Garson, J. (forthcoming) "Mechanisms, Phenomena, and Functions," in Glennan, S., and Illari, P. (Eds.), *Routledge Handbook of Philosophy of Mechanisms* (London: Routledge).

Mechanisms, Phenomena, and Functions[1]

Justin Garson

1. Introduction

A mechanism is always a mechanism for a phenomenon. The phenomenon that a mechanism serves is not somehow incidental to that mechanism, but constitutive of it: mechanisms are identified, and individuated, by the phenomena they produce. Thus, we can ask about mechanisms for blood clots, or demand-pull inflation, or social cohesion in naked mole rats. But it does not make sense to ask, "how many mechanisms are in the human body?," or even, "is the universe a mechanism?" These latter two questions are nonsensical because they do not specify a relevant phenomenon. That each mechanism has a phenomenon has become a platitude in the new mechanism literature in the philosophy of science (e.g., Glennan 1996, 52; Bechtel and Abrahamsen 2005, 423; Craver 2007, 122; Darden 2008, 960; Craver and Darden 2013, 52) – though as it turns out, there are different ways in which a mechanism "has" its phenomenon.

Sometimes, however, philosophers describe mechanisms not as having phenomena but as serving functions. "Mechanisms are identified and individuated by the activities and entities that constitute them, by their start and finish conditions, and by their functional roles" (Machamer, Darden and Craver 2000, 14). "A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization" (Bechtel and Abrahamsen 2005, 423). "Mechanisms are systems that produce some phenomenon, behavior or function" (Glennan 2010, 256). "The entities and activities that are part of the mechanism are those that are relevant to that function or to the end state, the final product that the mechanism, by its very nature, ultimately produces" (Craver 2013, 141). Sometimes, only parts of mechanisms are described as having functions; sometimes the mechanism as a whole is described as serving a function. In the following I will mainly focus on this latter use (though see Section 3).

The fact that people sometimes describe mechanisms in terms of their *functions*, rather than (or in addition to) their *phenomena*, raises an intriguing question: are these two ways of talking about mechanisms mere terminological variants of one another? Or, when we describe a mechanism as having a function, are we saying something more than, or other than, that it has a phenomenon? If so, what more, precisely, is being said? We can also frame the question from an ontological perspective. Compare the set of all mechanisms and the set of all mechanisms that serve functions. Are these two sets coextensional? Or is the set of mechanisms that have functions a proper subset of the set of mechanisms that have phenomena? In other words, is it that *all* mechanisms have phenomena, but for *some* of those mechanisms, those phenomena happen to be their functions, too?

In the following, I urge the view that there are at least two broad senses of mechanism at play in science and in the philosophy of science. (Technically, these should be thought of as two families of senses of mechanism, since each family may encompass several definitional variants.) According to the first sense of mechanism, which I will call, following Glennan (forthcoming), "minimal mechanisms," mechanisms are defined in terms of their phenomena, but there is no additional implication that these phenomena are in any intuitive sense functions of those mechanisms, where function is aligned with design, purposiveness, goal-directedness, or utility. Using Glennan's example, in this minimal sense, my car is mechanism for locomotion, but it is also a mechanism for melting chocolate bars, even though melting chocolate bars is not one of my car's functions. Its function is just to get me from place to place. (Perhaps it has other legitimate functions, for example, to serve as a status symbol.) In this minimal sense, the aggregation of misfolded proteins is part of the mechanism for Alzheimer's disease, even though misfolded proteins do not have the function of producing Alzheimer's disease.

The second sense of mechanism is what I call, following Garson (2013), the functional sense of mechanism. Here, mechanisms are identified by the functions they serve, where "function" is understood as having a connotation of teleology, purposiveness, or design – though as we will see, working out precisely what this sense of "function" amounts to is beset with controversy. Cast in an ontological vein, I urge that the class of functional mechanisms is an interesting proper subset of the class of minimal mechanisms. In this sense of mechanism, a car typically would not be described as a mechanism for melting chocolate bars, even though it does so frequently enough and even though we can explain how that works. The heart would not usually be considered a mechanism for causing hemorrhages; it is a mechanism for pumping blood, even though it does both of those things and we understand how both processes work.

One virtue of recognizing a distinct, functional sense of mechanism is that it helps us think about the normativity of mechanisms, that is, it helps us understand how mechanisms can break (see Section 4). A mechanism breaks when it cannot perform the function that defines it. In other words, the class of functional mechanisms is coextensive with the class of mechanisms that can break. It is not clear whether the minimal sense of mechanism has the resources to explain how a mechanism can break. A second virtue is that it helps us think about the distinction between a mechanism's target and its byproduct (e.g., Craver and Darden 2013, 69), for example, locomotion and melting chocolate bars in the case of the car. The target/byproduct distinction maps neatly onto the traditional function/accident distinction. A mechanism's target is just its function; a byproduct is any effect that is incidental to its discharging its function (in the functions literature, an "accident"). A third virtue of this functional sense of mechanism is that it helps us organize biomedical knowledge well (see Section 4; also see Chapter 25).

