of this chapter. This may not be the only plausible way for one to make it intelligible that actual regularities can be explained by mechanisms (as abstract, idealized models). But I think the account presented here provides a considerable metaphysical support for a scientific realist, who might want to argue against the mechanistic philosophers on the grounds of their ontological commitments.

4.4 EXISTENTS AND REALIST COMMITMENT: META-ONTOLOGICAL CONCERNS

All of what I say above points at the compatibility of dispositional realism with mechanisms being not-so-truthful, abstract and idealized models of real systems. Just to furnish this point a little more, I could add that Chakravartty writes about the “sociability” of property instances—a metaphor that stands for the clustering of dispositions that tend to be manifested altogether (2007, 170). By this metaphor, Chakravartty intends to make a contribution to the natural kinds debate in the literature; in the hands of a mechanistic realist, however, it could be used to assert that mechanistic explanations reveal which causal properties of the component objects tend to be manifested together in an instance of an explanandum phenomenon.

As implicated at the end of last chapter, this compatibility has the potential to backfire on my stance, that is, mechanistic non-realism. Following the line of thought summarized above, a mechanistic realist could happily endorse the dispositional realism defended here (frankly speaking, I would appreciate that), but then also plausibly argue that the repeatably cohering sociable causal properties, those whose manifestations are subject to mechanistic explanation, are the ontic structures that deserve a *mechanistic realist* commitment. Put differently, a mechanistic realist still has the right to question my non-realism. For instance, drawing on his own paper on the ontic account of explanation (2014), Craver could argue that I have given an account of how models of mechanisms differ from the real systems of concern (due to the latter having many non-machine-like features), but I have still not provided any argument showing why the cohering causal properties *cannot* be the very structures, the *mechanisms* simpliciter so to say, that ontically explain the phenomena.

Any mechanistic realist voicing this concern would be right, I admit. Nonetheless I will not be providing the sort of argument he requests from me. This is because my non-realism is not much about the objects of existential commitment; it is rather about distinguishing the *fundamental* existents from the *derivative* ones. This involves engaging in a meta-ontological debate, which I now begin by explicating the content of the mechanistic realist commitment.

I say that one is a mechanistic realist if he claims either or both of the following claims:

SR (Scientific Realism about mechanisms): “Mechanism models represent real, objectively existing, repeatable, concrete structures. It is these structures that ontically explain the phenomena and that the scientists discover, and my commitment to mechanisms as real features of the world amounts to a commitment to these (be them machine-like or non-machine-like).”

MPR (Metaphysical Realism about mechanisms): “Mechanism models are abstract objects, useful in scientific inquiry. Just like with other abstract notions (e.g., numbers, sets, universals), their being abstract does not imply that they are not real. By some arguments for their indispensability, it can be shown that one can commit to them.”

My mechanistic non-realism will be put against SR on the grounds that one's scientific realist commitments should actually be about the *fundamental* denizens of reality, and if we are to take mechanisms as the concrete structures underlain by sociable causal properties and modeled through mechanistic explanations, then at best we are assigning them a *derivate* ontic status. In a similar vein, against MPR, it will be asserted that the existential commitment to mechanisms as abstract objects is again a commitment to derivative existents which does not have a bearing on scientific realist ontology. The important point to be elaborated below is that, my non-realism does not involve arguments against the existence of mechanisms construed in either of the senses.

4.4.1 Defending Mechanistic Realism by Psillos' Realism

As far as I am concerned, when the mechanistic literature is surveyed, one finds the best exposition of a claim approximating to SR in Illari and Williamson 2011 and Craver 2014. MPR, on the other hand—to my knowledge—has not been defended by any mechanistic philosopher yet. As a matter of fact, I think the most rigorous defense of both SR and MPR can be put forward by

drawing upon Psillos (2005 and 2011) who is not a mechanistic realist, but whose work on scientific realism in general has a clear relevance to the present discussion. In his paper “Scientific Realism and Metaphysics” (2005) Psillos distinguishes two conceptions of reality which he labels as “factualist” and “fundamentalist”. It is the former conception that Psillos advocates (for reasons to be explored below); this preference is motivated by a need to maintain a scientific realist attitude without being committed to a particular metaphysical stance. Here Psillos is responding to the challenge raised by Brian Ellis, who—in Psillos' words—“argues that scientific realism should be seen as a rich metaphysical world-view that commits its proponents to physicalism and non-Humeanism.” (2005, 386) I shall leave his debate with Ellis aside and instead focus on the argument Psillos pursues, because it is the same argumentation by which a mechanistic realist could justify SR. Likewise Psillos' other paper “Living with the Abstract: Realism and Models”, which “argues that scientific realism can happily co-exist with models *qua* abstracta” (2011, 4) could be used by a mechanistic realist to defend MPR.

