Beate Krickel

The Mechanical World

The Metaphysical Commitments of the New Mechanistic Approach

Springer

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

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

Library of Congress Control Number: 2018961738

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

Contents

Chapter 3
Types of Mechanisms: Ephemeral, Regular, Functional

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

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

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

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

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

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

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

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

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

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

In the following sections I introduce a taxonomy of mechanisms that goes beyond the minimal notion. First, I introduce the notion of a *functional mechanism*: one can distinguish between those mechanisms that fulfill a (biological) function, and those that do not (Garson 2013; Piccinini 2015; Maley and Piccinini 2017). Indeed, combining the notion of a mechanism with that of a function seems to be promising with regard to making sense of the *normativity* of mechanism-talk: a mechanism that has a certain function is *supposed* to fulfill that function and might *fail* to fulfill it. In what follows, I discuss different suggestions for how to characterize functional mechanisms. It will turn out that neither of these notions successfully accounts for the normativity of mechanism-talk unless the second and third sub-types of mechanisms are taken into account. I will call the second type *regular mechanism*; the third type I will call *reversely regular mechanism* (Krickel 2018). Both notions rest on the idea that one can distinguish between one-off mechanisms and mechanisms

that establish some kind of regularity (Andersen 2012). Regular mechanisms, as I will show, have to be understood as mechanisms that bring about a particular phenomenon more often than they bring about any other phenomenon. Reversely regular mechanisms are mechanisms that bring about a particular phenomenon that is more often brought about by that mechanism than by any other mechanism. I will show how these two notions of regularity together are necessary and sufficient for grounding type-level mechanistic explanations (see also Krickel 2018), and when combined with the functional notion of a mechanism, can solve the problem of accidental goal contributions, which afflicts the most promising account of functions as discussed in the next section.

3.1 Functional Mechanisms

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

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

Today, scientists and philosophers commonly accept that biological mechanisms are not the result of the work of an intelligent designer. Furthermore, it is commonly denied that there are irreducible goals or purposes in nature. Nevertheless, the term ‘function’ is still ubiquitous in biology and philosophy of biology. Statements like ‘The heart has the function of pumping blood’ are, at least prima facie, accepted as

valid claims about hearts (see Allen 2009 for an overview of the debate about teleological terms in biology). Furthermore, the assumption that biological systems have functions manifests in normative claims such as 'A heart is a bad heart if it does not pump blood properly.' Therefore, combining function-talk with mechanism-talk seems to be promising with respect to making sense of the normativity of mechanism-talk (Garson 2013, n. 325).

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

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

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

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

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

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

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

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

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

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

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

that are not of explanatory interest at all.[1] If we cannot make sense of this distinction, a perspectivalist view of functional mechanisms does not constitute an interesting sub-type of mechanisms over and above minimal mechanisms.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.2 Regular Mechanisms[3]

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

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

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

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

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

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

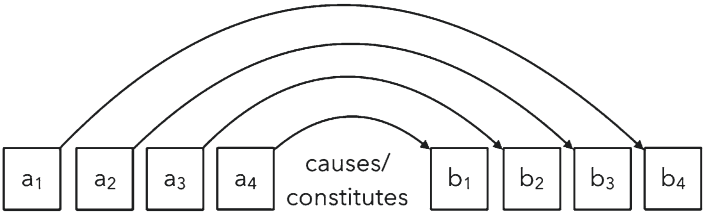

Fig. 3.1 *Factual regularity:* The relationship between types A and B is factually regular since A has multiple instances a_1–a_4 that bring about (cause or constitute) instances of B b_1–b_4