In the following, I will do four things. First, I will provide an overview of the different ways that a mechanism can have a phenomenon (Section 2). Second, I will provide a survey of what various philosophers have had to say about a mechanism's function (Section 3). Even when philosophers agree that there is some intimate connection between mechanisms and functions, they disagree about the nature of that relationship and about the underlying concept of function. (This is why I say that the functional sense of mechanism is actually a family of senses of mechanism.) Third, I will discuss some benefits of recognizing a distinct, functional sense of mechanism (Section 4). In closing, I will return to the problem I started with: what is the relationship between these two ways of talking about mechanisms? The most economical solution is to think of functional mechanisms as constituting an interesting subset of minimal mechanisms.

2. How Mechanisms Have Phenomena

There are two, and arguably only two, ways that a mechanism has a phenomenon. First, a mechanism can have a phenomenon in the sense that it causes the phenomenon to occur. That is, the phenomenon is the effect of the mechanism's operation. Craver and Darden (2013, 65) refer to this as the "producing" relation. Second, a mechanism can have a phenomenon in the sense that the operation of the mechanism somehow constitutes, or realizes, the phenomenon. Craver and Darden (ibid.) refer to this as the "underlying" relation (though there are open questions about what, precisely, this constitution relation amounts to; see, e.g., Couch 2011; Romero 2015; Kaiser and Krickel 2016). The fact that mechanisms can have a phenomenon in one of these two ways gives rise to two different styles of explanation, etiological mechanistic explanation and constitutive mechanistic explanation (see Craver 2001, 69). Kästner (in prep) explicitly makes this connection between the distinction between producing/underlying, on the one hand, and the distinction between etiological/constitutive explanations, on the other. I will describe each of these relations in turn.

Consider the genetic mechanisms involved in cystic fibrosis (for now we will discuss cystic fibrosis as "having" a mechanism or mechanisms, rather than arising from the breakdown of a mechanism). Cystic fibrosis is characterized by debilitating and even fatal respiratory blockages, among other problems. When we describe the genetic mechanism involved in the respiratory blockade characteristic of this disease, we typically do so by describing a certain cause-and-effect sequence. Although there are many biochemical pathways underlying cystic fibrosis, one very common pathway begins with the deletion of three nucleotides in a certain segment of DNA (the delta F-508 mutation in the CFTR gene). This causes the loss of an amino acid, phenylalanine, in the corresponding protein sequence (the CFTR protein). This makes the protein misfold, which, in turn, disrupts the normal passage of chloride across cell membranes. This can lead to a buildup of sticky mucus in the lungs and elsewhere, though there is still some uncertainty about the precise pathways involved. In this case, the delta F-508 mutation is part of a mechanism for the disruption of chloride transportation in this first, "producing" sense. Disruption of chloride transport is a late stage in a long and complex cause-and-effect sequence.

In contrast, consider the mechanism for the patellar (knee-jerk) reflex. A tap to the patellar tendon causes a muscle spindle in the quadriceps muscle to stretch, which sends a sensory signal to the ventral horn of the spinal cord. The sensory neuron synapses directly onto a motor neuron, which sends a command back to the quadriceps to contract, causing the leg to kick. (This reflex is an example of a "monosynaptic" reflex because it only involves a single synapse.) Importantly, this mechanism does not *cause* the patellar reflex. The reflex is not the terminal stage of a cause-and-effect sequence. Rather, the patellar reflex is constituted by the whole sequence of events. When one explains how this mechanism works, one offers a constitutive etiological explanation. (In contrast, if one wants to describe the mechanism that causes the kick, we are back to the producing relation, since kicking is the terminal stage of a causal sequence that begins with tapping the patellar tendon.)

Craver and Darden (2013) consider a third potential way that a mechanism can have a phenomenon, which they call the "maintaining" relationship. This is exemplified by the maintenance of homeostatic set points, like the water level of the blood. Consider the mechanism for dopamine homeostasis. Mesolimbic dopamine neurons originate in the ventral tegmental area of the midbrain and terminate in the nucleus accumbens. They play a role in mediating reward, although the precise function of these dopamine signals is unclear (Berridge and O'Doherty 2013). These neurons utilize a homeostatic mechanism for maintaining a more-or-less constant level of dopamine in the synapse, usually at concentrations of between 20-50nM (Grace 2000, 336). To do this, the dopamine neuron's axon terminal has autoreceptors to monitor the level of dopamine outside the cell. When the level rises significantly above this "set point," the autoreceptor sends a signal to decrease the synthesis of dopamine (by inhibiting the enzyme tyrosine hydroxylase). When it drops significantly below this level, the autoreceptor activates that same enzyme and causes more dopamine to be synthesized inside the neuron. Sometimes, scientists wish to know how such "set points" are maintained over the long run in the face of fairly regular perturbations.