According to Psillos, *the factualist conception of reality* amounts to taking reality to be comprised of truth-makers without regarding whether they are irreducible, basic or fundamental (2005, 387-388). A scientific realist working with this conception is to take mature scientific theories “at face-value” and be at ease in committing to the posits of them (p. 385) because of her contention that what are represented in a true theory are facts about existents after all. She might want to make some claims about the facts themselves, but in that case those claims would be motivated on grounds other than scientific realism. For instance, a factualist scientific realist could easily assert that a biological property genuinely exists. To do this, all she needs would be showing that positing this property is *indispensable* in explaining well-founded phenomena (p. 398) and endorsing the realist attitude that by positing it, the theory represents some facts inhabiting the world. This is enough for her to work as a scientific realist. If, however, she held that biological facts are reducible to some more fundamental (say, physical) facts, and felt an urge to track her commitment down to the fundamental level, this would neither impair nor contribute to her scientific realism. For she would be working with *the fundamentalist conception of reality* (p. 388), which is compatible with scientific realism, because if and how non-fundamental facts are reducible to fundamental ones is a separate (metaphysical) issue and reducibility does not imply that there are no non-fundamental existents (p. 396):

The right (realist) attitude for a given set of contested propositions is to start with a commitment that it *does* represent genuine facts and then to engage in the independent debate about whether they are *sui generis* or not. If it is shown that these genuine facts are not *sui*

generis, if, that is, there are some more fundamental facts that render the contested propositions true, this might revise our deep metaphysical commitments but not our claims to truth and reality. (p. 391 [original emphases])

Psillos forwards a similar view with regards to taking models as real (abstract) existents in his 2011. Importance is given again to the criterion of explanatory indispensability: If x meets this criterion (i.e., if x is indispensable in explaining well-founded phenomena), then there is no impediment to counting x as real and committing to it:

This is a permissive criterion. It does not dictate the status of entities that are explanatorily indispensable; in particular it does not disallow abstract entities from being real. Differently put, it is one thing to say that x is real because it meets the explanatory criterion; it is quite another thing to say that x is concrete, or abstract or physical or mental and so on. (2011, 15)

Psillos uses this to resolve a dilemma scientific realists face in considering the truth of scientific theories which is usually compromised due to the idealizations involved in model-building. The practice of idealization, by involving false assumptions about the world, seems to disallow the realists from taking theories at face-value. Therefore the “central dilemma” for the realist is the following: She should either take scientific theories as false or give up on understanding theories literally, which means giving up on realism (2011, 6). Psillos notes that the dilemma arises when the *process* aspect of model-building is foregrounded—in other words, the aspects by which a model deviates from reality is considered. Nevertheless, Psillos contends, if we instead turn our attention to the *product*, that is the idealized model itself, then we see a way out of the central dilemma: On this perspective, we should recognize that theories are indeed true of models being abstract objects (besides being true of at least some concrete objects) (p. 9). For in the final analysis, the model works by making true statements about idealized settings. This does not imply that models are unrelated with the systems they represent, because they are described so as to “[isolate] the explanatorily relevant features of the represented system with respect to the behaviour under study”; therefore there is still a match holding between the models and the real systems. (p. 16) This is also how they become not only representative but also explanatory. And for bearing such an indispensable role they deserve to be counted as real, that is, existing among other abstract objects such as numbers, point masses, frictionless surfaces, or properties (assuming that the latter are also indispensable in explanation).

I hold that Psillos' views put forward in these papers would offer an unmistakable support to a mechanistic realist, who would like to argue for SR and MPR. Here is the argumentation he could pursue: In support of SR (“my commitment to mechanisms as real features of the world amounts to

a commitment to some concrete, ontic structures”), he could first identify my mechanistic non-realism coupled with dispositional realism as a stance in line with the fundamentalist conception of reality that Psillos discusses. (To note in advance, he would be right.) He could then grant that some facts might turn out to be more fundamental than the facts about causally organized occurrences of concern in mechanistic explanation. But, he would note, there being facts about the latter already provides the foothold a mechanistic realist requires: These facts could be cashed out in terms of dispositions-talk, but they are facts anyway; how to understand them metaphysically, or whether it is worthwhile to look for more fundamental facts about them are issues independent of scientific realist commitments. Hence he could conclude that mechanistic realism is based on this factualist conception of the portions of the objective world that is the subject of mechanistic explanation, allowing him to maintain that whichever facts are represented by good mechanistic explanations, they deserve to be seen as the real mechanisms existing in the world. On top of this—in support of MPR (“Mechanism models are abstract objects, indispensable in scientific explanation; so I can commit to them”)—the mechanistic realist could go on to say that even as abstract objects, mechanisms *as models* are worth committing to. Their abstractness does not entail their non-existence; given that they are known to bear explanatory indispensability, they could be taken as real as any other posit in mature scientific theories about the mind-independent, real world.

4.4.2 The Meta-Ontological Argument Against Mechanistic Realism

The thrust for the arguments for SR and MPR is the conviction that the main question for a realist ontology is “What exists?”. As we have seen, explanatory indispensability plays the key role in assigning a real status to the posits of concern, be them concrete or abstract; and once a posit is shown as indispensable, the main question is considered as answered. At the end of the day, this results in a quite permissive ontology. In this final section of my dissertation, I want to assert that this much permissiveness is susceptible to the charge from triviality. Schaffer's influential essay “On What Grounds What” (2009) provides all the resource I need to substantiate my responses against the mechanistic realists claims, SR and MPR. These will exhibit a fundamentalist conception, in the sense envisaged in Psillos 2005. This should imply at the outset that I do not agree with Psillos on the independence of metaphysical (and meta-ontological) concerns from the main task of a scientific realist.

Schaffer's thesis in his 2009 is that there is a *neo-Aristotelian* alternative to the established way

of doing ontology, that which is *Quinean*. He distinguishes these two meta-ontological stances according to the central discussions they engage in: While a Quinean sees the question of ontology as “What is there?”, for the neo-Aristotelian, it is “What grounds what?” According to Schaffer, the Quinean enterprise displays the subject matter of ontology in a “flat structure”, by which a list of existents is provided. This is not much satisfactory on the neo-Aristotelian framework, though, because—taking inspiration from Aristoteles' *Categories*—a neo-Aristotelian takes a “permissive disinterest” in what exists and instead focuses on “*how* something exists” (2009, 352). This is to say, the neo-Aristotelian meta-ontology champions the revealing of the hierarchical nature of the existents in an “ordered structure”, by way of describing the *grounding relations* between the existents. Displaying the ordered structure of the world of existents is therefore an activity exceeding the listing of the items, it is also an engagement with the goal of showing how the existence of the items on the list are hierarchically related (pp. 354-355).