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

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

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

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

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

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

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

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

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

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

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

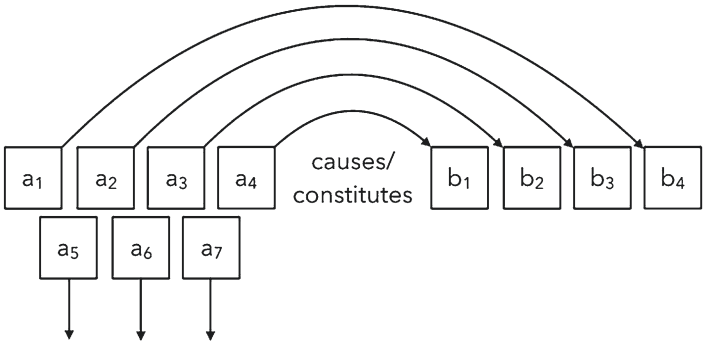

Fig. 3.2 *Stochastic regularity:* The relationship between type A and type B is stochastically regular since some but not all instances of A a_1–a_7 bring about (cause or constitute) instances of B

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

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

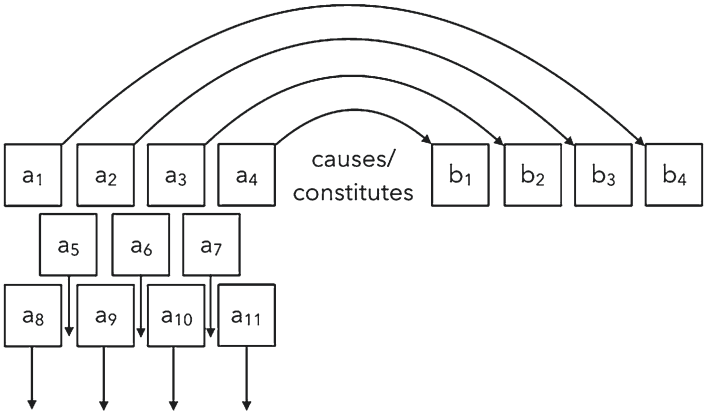

Fig. 3.3 *High-failure mechanism*: Type A has a high failure rate with respect to bringing about instances of type B since most of its instances a_1–a_{11} do not bring about (cause or constitute) an instance of B

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

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

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

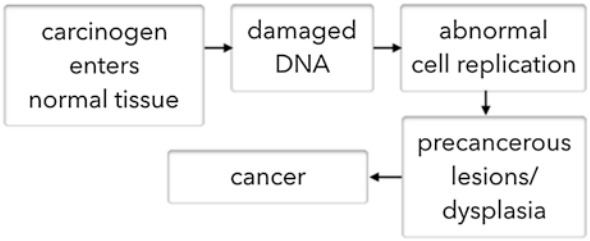

Fig. 3.4 Illustration of the cancer mechanism. (Illustration inspired by Grundmann 2000, Chap. 8)

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

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

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

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

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

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

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

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

> (*Frequented Regularity*) The relationship between a mechanism A and a phenomenon B is frequentedly regular iff there is a consistent percentage of times where instances of A bring about instances of B.
>
> (*Interrupted Regularity*) The relationship between a mechanism A and a phenomenon B is interruptedly regular iff every time when an instance of A does not cause/constitute an instance of B interfering factors can be identified.

I agree that an interruptedly regular mechanism can ground type-level explanation. Interrupted regularity implies that there is a mechanism that always brings about a phenomenon *ceteris paribus* (i.e., except for cases in which certain interfering factors occur). The cp-clause does not trivialize the assumption of a deterministic