On reflection, however, it is not clear that we need to recognize a third, distinct, category for how mechanisms "have" phenomena in order to make sense of homeostatic phenomena like dopamine homeostasis. Rather, dopamine homeostasis is just another sort of phenomenon that we might wish to give a mechanistic explanation for. When we do provide a mechanistic explanation, we will either identify a producing relation, or an underlying relation, depending on the details of our analysis (Kästner in prep). Plausibly, if you were to ask, “how does the dopamine neuron maintain a fairly steady extracellular dopamine level despite frequent disruptions?” and I tell you about the dopamine neuron’s autoreceptors and how they regulate tyrosine hydroxylase, I am describing the mechanism that underlies dopamine homeostasis, and giving a constitutive mechanistic explanation. If, instead, I were to give you a developmental account of how genetic and early environmental factors interact to create dopamine homeostasis, I am providing an etiological mechanistic explanation.

It would be easy to form the impression that when scientists pursue a mechanistic explanation, they first fix the phenomenon clearly and then discover its mechanism. But things are rarely that simple. Rather, there is often a back-and-forth movement between the way we describe a phenomenon, and the way we describe its mechanism. This movement is important for documenting the role of mechanisms in the process of scientific discovery (see Chapter 19). This is what Bechtel and Richardson (1993, 173) call “reconstituting the phenomenon;” Glennan (2005) and Craver and Darden (2013, Chapter 4) also discuss it. Sometimes in the history of science, scientists have a certain conception of the phenomenon they are after, and they start to identify the mechanisms underlying it. What they discover about those mechanisms changes how they think about the phenomenon itself. Consider how the concept of memory has changed over the last 50 years because of advances in psychology and neuroscience (Bechtel 2008). Or consider how the notion of schizophrenia has changed over the last century. Garson (forthcoming) describes how the study of recreational amphetamine use in the United States helped to "reconstitute the phenomenon" of schizophrenia in the 1960s and 1970s. The Swiss psychiatrist Eugen Bleuler described the fundamental phenomenon as a "loosening of associations;" today, most Western psychiatrists would consider this "loosening of associations" as representative of, at best, one subtype of schizophrenia, as a result of ongoing research.

3. Mechanisms and Functions

Several philosophers have recognized an intimate connection between mechanisms and functions, where "function" has the connotation of teleology, design, or purpose. Consider a statement like, "the function of zebra stripes is to deter biting flies," or "the function of eyespots on butterfly wings is to deflect attack away from vital organs." Here, "function" connotes teleology or purpose: deflecting attack is in some loose sense what the eyespots are "there for." Philosophers of biology have developed many different theories about what functions are (Garson 2016). The point I wish to make here is that some philosophers and scientists have explicitly defined the notion of mechanism in terms of a rich, teleological notion of function. These philosophers include Craver (2001; 2013), Piccinini (2010), Piccinini and Craver (2011), Moss (2012), Garson (2013), and Rosenberg (2015), though they disagree about the precise relationship between mechanisms and functions and they disagree about what functions themselves are.

Why would anyone want to tie functions and mechanisms together in this way? One way to motivate this narrower, functional, sense of mechanism is by bald appeal to intuition. In most everyday contexts, it seems natural to say that a car is a mechanism for locomotion. In contrast, it seems strange to say that a car is a mechanism for melting chocolate bars. It seems equally strange to say that a car is a mechanism for spewing carbon dioxide. But why does it seem strange? All three of those phenomena (locomotion, melting chocolate bars, emitting carbon dioxide) are very common, and they are completely explicable in terms of physics and engineering. In everyday contexts, however, only the first description seems natural or normal; the latter two sound unusual or strained. Why?

The same point can be made in the biological context instead of the context of artifacts and tools. In the minimal sense of mechanism, the heart is a mechanism for pumping blood, and it is also a mechanism for making beating sounds that one can listen to through a stethoscope. But we typically say that the heart is a mechanism for pumping blood. It sounds strange to say that it is a mechanism for making beating sounds, even though it does both and both are entirely explicable in terms of basic physics and biology. What is going on here?

One way of articulating this intuition (rather than explaining it) is in terms of the distinction between a mechanism's target and its byproduct (Craver and Darden 2013, 69). In the case of the car, it seems natural enough to identify locomotion as the mechanism's target and the emission of CO_2 as a byproduct. In the case of the heart, it seems natural enough to identify the mechanism's target as pumping blood and the beating sounds as a byproduct. The fact that these attributions are fairly stable in most everyday contexts raises the question of why we speak this way.