Schaffer raises three arguments for motivating the neo-Aristotelian way of doing ontology (over its Quinean rival). The third of these arguments ("The Quinean method presupposes Aristotelian structure" [2009, Section 2.3, 366ff]) is directly about the Quinean method. I thus think that it is the least important one in the context of the actual discussion I am pursuing, and want to take a quick peek at it. Schaffer surveys the Quinean method in several stages, these are: identification of the best theory, identification of the canonical logic, translation of the best theory into canonical logic, determining the domain of quantification that will make the translation true, and evaluation of the existence commitments of what results. With regards to all of these stages, Schaffer contends that the Aristotelian ordered structure of Being is either already invoked, or would lead to a better ontology even in the hands of a Quinean (pp. 372-373).

Let us now look at Schaffer's other—more essential—arguments:

S1. Existence questions are trivial (2009, Section 2.1)

S2. “Central metaphysical questions are not questions about whether entities exist, but only about how they do.” (2009, Section 2.2, 363)

I first examine the argument for S1. Schaffer's point here is that, the existence of many items of the ontology can be deduced from simple and obviously true propositions about them. This is why he claims that existence questions are trivial. One example he gives is the following (2009, 357):

1. There are prime numbers.
2. Therefore there are numbers.

Schaffer reminds that 1 is a mathematical truism, which is “more credible than any philosopher’s argument to the contrary.” (*ibid.*) On considering a possible charge from numbers being fictions, he writes that 1 makes no claims about fictions and such a charge would not touch the existence argument. (He also considers two other possible replies, which I skip.) The other examples he gives (ones about the existence of properties, mereological composites and fictional characters) share the same simplicity in argumentation, and all underline the following crucial point: that permissivism about existents is plausible, but it is philosophically more worthwhile to show which existents are *fundamental* and which others are *derivative* (p. 361). This point is put forward most vividly (in my opinion) in connection with the debate on the existence of God. After recalling that an atheist takes God as a fictional character (and that she does not need to commit to the inexistence of fictional characters), Schaffer writes,

if the theism debate were about the existence of God, then the following would count as a defense of theism: (i) God is a fictional character, and (ii) fictional characters exist, hence (iii) God exists. But obviously that is no defense of theism! Hence the theism debate is not about existence. (2009, 359)

Schaffer's S2 is built upon this point. For this claim, Schaffer states that many important discussions in metaphysics are indeed not about existence/inexistence, but about whether *x* is mind-dependent or mind-independent. He sees this as an issue about ontological priority, that is, about what grounds what. Some of the discussions he views under this light are metaphysical realism vs. metaphysical idealism, realism vs. nominalism about universals, mind vs. matter, and the debate over substantival space (whether space is grounded in its occupants) (2009, 362-363). Regarding all these debates, Schaffer points out that there is “little disagreement about what exists, but profound dispute over what is fundamental.” (p. 362)

This summary view of Schaffer's defense of a neo-Aristotelian meta-ontology should suffice for me to object to mechanistic realism expressed in terms of MPR, that is, realism about mechanisms construed as abstract objects. (My objection to SR, however, will require additional arguments to show that the concrete structure underlain by causal properties has an ontological priority over mechanisms as ontic structures explaining phenomena.) The first step in my objection is the acknowledgment that an existent is to be considered as mind-dependent if it is abstract (and mind-independent if it is concrete). This we can see as a truism, and accordingly, I assume that there is nothing debatable about taking mechanism models (alongside other scientific models) as mind-dependent. Then, given the scientific realist stance adopted both by me and the mechanistic realist, there should be no disagreement about the ontological status of the models either: They must be

ontologically derivative, because for a scientific realist the mind-independent world has a priority over how we conceptualize, come to know, and convey information about it. It is the belief in this priority that makes Craver underline the ontic mode of scientific explanation (instead of the communicative, representative and cognitive modes) and state that the goodness of an explanation is to be evaluated against how much of the mind-independent world is revealed (2014). It should then go almost without saying that a mechanistic realist who argues for MPR is already committed to an ordered structure of Being: He must be arguing that he first and foremost commits to the objective truthmakers of his posits, and *then* to the models dependent on (grounded in) these. Put differently, he must be holding that the models derive their existence from the existence of the objective structures, making the former derivative and the latter fundamental. In saying these, then, I am not positioning myself against the mechanistic realist with regards to his commitment to the abstract, idealized model (bearing an explanatory indispensability). I am instead asserting that mechanisms as models are *ontologically free lunch*: They figure in our ontology *because of* the mind-independent world in the first place, and because of the mental activity involved in understanding *this world* in the second place. The concrete existents out there in the world could *be* without these abstract objects, but the converse would not be true (given scientific realism).

Sharing a scientific realist stance with the mechanistic philosopher nicely aids in making my response to MPR compact, for the in-built fundamentalism about the mind-independent world. Concerning SR, though, I obviously need an extra argument to convince the mechanistic realist that mechanisms construed as ontic structures are derivative, too. Let me pause here to make one thing clear: Until this chapter, I wrote against taking mechanisms as ontic structures. While I still would like to insist on this (that is, on construing them as not concrete but abstract objects in our ontology) I assume their concrete existence for the sake of the argument. In any case, this much is a terminological issue: My mechanistic non-realism acknowledges that there are concrete, repeatable patterns in the world that can be revealed through mechanistic explanation. I am more inclined to label these as real mechanisms only when the causal systems they make up are actually machine-like (such as a moka pot). But one could be more liberal in his use of the label “real mechanisms”.