This is where functions come in. An obvious difference between locomotion and melting chocolate bars (or spewing CO_2) is that locomotion is the car's *function*. That is what cars are designed to do. Emitting CO_2 is a (rather unfortunate) byproduct of its function. That is not why cars exist. A similar point can be made about the heart. The reason the heart is a mechanism for pumping blood, rather than making beating sounds, is that pumping blood is the heart's function. That is what the heart is for. Perhaps there are some contexts where people are willing to say that the heart is a mechanism for making heart sounds (for example, in a diagnostic context). But to the extent that there is some context-dependency in the way people talk about mechanisms, that might be explicable in terms of the context-dependency in the way people talk about functions. Quite generally, according to this functional sense of mechanism, in order for *X* to be (part of) a mechanism for *Y*, *Y* must be its function. As I will show in Section 4, this construal of mechanism also helps us understand the normativity of mechanisms, that is, how mechanisms break.

Suppose we accept that, in some contexts, mechanisms are defined in terms of the function they serve. That raises an obvious question: what are functions? Philosophers have explored this question for the last forty years with little agreement (see Garson 2016 for a recent overview). There are three mainstream theories of function on the market, the selected effects (SE) theory, the causal role (CR) theory, and the biostatistical theory (BST), though this list is not exhaustive. Roughly, according to the SE theory, the function of a trait is the reason it evolved by natural selection. The function of the heart is to pump blood, rather than make beating sounds, because it evolved by natural selection for pumping blood. BST characterizes a trait's function in terms of its current-day contribution to survival and fitness, rather than in terms of selection history. Like SE, BST defines function in terms of evolutionary considerations (that is, present-day fitness) but remains neutral about history. According to the CR theory, the function of a system *part* consists in its contribution to some system capacity that an investigator has picked out as especially worthy of attention. Unlike the other views, CR emphasizes the contextual and perspective-dependent aspects of functions.

I am not interested, in this place, in characterizing these theories of function more rigorously or discussing alternatives. Rather, the point I wish to make is that people might think about mechanisms slightly differently depending on how they think about functions. As Piccinini (2010, 286) put the point, "different notions of mechanism may be generated by employing different notions of function." That is why I say that the functional sense of mechanism is really a family of different senses of mechanism. We can talk about the SE-functional sense of mechanism, the CR-functional sense of mechanism, and so on.

For example, one fairly austere way of thinking about mechanisms is to define mechanism in terms of function, and then define functions in terms of selected effects (that is, using the SE theory of function). Some philosophers and scientists have done just that. This is a fairly narrow way of thinking about mechanisms since it implies that in order for *X* to be a mechanism for *Y*, *X* must have been shaped by natural selection for doing *Y*. We can refer to this as the *SE-functional sense of mechanism*. In this sense, the heart is a mechanism for pumping blood, and not making beating sounds, because it was shaped by natural selection for pumping blood.

In the 1960s, the evolutionary theorist G. C. Williams (1966, 9) recommended this way of talking about mechanisms. As he put it, "the designation of something as a *means* or *mechanism* for a certain *goal* or *purpose* will imply that the machinery involved was fashioned by selection for the goal attributed to it." Here, mechanisms are defined in terms of function ("purpose"), which, in turn, is defined in terms of natural selection. Williams thought there was something deeply counterintuitive about describing something as a "mechanism" for an effect that it was not plausibly selected for. "Should we therefore regard the paws of a fox as a mechanism for constructing paths through snow? Clearly we should not (13)." By the same token, I suppose that Williams would also be loath to say that, "a car is a mechanism for melting chocolate bars," or that, "the heart is a mechanism for causing hemorrhages."

Some of the contemporary evolutionary psychologists have adopted Williams' usage in the way that they characterize psychological mechanisms (see Garson 2013, 322 for discussion). Among philosophers, Rosenberg (2015) seems to endorse this SE-functional sense of mechanism. In (Garson 2013) I endorse the functional sense of mechanism, but do not specify which theory of function is correct. However, I emphasize (p. 319) that function should be defined somewhat narrowly in terms of selection, fitness, or design.

A reservation one might have about the SE-functional sense of mechanism is that it might be *too* narrow, that is, it might fail to account for the wide spectrum of mechanisms out there. For example, sometimes neuroscientists talk about neural "mechanisms" for various activities, even when they do not think those activities are their evolved functions (Craver 2013, 141). It also would not apply to mechanisms outside of the contexts of biology and engineering, such as in physics, chemistry, or geology. As Glennan (2005, 445) points out, one may wish to talk about mechanisms underlying the eruption of geysers, even though geysers have no evolved functions. However, there is a broader sense of function that can encompass these diverse contexts. This is the CR theory of function, to be discussed shortly.

Craver (2001; 2013) explored the connection between mechanisms and functions in substantial detail. He says that mechanisms are defined and individuated in terms of the functions they serve. As he puts it, "the entities and activities that are part of the mechanism are those that are relevant to that function or to the end state, the final product that the mechanism, by its very nature, ultimately produces" (2013, 141). Here, the "function" of a mechanism is equated with its "end state" or "final product." This way of thinking about mechanisms also makes an appearance in Machamer, Darden and Craver's well-known paper, which states that, "to the extent that the activity of a mechanism as a whole contributes to something in a context that is taken to be antecedently important, vital, or otherwise significant, that activity too can be thought of as the (or a) function of the mechanism as a whole" (2000, 6).

However, instead of embracing the SE theory of function, Craver accepts the CR theory of function. We can label the whole package of ideas the *CR-functional sense of mechanism*. One distinctive feature of CR functions is that they have an explicitly perspectival, or contextual, character. For Craver, the function of the "uppermost" system in a mechanistic analysis is simply some capacity that a researcher or research team has found especially worthy of attention. To say the system has a function does not imply that it has an intrinsic goal or purpose. To the extent that CR-functions can be associated with teleology or purpose, they have a *derived* teleology and *derived* purpose – derived, that is, from the interests and goals of the researchers who investigate them. As Craver puts it, "This teleological feature of mechanistic description is also implicit in the fact that mechanisms such as the NMDA receptor are bounded: a judgment has been made about which entities, activities, and organizational features are in the mechanisms and which are not" (Ibid., 140). Craver's view about mechanisms raises complex questions about realism and antirealism about mechanisms (see Chapter 9).

One of the benefits of Craver's expansive way of thinking about mechanisms is that it allows us to make sense of mechanism-talk outside of the contexts of biology and engineering (for more on the topic of mechanisms in engineering, see Chapter 35). Nothing prevents us from giving a CR-functional analysis of, say, El Niño phenomena, or demand-pull inflation, or even the way that atoms aggregate into molecules.

Lenny Moss (2012) also thinks there is a tight connection between mechanisms and functions (or "goals"). He points out that, often in the life sciences, biologists only attribute mechanisms to systems that are in some sense goal-directed: "To count as a biological mechanism the phenomenon in question thus must be perceived as being an expression of the ostensible 'purposiveness' of the living cell or organism" (165). Note that to say that an organism or biological system is "goal-directed" is different from saying that it has functions (see Garson 2016, Chapter 2 for discussion). The point here is that Moss agrees with Craver (2013) and Garson (2013) that mechanisms have a teleological dimension. Part of Moss' evidence for this view is that, in many cases, we would refrain from calling something a mechanism if it were merely an artifact of laboratory procedure and not somehow expressive of an organism's goal-directedness: "if cells stick to tissue culture plastic because of a chemical reaction with the plastic that resulted in the happenstance chemical production of an epoxy resin this too would be registered as an artifact and not as a 'mechanism of adhesion'….[I]t is strictly the teleological aspect which makes the difference" (ibid.)

There is a potential source of ambiguity that comes up when we discuss the relation between mechanisms and functions. There is a difference between attributing a function to a mechanism, qua whole, and attributing functions to the mechanism's parts. Up until now, the philosophers that I have discussed hold that, in order for *X* to be a mechanism for *Y*, *Y* must be *X*'s function (or at least, *X* must be part of a goal-directed system, as in Moss's view). In other words, they see the mechanism as a whole as the thing that has a function, purpose, or goal. Other theorists tend to attribute functions, first and foremost, to the parts of mechanisms, rather than (or in addition to) the whole mechanism. This seems to be a fairly common stance. For example, Craver (2001) emphasizes the way that parts of mechanisms have functions. He refers to these part-functions as "mechanistic role functions." At times, Bechtel and Richardson (1993) also use "function" specifically when they are describing the activities of a mechanism's *parts*, rather than a mechanism's phenomenon. For example, they define "mechanistic explanations" as explanations that "account for the behavior of a system in terms of the functions performed by its parts and the interactions between these parts (17)." Piccinini and Craver (2011) also describe the way that mechanisms have functions, but if I understand them correctly, they typically focus on the way that parts of mechanisms have functions, rather than the whole mechanism.

Craver (2001) makes an interesting distinction regarding these mechanistic role functions. He observes that when we attribute a function to a component of a system, we can point to the item's activity without making any reference to the role that it plays in the broader system, or we can emphasize its role, irrespective of the specific activity by which it performs this role. A simple example will clarify the distinction. Consider the function of the heart. On the one hand, we can describe the heart as having the function of *beating*. This just describes what it does irrespective of the contribution it makes to the system (circulating blood). On the other hand, we can describe the heart as having the function of *helping to circulate blood*. This describes the role it plays in the circulatory system without telling us the specific activity it performs (beating) Craver (2001, 65) refers to these two roles as the item's "isolated activity" and "contextual role" respectively. In his view, "a complete description of an item's role would describe each of these."