In the discussion on machine-likeness in Section 3.4.1, I mentioned that there are non-machine-like systems which may not be explained by mechanistic explanations at all. At several points, I also referred to Craver admitting that not all explanations might be mechanistic. These should entail that, the concrete structure underlain by causal properties might not be covered completely by (ontic) mechanisms. This is to illustrate the *part-whole* relationship obtaining between mechanisms and the

concrete structure, the whole having some other parts that are not mechanisms. Or, even if it might turn out that mechanisms are everywhere, one mechanism being spatiotemporally located next to another one to cover the whole concrete structure of the world, it would still be the true that mechanisms are *parts* of this world. For mechanisms, as we know, are mechanisms for different phenomena, and since the world is comprised of many phenomena, there cannot be one mechanism that would overlap with the whole world. Now, if I can show that the whole has an ontological priority over its parts, then I am done with showing that it is the concrete structure that is fundamental and its parts are derivative.

For this, it should suffice to draw on the *holism* involved in Chakravartty's dispositional realism, and on the *priority monism* endorsed by Schaffer. In a discussion regarding the difference between how dispositions are *identified* and how they are *constituted*, Chakravartty acknowledges that his dispositional identity thesis entails a holism (“with respect to the natures of causal properties”). This is because what identifies a causal property is the dispositions it confers on its bearers to enter into relations with other causal properties. It is therefore the case that “the natures of causal properties taken as a whole are constituted by a vast network of potential relations. The natures of individual properties are thus linked to one another via closed loops of potential relations.” (2009, 140) Of course more needs to be said about this holism to be fair to Chakravartty's exposition, but I contend that this much should be enough to highlight that the concrete structure, as understood within this dispositional realist framework, is not a “mere aggregate” but an “integrated whole”.

These two notions are from Schaffer (2009, 354), who defines them as follows:

Integrated whole: x is an integrated whole =_{df} x grounds each of its proper parts.

Mere aggregate: x is a mere aggregate =_{df} each of x 's proper parts ground x .

What Schaffer intends to do with these definitions is actually to showcase the use of the notion of grounding in ontology. According to him, this is a “natural and intuitive” (p. 375) notion that is a “metaphysical primitive” (p. 376) needed for the ordering of the ontological layers of the world. From here, Schaffer moves on to give a taster of his position (2009, 378—more thoroughly introduced in his 2010):

Priority Monism: There is exactly one substance, the whole concrete cosmos.

where “priority” stands for the rank in the metaphysical ordering. On this account, then, the whole concrete cosmos stands at the most fundamental level of Being, all of its parts being derivative to it.

It would be quite an important task to show how Chakravartty's dispositional realism merges

well with Schaffer's monism, but let it be left to another study. Staying on our topic, I think I have come to a good position to make the following claim: If the mechanistic realist agrees to adopt the dispositional realist stance, then he is urged to accept that the concrete structure is more fundamental than its parts. For the former is an integrated whole, due to how dispositions are identified. Mechanisms must then be seen as (arguably some of) the parts grounded in the concrete structure underlain by causal properties. In the final analysis, the mechanistic realist can freely commit to mechanisms as real features of the world ontically explaining how phenomena are produced or constituted. Yet committing to them, I stress, does not amount to a worthwhile scientific realist position. For by exploiting the notion of explanatory indispensability, it becomes trivial to make existential claims. What is rather required from a scientific realist is giving an account of the ordered structure of the mind-independent world.

On offer here is, obviously, a fundamentalist conception of reality. A thing worthy to question: Is it the same fundamentalism that Psillos discusses? One could think that my fundamentalism *differs* from Psillos' discussants for not championing physicalism, that is, for not attributing an ontological priority to physical properties over, say, biological properties. This would be a right observation, because I favor a categorization where the ontic status of an existent is evaluated against its being concrete or abstract, and its being a part or a whole. This is categorization that makes use of metaphysical notions, and that looks less scientific for not distinguishing different levels of scientific investigation. But also, perhaps, the scientific categorization could be shown to be not incompatible with the metaphysical one: For instance, one could argue that biological properties are the parts of a physical whole, and conclude that the former are the derivative properties. I have not found the chance to think more on this, though, so I could only *guess* at this point that Psillos would view my stance as fundamentalism anyway, despite its first look, for advocating an ontological hierarchy in one way or another.

CONCLUSION

This study comprised a critique of the ontological commitment to mechanisms, and a suggestion to replace the scientific realist attitude towards mechanisms with dispositional realism. Drawing on Chakravartty's work, I indicated how dispositional realism can be put in use for making philosophical sense of scientific mechanism-talk. I concluded the discussion by arguing for a fundamentalism, where the whole concrete causal structure underlain by dispositional properties is given an ontological priority over its parts and its partial representations. Mechanisms can be construed as some parts of the whole causal structure or just as models representing the former. (I am more inclined towards the latter construal, towards taking mechanisms as abstract, idealized models.) In both cases it is possible to take them as real, but due to the meta-ontological argument provided in Chapter 4, this should not entail mechanistic realism. For, it has been argued, no brand of scientific realism can stop at the point when the existents are decided. Scientific realists should be able to make statements regarding the hierarchical structure of the mind-independent world. The point I want to emphasize in this regard is that, my being against mechanistic realism is due to its not involving such a fundamentalist ontological conception. I have argued for the compatibility of mechanistic explanation with dispositional realism, and attempted to show that mechanistic philosophers do have a ground to state that by their commitment to mechanisms, they are actually committing to the existence of causal properties.