There is one further source of ambiguity that arises when we think about functions. I suspect that some writers in the new mechanism tradition use the term "function," purely synonymously with "phenomenon." That is, in some cases I suspect that "function" represents nothing more than a terminological variation on "phenomenon" or "behavior," and hence the class of mechanisms that have "functions" in this weak sense is coextensional with the class of what Glennan calls "minimal mechanisms." For example, Glennan (2002, 127, fn. 6) notes that, "it is tempting to use 'function' in place of the term 'behavior'" Ultimately, he suggests there that we resist that temptation, because 'function' may carry inappropriate connotations of design, Bechtel and Abrahamsen (2005, 433), similarly, characterize function in terms of the familiar function/structure distinction, which is family devoid of connotations of teleology or purpose. However, I see a potential problem here. As I will describe in the next section, we often want to describe a mechanism as "broken," and one good way to make sense of that is by appealing to function, in some appropriately restrained sense of that term.

4. When Mechanisms Break

Biologists and biomedical researchers employ a colorful vocabulary to talk about how mechanisms break. A mechanism can have a "breakdown." It can be "usurped," "coopted," or "hijacked." It can be "interfered with," "disrupted," "impaired," or "disabled." It can simply "fail." For example, "drugs of abuse can hijack synaptic plasticity mechanisms in key brain circuits" (Kauer and Malenka 2007, 844). "Potentially irreversible impairments of synaptic memory mechanisms in these brain regions are likely to precede neurodegenerative changes…" (Rowan et al. 2003, 821). The idea that mechanisms can break plays an important role in the way that biomedical researchers organize knowledge. Many philosophers in the new mechanism tradition have also emphasized how the idea of a mechanism's breaking is crucial for understanding causation, identifying components of mechanisms, and explaining disease (e.g., Bechtel and Richardson 1993, 19; Craver 2001, 72; Glennan 2005, 448; Darden 2006, 259.)

So, what is it for a mechanism to break? At first pass, it is tempting to say that a mechanism breaks just when it cannot yield the phenomenon that defines it. My truck is a mechanism for locomotion, but because of a rusty spark plug, it cannot start. Sophia's immune system is a mechanism for fighting off harmful pathogens, but because of a low white blood cell count (leukopenia) due to a virus, it cannot do that (or not as effectively as usual). It is broken.

But this way of speaking raises a deeper question. Suppose my truck cannot start. Why would we say that the truck is a broken mechanism for locomotion, rather than simply that it is not currently (or no longer) a mechanism for locomotion? What license do we have (no pun intended) to say that my truck is a mechanism for locomotion, that is, that locomotion is its phenomenon, even when it cannot actually yield this phenomenon? Why not just say it is no longer a mechanism for locomotion?

Indeed, some of the standard definitions of mechanism, read strictly, do not allow us to speak of broken mechanisms. The set of entities and activities constituting my broken truck are *not* "organized such that they are productive of regular changes [that is, those constituting locomotion]…" My busted truck is not a "complex system that produces the behavior [of locomotion] by the interaction of a number of parts…" This suggests that we should find some way of loosening up those definitions to allow my truck to be a broken mechanism for locomotion, rather than simply not a mechanism for locomotion at all.

There are a few ways to respond to this situation. One way is to bite the bullet and say that, strictly speaking, there is no such thing as a broken mechanism. For example, in Glennan's minimal sense of mechanism, one would have to say that my broken truck is not a mechanism for locomotion, since it cannot actually perform that activity (though it might still be a mechanism for something else such as melting chocolate bars). This maneuver has the disadvantage that it seems to run against much of biomedical usage, as indicated in the quotations that open this section. Another way of accounting for the possibility of broken mechanisms is to explain it in terms of a purely statistical or epistemic norm. For example, one might hold that, when we say that the truck is a "broken" mechanism, all we mean is that it is acting in a way that is atypical or unexpected. This is a step in the right direction, but we should not equate brokenness with atypical or unexpected behavior. The medulla oblongata is a mechanism for helping us breathe, and it is also a mechanism for initiating the gag reflex. The latter is atypical, and perhaps unexpected, but one would not want to say it is broken when it causes us to gag.

My suggestion here is that a mechanism breaks when it cannot perform the function that defines it. The function of my car is locomotion, but it cannot perform that function because of a rusty spark plug, so it is broken. The function of Sophia's immune system is to fight pathogens, but it cannot do so because of an unusually low white blood count so it is broken. Perhaps there are other ways of explaining what it is for a mechanism to break; my only claim is that this is an obvious option. I also do not want to restrict functions to selected effects or design (as in artifacts); as I indicated in the last section, one might construe function broadly in terms of the CR theory of function. My point here is that when we say that a mechanism is broken, we seem to imply that it has some function that it is unable to perform, even if we disagree about precisely what functions are.

Why are functions uniquely suited to this role? Because functions, at least on standard analyses, have a normative character. To say that functions are "normative" just means it is possible for a token system to *have* a function that it cannot *perform*. Functions have the remarkable property that they linger, as it were, even in the absence of the corresponding capacity. Standard analyses of function seem to provide a reasonable explication of this normative character, though there is ongoing debate about how exactly the different theories of function should account for normativity. For example, on the standard SE theory, the function of a trait is the reason it evolved by natural selection. So, the function of a trait has to do with its history, rather than its current-day capacities. As a consequence, it is easy to understand how a trait can possess a function (owing to its history) that it is nonetheless unable to perform.

Mechanisms, too, at least in standard biological usage, have a normative character. Mechanisms are the sorts of things that can break. Put simply, it is possible for *X* to be a mechanism for *Y* even when *X* cannot do *Y*. In my view, the normativity of mechanisms derives from the normativity of functions. On reflection, this is not such a radical philosophical move. Philosophers of biology have often appealed to the normativity of functions in order to make sense of the normative character of other biological categories. Dretske (1986) and Neander (1995) appeal to the normativity of function in order to make sense of the normative character of representations. Neander (1991) and Rosenberg and Neander (2009) appeal to the normativity of function in order to make sense of the way that types of traits are individuated. Appealing to function is a natural, even obvious, way to think about the normative character of mechanisms.

The idea that mechanisms serve functions not only helps us think about how mechanisms break, but it is useful for organizing biomedical knowledge. Suppose we accept that mechanisms serve functions, in some rich, teleological sense of that term. Then, we (generally) would be prohibited from talking about mechanisms for disease. For example, in the functional sense, there is no mechanism for anencephaly. Rather, anencephaly is an explicable result of the breakdown of a mechanism for neurulation. Moghaddam-Taaheri (2011, 608-10) makes a similar point in her discussion of medicine, though she does not situate this insight within a general theory of mechanism. Neander (forthcoming, Chapter 3) argues that diseases are generally best explained in terms of deviations from proper function, though she does not specifically relate this point to the concept of mechanism.

Let me clarify the point by contrasting two different ways of talking about diseases. On the first way of talking about disease, all diseases have mechanisms. There is a mechanism for spida bifida, and anencephaly, and encephalocele, amongst others. Biomedical research would be the organized attempt to explicate the mechanisms for these diverse diseases, along with their spatial, temporal, and hierarchical features. On the second way of talking about disease, diseases generally do not have mechanisms. Rather, there are only mechanisms for functions. Here, there is only one functional mechanism involved, namely, the mechanism for neural tube formation. Different diseases result when this mechanism breaks in various ways.

Biomedical researches often adopt this latter way of speaking, though see Garson (2013) for important qualifications. The reason is that the latter way of speaking is highly informative, and it also provides a useful heuristic for discovery. First, when I describe anencephaly as the result of a breakdown in a mechanism for neurulation, rather than as having its own mechanism, I convey critical information about its etiology. I am guiding you to the root problem, as it were, underlying anencephaly. Second, I set up a heuristic for future biomedical discoveries. Anencephaly results from disrupting neural tube folding at the *anterior* neuropore. What happens if folding is disrupted at the *posterior* neuropore instead? These are the kinds of questions that come up when we frame diseases in terms of broken mechanisms.

In conclusion, I accept a form of mechanism pluralism. There are different concepts of mechanism at play in biology and in philosophy, and my goal is not to argue for the unique correctness or superiority of a single one. In one sense (Glennan's "minimal mechanisms"), the notion of mechanism serves as a way of understanding causation and constitution quite generally, even outside of biology and engineering. For example, it can help us to understand mechanism-talk in areas such as physics, chemistry, and geology, where talk about functions is generally out of place. In another sense (the "functional" sense), mechanisms are defined in terms of the functions they serve, where function is thought of in a suitably rich teleological sense. Mechanisms in this functional sense constitute an interesting subset of mechanisms in this minimal sense. Perhaps there are other senses of mechanism as well (see Chapter 7). This functional sense of mechanism can help us understand certain aspects of mechanisms, such as the distinction between a mechanism's target and byproduct, the normativity of mechanisms, and the role of mechanisms in biomedical research.

Note

1. I am grateful to several people for discussions that helped me develop the ideas in this chapter. These include Carl Craver, Lindley Darden, Lenny Moss, Karen Neander, Gualtiero Piccinini, Anya Plutynski, and Sahotra Sarkar. I am also grateful to Phyllis

Illari and Stuart Glennan for inviting me to contribute a chapter to this volume and for their valuable editorial comments.

References

Bechtel, W. (2008) *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*, New York: Routledge.

Bechtel, W., and Abrahamsen, A. (2005) "Explanation: A Mechanist Alternative," *Studies in the History and Philosophy of Biological and Biomedical Sciences* 36: 412–441.

Bechtel, William, and Richardson, R. C. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton, NJ: Princeton University Press.