To make another point, let me return back to the semantic ambivalence in “mechanism”. Concerning mechanism-talk itself, I proposed a disambiguation of the two senses of the notion on the basis of the number of behaviors machine and causal mechanisms are associated with. Strictly speaking, causal mechanisms must be identified by a single behavior, especially because we expect them to be the explanantia of that very behavior, whereas it is not so with the machine mechanisms. I described the latter as the class of objects which are parts of the investigated whole. What sets an element of a machine mechanism apart from the other parts of the whole is its relevance to the investigated behavior of the whole. Relatedly, given that machine mechanisms can also be wholes, we can distinguish machine mechanisms from the investigated wholes by constituents-talk. (Recall how we distinguish machine mechanisms from machines, the assembly of the bottom and collecting chambers from the moka pot.) Constituents-talk can be an intuitive and efficient way to make some points about machine mechanisms and real machines (artifacts). I conjecture that one is tempted to apply this way of thinking to distinguish *causal* mechanisms from real systems, by way of

interpreting abstraction in mechanistic modeling as a case of constituents-talk—that is, as a case of informed picking of this and that local states of affairs within the system (and *not* picking some others).

It is *not* that this is a wrong interpretation, or that causal mechanisms and real systems cannot be subject to a version of constituents-talk. Nonetheless, regarding mechanistic explanation, abstraction (that is practiced by determining the relevant factors in the system as mechanism components) is not the end of the story. The explanation works by weaving a causal narrative into the class of picked occurrences, and with this narrative we are given an idealization. With this idealization the production of the explanandum phenomenon becomes much more intelligible, due to simplifying the complex network of causal relations down to the local-scale mechanical interactions. Again, it is *not* that no such interaction obtains; I do grant that there are concrete objects actually entering into mechanical relations. However their actual causal organization is not the way they are portrayed by mechanisms as models. The actual causal organization of concern differs from the configuration of components (forming the mechanism) not only for involving more causal intermediaries but also due to the effects of the global-scale properties of the system. The latter is irreducible to (or not translatable into) the mechanistic-causal narrative. Therefore mechanisms themselves are assertions of falsehoods, making their truthfulness questionable.

On the other hand, as I argued upon finishing the last chapter, mechanisms are about the concrete structures that we the scientific realists can confidently be committed to. For these structures are the networks formed by the quantitatively determinate properties that we are in causal contact with. The dispositions these properties confer on their bearers are manifested in the obtaining of a causal setting. Dispositional realists think that these dispositions are occurrent properties of objects, regardless of being manifested or not. For them to be of worthy of scientific interest, they need *not* be manifested frequently (most of the times). But the scientists' theoretical or practical concerns can and do bring them in causal contact with the repeatable patterns the manifestations of these dispositions constitute. The patterns are privileged over others (by the scientific community) because of these concerns, but they exist independently of one's knowledge about them—unlike mechanisms, which depend a lot on human's rational and intellectual capacities. Our limitations, it seems, makes us inclined to regard natural systems as machine-like structures, and/or to explain natural phenomena by mentally decomposing them and weave a causal narrative with certain aspects.

A point to be made in connection with this is due. Throughout the dissertation I only wrote about

causal explanation; and in the case of the Hodgkin-Huxley mathematical model of the action potential, too, I sided with philosophers who assert that this model has a causal (and even, in some sense, a “mechanistic”) content. But without doubt, in modeling real systems, (life) scientists are not limited with mechanisms and causal narratives. They do employ other means and heuristic devices to provide an understanding and intelligibility of the external world. Therefore none of what I say so far should promote a short-sighted look on scientific modeling. It is due to my criticizing the metaphysical implications of a certain brand of causal modeling that I had to constrain myself.

In Chapter 4 I contrasted scientific explanation with metaphysical explanation; this was a point made in acknowledgment of the compromise of truthfulness in scientific modeling. Unfortunately I do not have much to add on top of this, but I can invite my reader to look at the distinction between *depictive* and *denotative* one more time. This distinction is intended to express how models (like mechanisms) can represent the causal structure without being similar to whatever it is out there that they represent. Following Chakravartty, I claimed that models that fictionalize the causal relations between actual objects still make existential claims with regards to the repeatable features of the world. Once more resorting to him, I want to make use of an analogy that Chakravartty draws between the representational contents of scientific models and modern paintings.

Chakravartty asks us to recall Pablo Picasso's famous *Guernica* and consider how the painting can be about reality despite being not a realistic depiction of the actual states of affairs:

Aspects of the work –figures of a bull, a dead baby in the arms of a screaming woman, a speared horse, the broken body of a soldier, etc. – represent various things with greater and lesser degrees of realism. The painting taken as a whole also has representational content; one thing it represents is the rising threat of European fascism . . . In so far as it does this, however, it is not depictive, but merely denotative. It does not furnish much in the way of “description” beyond making an existential “claim” about the presence of a terrifying danger. (2007, 229)

Let us acknowledge that how paintings do represent their subject matter is a problem in philosophy of art as well—but stop right at this point, for the nature of the representation relation (or the explanans-explanandum relation) is an issue that is out of the scope of my study. My space and resources allowed me to critically expand only on the scientific realism of mechanistic philosophers. But if my argument was successful, then I can also hope to have directed the attention to relatively unexplored areas in the philosophical work on mechanisms: Namely, the nature of reality that scientific mechanism-talk is about, and the normative and epistemic content of the denotative product itself, the mechanism.