Berridge, K. C., and O'Doherty, J. P. (2013) "From Experienced Utility to Decision Utility," in P. W. Glimcher, and Fehr, E. (Eds.) *Neuroeconomics: Decision Making and the Brain, 2nd ed.* Amsterdam: Elsevier, 335-351.

Couch, M. B. (2011) "Mechanisms and Constitutive Relevance," *Synthese* 83: 375–388.

Craver, Carl F. (2001) “Role Functions, Mechanism, and Hierarchy,” *Philosophy of Science* 68:53–74.

———. (2007) *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, New York: Oxford University Press.

———. 2013. “Functions and Mechanisms: A Perspectivalist View.” In P. Huneman (ed.) *Functions: Selection and Mechanisms*. Dordrecht: Springer, 133-158.

Craver, C. F., and Darden, L. (2013) *In Search of Mechanisms: Discoveries Across the Life Sciences*, Chicago: University of Chicago Press.

Darden, Lindley. (2006) *Reasoning in Biological Discoveries*, Cambridge: Cambridge University Press.

———. (2008) “Thinking Again about Biological Mechanisms,” *Philosophy of Science* 75: 958–969.

Dretske, Fred. (1986) “Misrepresentation.” In R. Bogdan (ed.), *Belief: Form, Content, and Function.* Oxford: Clarendon, 17–36.

Garson, J. (2013) “The Functional Sense of Mechanism,” *Philosophy of Science* 80: 317-333.

———. (2016) *A Critical Overview of Biological Functions*, Dordrecht: Springer.

———. (forthcoming) "'A Model Schizophrenia': Amphetamine Psychosis and the Transformation of American Psychiatry." In S. Casper and D. Gavrus (eds.), *Technique in the History of the Brain and Mind Sciences*. Rochester: University of Rochester Press.

Glennan, Stuart. (1996) "Mechanisms and the Nature of Causation," Erkenntnis 44:49–71.

———. (2002) "Contextual Unanimity and the Units of Selection Problem," *Philosophy of Science* 69:118–37.

———. (2005) "Modeling Mechanisms," *Studies in History and Philosophy of Biological and Biomedical Sciences* 36:443–64.

———. (2010) "Ephemeral Mechanisms and Historical Explanation," *Erkenntnis* 72:251–266.

———. (forthcoming) *The New Mechanical Philosophy*. Oxford: Oxford University Press

Grace. A. A. (2000) “Gating of Information Flow within the Limbic System and the Pathophysiology of Schizophrenia,” *Brain Research Reviews* 31: 330-341.

Kaiser, M. I., and Krickel, B. 2016. “The Metaphysics of Constitutive Mechanistic Phenomena,” *British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axv058

Kästner, L. (in prep) “On the Curious Trinity of Mechanisms.”

Kauer, Julie A., and Robert C. Malenka. (2007) “Synaptic Plasticity and Addiction,” *Nature Reviews Neuroscience* 8:844–58.

Machamer, Peter, Lindley Darden, and Carl F. Craver. (2000) “Thinking about Mechanisms,” *Philosophy of Science* 67:1–25.

Moghaddam-Taaheri, Sara. (2011) “Understanding Pathology in the Context of Physiological Mechanisms: The Practicality of a Broken-Normal View,” *Biology and Philosophy* 26:603–11.

Moss, Lenny. (2012) “Is the Philosophy of Mechanism Philosophy Enough?” *Studies in History and Philosophy of Biological and Biomedical Sciences* 43:164–72.

Neander, Karen. (1991) “Functions as Selected Effects: The Conceptual Analysts’ Defense,” *Philosophy of Science* 58:168–84.

———. (1995) “Misrepresenting and Malfunctioning,” *Philosophical Studies* 79:109-141.

———. (Forthcoming) *Informational Teleosemantics,* Cambridge, MA: MIT Press.

Piccinini, Gualtiero. (2010) “The Mind as Neural Software? Understanding Functionalism, Computationalism, and Functional Computationalism,” *Philosophy and Phenomenological Research 81*:269–311.

Piccinini, G., and C. Craver. (2011) “Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches,” *Synthese* 183: 283–311.

Romero, F. (2015) “Why There isn’t Inter-Level Causation in Mechanisms,” *Synthese* doi: 10.1007/s11229-015-0718-0.

Rosenberg, A. (2015) “Making Mechanisms Interesting,” *Synthese* doi: 10.1007/s11229-015-0713-5.

Rosenberg, Alex, and Karen Neander. (2009) “Are Homologies Selected Effect or Causal Role Function Free?” *Philosophy of Science* 76:307–34.

Rowan, M. J., I. Klyubin, W. K. Cullen, and R. Anwyl. (2003) “Synaptic Plasticity in Animal Models of Early Alzheimer’s Disease,” *Philosophical Transactions of the Royal Society of London B* 358:821–28.

Williams, George C. (1966) *Adaptation and Natural Selection*, Princeton, NJ: Princeton University Press.