REFERENCES

- Andersen, H. 2011. Mechanism, Laws, and Regularities. *Philosophy of Science* 78: 325-331.
- Andersen, H. K. 2012. The Case for Regularity in Mechanistic Causal Explanation. *Synthese* 189: 415-432.
- Barros, D. B. 2008. Natural Selection as a Mechanism. *Philosophy of Science* 75: 306-322.
- Baumgartner, M., and A. Gebharter. 2015. Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *The British Journal For The Philosophy of Science* 0: 1-26. (Published online, February 20)
- Beatty, J. 1995. The Evolutionary Contingency Thesis. In *Concepts, Theories, and Rationality in the Biological Sciences*, edited by G. Wolters and J.G. Lennox. Pittsburgh: University of Pittsburgh Press, pp. 45-83.
- Bean, B. P. 2007. The Action Potential in Mammalian Central Neurons. *Nature Reviews Neuroscience* 8: 451-465.
- Bechtel, W. 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York: Routledge.
- Bechtel, W. 2011. Mechanism And Biological Explanation. *Philosophy of Science* 78: 533-557.
- Bechtel, W., and A. Abrahamsen. 2005. Explanation: A Mechanist Alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 421-441.
- Bechtel, W., and A. Abrahamsen. 2010. Dynamic Mechanistic Explanation: Computational Modeling of Circadian Rhythms as an Exemplar For Cognitive Science. *Studies in History and Philosophy of Science Part A* 41: 321-333.
- Bechtel, W., and R. C. Richardson. 1993 [2010]. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA.: The MIT Press.
- Bickle, J. 2006. Reducing Mind to Molecular Pathways: Explicating the Reductionism Implicit in Current Cellular and Molecular Neuroscience. *Synthese* 151: 411-434.
- Bogen, J. 2005. Regularities and Causality; Generalizations and Causal Explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 397-420.
- Bogen, J. 2008. The Hodgkin-Huxley Equations and the Concrete Model: Comments on Craver, Schaffner, and Weber. *Philosophy of Science* 75: 1034-1046.
- Brandon, R. N. 1984. Grene on Mechanism and Reductionism: More Than Just a Side Issue. *PSA* 1984: 345-353.

- Cartwright, N. 1983. *How The Laws of Physics Lie*. Oxford: Clarendon.
- Catterall, W. A. 2000. From Ionic Currents to Molecular Mechanisms: The Structure and Function of Voltage-Gated Sodium Channels. *Neuron* 26: 13-25.
- Chakravartty, A. 2007. *A Metaphysics For Scientific Realism: Knowing the Unobservable*. Cambridge: Cambridge University Press.
- Chakravartty, A. *forthcoming*. Saving the Scientific Phenomena: What Powers Can and Cannot Do. In *Putting Powers to Work*, edited by J. D. Jacobs. Oxford University Press, pp. 14-27.
- Clay, J. R., and L. J. DeFelice. 1983. Relationship Between Membrane Excitability and Single Channel Open-Close Kinetics. *Biophysical Journal* 42: 151-157.
- Colombo, M., S. Hartmann, and R. van Iersel. 2015. Models, Mechanisms and Coherence. *The British Journal For The Philosophy of Science* 66: 181-212.
- Commoner, B. 2014. Unraveling the DNA Myth. Last modified May 30. Accessed June 19, 2015. <http://www.iatp.org/files/Unraveling_the_DNA_Myth_The_Spurious_Foundatio.htm>
- Craver, C. F. 2002. Structures of Scientific Theories. In *The Blackwell Guide to the Philosophy of Science*, edited by P. Machamer and M. Silberstein. Malden: Blackwell, pp. 55-79.
- Craver, C.F. 2006. When Mechanistic Models Explain. *Synthese* 153: 355-376.
- Craver, C.F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Science*. New York: Oxford University Press.
- Craver, C. F. 2008. Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential. *Philosophy of Science* 75: 1022-1033.
- Craver, C. F. 2014. The Ontic Account of Scientific Explanation. In *Explanation in the Special Sciences*, edited by M. I. Kaiser, O. R. Scholz, D. Plenge, and A. Hütteman. Dordrecht: Springer, pp. 27-52.
- Craver, C. F., and W. Bechtel. 2007. Top-Down Causation Without Top-Down Causes. *Biology and Philosophy* 22: 547-563.
- Craver, C. F., and M. I. Kaiser. 2013. Mechanisms and Laws: Clarifying the Debate. In *Mechanism and Causality in Biology and Economics*, edited by H. Chao, S. Chen, and R. L. Millstein. Dordrecht: Springer, pp. 125-146.
- Crick, F. 1988. *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books.
- Cummins, R. 1975. Functional Analysis. *Journal of Philosophy* 72: 741-764.
- Cummins, R. 1983. *The Nature of Psychological Explanation*. Cambridge, MA.: The MIT Press.

- Darden, L. 2007. Mechanisms and Models. In *The Cambridge Companion to the Philosophy of Biology*, edited by D. L. Hull, and M. Ruse. Cambridge: Cambridge University Press, pp. 139-159.
- Darden, L. 2008. Thinking Again About Biological Mechanisms. *Philosophy of Science* 75: 958-969.
- Darden, L., and C. F. Craver. 2002. Strategies in the Interfield Discovery of the Mechanism of Protein Synthesis. *Studies in History and Philosophy of Biological and Biomedical Sciences* 33: 1-28.
- DesAutels, L. 2011. Against Regular and Irregular Characterizations of Mechanisms. *Philosophy of Science* 78: 914-925.
- Dupré, J. Living Causes. *Proceedings of the Aristotelian Society* Supplementary Volume LXXXVII:19-37
- Elgin, M. 2006. There May Be Strict Empirical Laws in Biology, After All. *Biology and Philosophy* 21: 119-134.
- Feyerabend, P. K. 1975. *Against Method: Outline of An Anarchistic Theory of Knowledge*. London: NLB.
- Fleming, A. J. 2002. The Mechanism of Leaf Morphogenesis. *Planta* 216: 17-22.
- Furst, B. 2015. The Heart: Pressure-Propulsion Pump or Organ of Impedance? *Journal of Cardiothoracic and Vascular Anesthesia*. (Available online, February 20)
- Glennan, S. 1996. Mechanisms and the Nature of Causation. *Erkenntnis* 44: 49-71.
- Glennan, S. 2002. Rethinking Mechanistic Explanation. *Philosophy of Science* 69: S342-S353.
- Glennan, S. 2005. Modeling Mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 375-388.
- Glennan, S. 2010. Ephemeral Mechanisms and Historical Explanations. *Erkenntnis* 72: 251-266.
- Godfrey-Smith, P. 2009. Abstractions, Idealizations, and Evolutionary Biology. In *Mapping the Future of Biology: Evolving Concepts and Theories*, edited by A. Barberousse, M. Morange, T. Pradeu. Boston: Springer, pp. 47-55.
- Griffiths, P. 2011. Philosophy of Biology. In *Stanford Encyclopedia of Philosophy (Winter 2014 Edition)*, edited by E. N. Zalta. <<http://plato.stanford.edu/archives/win2014/entries/biology-philosophy>>
- Hempel, C. G. 1965. *Aspects of Scientific Explanation*. New York: Free Press.
- Hempel, C. G. 1967. Scientific Explanation. In *Philosophy of Science Today*, edited by S. Morgenbesser. New York: Basic Books, pp. 79-88.
- Hempel, C.G., and P. Oppenheim. 1948. Studies in the Logic of Explanation. *Philosophy of Science* 15: 135-175.

- Hodgkin, A. L. 1963. The Ionic Basis of Nervous Conduction. *Nobel Lecture (Delivered on December 11)*. Retrieved from <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1963/hodgkin-lecture.pdf> (June 27, 2014)
- Hodgkin, A. L., and B. Katz. 1949. The effect of sodium ions on the electrical activity of the giant axon of the squid. *The Journal of Physiology* 114: 37-77.
- Hodgkin, A. L., and A. F. Huxley 1952. A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve. *The Journal of Physiology* 117: 500-544.
- Huxley, A. F. 1963. The Quantitative Analysis of Excitation and Conduction in Nerve. *Nobel Lecture (Delivered on December 11)*. Retrieved from <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1963/huxley-lecture.pdf> (June 27, 2014)
- Illari, P. 2013. Mechanistic Explanation: Integrating the Ontic and Epistemic. *Erkenntnis* 78: 237-255.
- Illari, P. M., and J. Williamson. 2011. Mechanisms are Real and Local. In *Causality in the Sciences*, edited by P. M. Illari, F. Russo, J. Williamson. Oxford: Oxford University Press, pp. 818-844.
- Illari, P. M., and J. Williamson. 2012. What is a Mechanism? Thinking About Mechanisms Across the Sciences. *European Journal for Philosophy of Science* 2: 119-135.
- Illari and Williamson 2013. In defence of activities. *Journal for General Philosophy of Science* 44: 69-83.
- Jones, M. R. 2005. Idealization and Abstraction: A Framework. In *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences*, edited by M. R. Jones, and N. Cartwright. Amsterdam: Rodopi, pp. 173-217.
- Kauffman, S. A. 1971. Articulation of Parts Explanation in Biology and the Rational Search for Them. In *PSA 1970*, edited by R. C. Buck, and R. S. Cohen. Dordrecht: Reidel, pp. 257-272.
- Kaiser, M., and B. Krickel. *forthcoming*. The Metaphysics of Constitutive Mechanistic Phenomena. *The British Journal for the Philosophy of Science*.
- Katz, L. N. 1954. The Sixty-Ninth Mary Scott Newbold Lecture: The Mechanism of Cardiac Failure. *Circulation* 10: 663-679.
- Kim, J. 1984. Concepts Of Supervenience. *Philosophy and Phenomenological Research* 45: 153-176.
- Kim, J. 1998. *Mind In A Physical World: An Essay On the Mind-Body Problem and Mental Causation*. Cambridge, MA.: The MIT Press.
- Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Leuchtag, H. R. 2008. *Voltage-Sensitive Ion Channels: Biophysics of Molecular Excitability*. Dordrecht: Springer.
- Leuridan, B. 2010. Can Mechanisms Really Replace Laws of Nature? *Philosophy of Science* 77: 317-340.

- Leuridan, B. 2012. Three Problems For the Mutual Manipulability Account of Constitutive Relevance in Mechanisms. *The British Journal For The Philosophy Of Science* 63: 399-427.
- Levy, A. 2013a . Three Kinds of New Mechanism. *Biology & Philosophy* 28: 99-114.
- Levy, A. 2013b. What Was Hodgkin and Huxley's Achievement? *The British Journal For The Philosophy Of Science* 0: 1-24. (Published Online, June 7)
- Levy, A. 2014. Machine-Likeness and Explanation by Decomposition. *Philosophers' Imprint* 14 (6): 1-15.
- Lewis, D. 1986. Causal Explanation. In *Philosophical Papers Volume II*. Oxford: Oxford University Press, pp. 214-240.
- Love, A. C., and M. J. Nathan. *forthcoming*. The Idealization of Causation in Mechanistic Explanation. *Philosophy of Science*. Retrieved from <http://www.du.edu/ahss/philosophy/faculty_staff/nathan_marco.html> (May 11, 2015)
- Machamer, P. 2004. Activities and Causation: The Metaphysics and Epistemology of Mechanisms. *International Studies in the Philosophy of Science* 18: 27-39.
- Machamer, P., L. Darden, and C. F. Craver. 2000. Thinking About Mechanisms. *Philosophy of Science* 67: 1-25.
- McMullin, E. 1985. Galilean Idealization. *Studies in the History and Philosophy of Science* 16: 247-273.
- Meunier, C., and I. Segev. 2002. Playing the Devil's Advocate: Is the Hodgkin–Huxley Model Useful? *Trends in Neurosciences* 25: 558-563.
- Mitchell, P. 1961. Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic Type of Mechanism. *Nature* Jul 8; 191: 144-148.
- Mitchell , S. D. 1997. Pragmatic Laws. *Philosophy of Science* 64: S468-S479.
- Mitchell, S. D. 2000. Dimensions of Scientific Law. *Philosophy of Science* 67: 242-265.
- Nathan, M. J. 2012. Causation by Concentration. *The British Journal For The Philosophy of Science* 0: 1-22. (Published online, April 20)
- Nathan, M. J. 2013. A Simulacrum Account Of Dispositional Properties. *Noûs* 0: 1-22. (Published Online, May 29)
- Nagel, E. 1961. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World.
- Nicholson, D. J. 2012. The Concept of Mechanism In Biology. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 152-163.
- Overton, J. A. 2011. Mechanisms, Types, and Abstractions. *Philosophy of Science* 78: 941-954.

- Ruse, M. 2005. Darwinism and Mechanism: Metaphor in Science. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 285-302.
- Salmon, W. 1984. *Scientific Explanation And The Causal Structure Of The World*. Princeton: Princeton University Press.
- Schaffer, J. 2009. On What Grounds What. In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by D. Manley, D. J. Chalmers, R. Wasserman. Oxford: Oxford University Press, pp. 347-383.
- Schaffer, J. 2010. Monism: The Priority of the Whole. *Philosophical Review* 119: 31-76.
- Shoemaker, S. 1980. Causality and Properties. In *Time and Cause: Essays Presented to Richard Taylor*, edited by P. van Inwagen. Dordrecht: D. Reidel, pp. 109-135.
- Skipper, R. A., and R. L. Millstein 2005. Thinking About Evolutionary Mechanisms: Natural Selection. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 327-347.
- Smart, J. J. C. 1963. *Philosophy and Scientific Realism*. London: Routledge & Kegan Paul.
- Sober, E. 1997. Two Outbreaks of Lawlessness in Recent Philosophy of Biology. *Philosophy of Science* 64: S458-S467.
- Soom, P. 2012. Mechanisms, Determination and the Metaphysics of Neuroscience. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 655-664.
- Steel, D. 2007. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.
- Strassberg, A. F., and L. J. DeFelice. 1993. Limitations of the Hodgkin-Huxley Formalism: Effects of Single Channel Kinetics on Transmembrane Voltage Dynamics. *Neural Computation* 5: 843-855.
- Strevens, M. 2008. *Depth: An Account of Scientific Explanation*. Cambridge, MA.: Harvard University Press.
- Swoyer, C. 1982. The Nature of Natural Laws. *Australian Journal of Philosophy* 60: 203-223.
- Tabery, J. G. 2004. Synthesizing Activities and Interactions in the Concept of a Mechanism. *Philosophy of Science* 71: 1-15.
- Tabery, J. 2014. *Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture*. Cambridge, MA.: The MIT Press.
- Thagard, P. 2003. Pathways to Biomedical Discovery. *Philosophy of Science* 70: 235-254.
- Thompson, P. 2001. Biology. In *A Companion to the Philosophy of Science*, edited by W. H. Newton-Smith. Malden: Blackwell, pp. 16-25.

- Waters, C. K. 1998. Causal Regularities in the Biological World of Contingent Distributions. *Biology and Philosophy* 13: 5-36.
- Weber, M. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.
- Weber, M. 2008. Causes Without Mechanisms. *Philosophy of Science* 75: 995-1007.
- Weisberg, M. 2007. Three Kinds of Idealization. *Journal of Philosophy* 104: 639-659.
- Wimsatt, W. C. 1976. Reductive Explanation: A Functional Account. In *PSA 1974*, edited by R. S. Cohen, C. A. Hooker, A. C. Michalos, and J. W. v. Evra. Dordrecht: Reidel, pp. 671-710.
- Wright, C. D. 2012. Mechanistic Explanation Without the Ontic Conception. *European Journal for Philosophy of Science*: 1-20. (Published online, March 29)
- Woodward, J. 2002. What Is a Mechanism? A Counterfactual Account. *Philosophy of Science* 69: S366-S377.
- Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Woodward, J. 2013. Mechanistic Explanation: Its Scope and Limits. *Proceedings of the Aristotelian Society* Supplementary Volume LXXXVII: 39-65.
- Ylikoski, P. 2013. Causal and Constitutive Explanation Compared. *Erkenntnis* 78: 277-297.