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

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

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

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

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

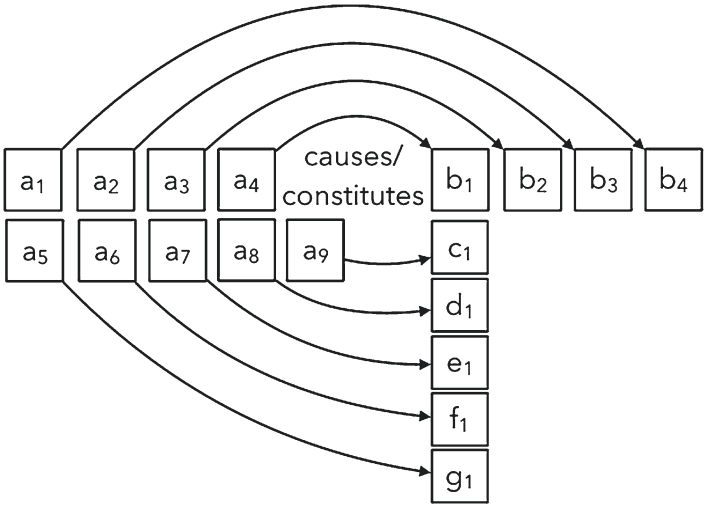

Fig. 3.5 *Comparative regularity*: The relationship between types A and B is comparatively regular since there are more instances of A a_1–a_9 that bring about an instance of B than instances of A that bring about a particular other type—the instances that do not bring about an instance of B bring about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

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

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

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

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

3.3 Reversely Regular Mechanisms

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

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

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

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

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

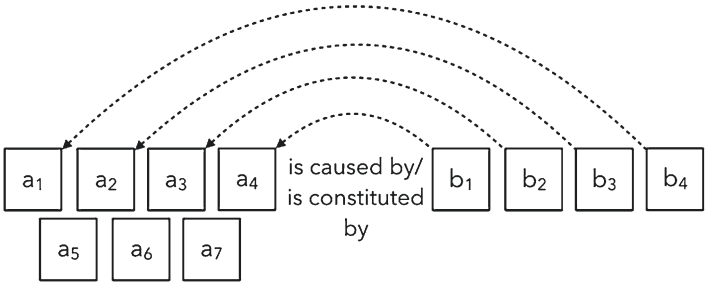

Fig. 3.6 *Reverse regularity*: The relationship between types A and B is reversely regular since all instances of B b_1–b_4 are caused/constituted by instances of A

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

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

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

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

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

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

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

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

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

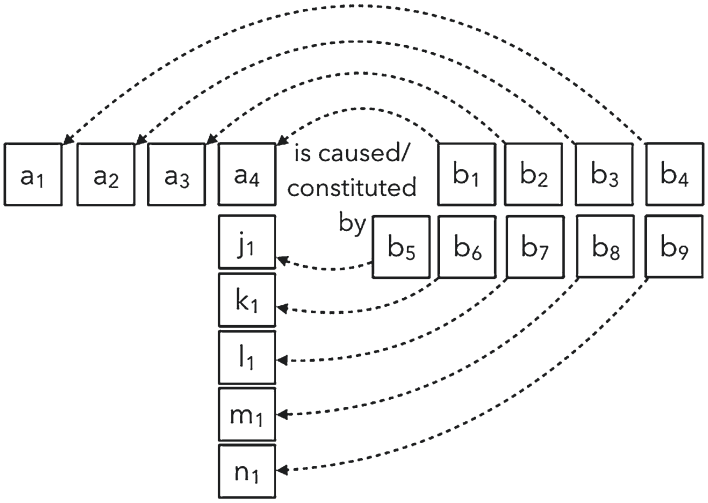

Fig. 3.7 *Comparative reverse regularity*: The relationship between types A and B is comparatively reversely regular since more instances of B are brought about by an instance of A than by any other type—the instances of B that are not brought about by an instance of A are due to instances of various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

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

3.4 Individuating Mechanism Types

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

when I speak of the individuation of mechanism types the reader should keep in mind that I assume a reductionism concerning mechanism types to similarities between token mechanisms.

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

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

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

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

The individuation of mechanism types, as it is understood here, has metaphysical as well as pragmatic aspects. Whether a particular causal sequence type is a mechanism or not depends on whether it instantiates a comparatively regular or

comparatively reversely regular relationship to the phenomenon—independently of whether we are able to detect that relationship. But, of course, we are in principle able to empirically discover mechanism types. Still, pragmatic considerations will determine which similarities between tokens are taken to be interesting and relevant such that we form type descriptions that categorize mechanism tokens.

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

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

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

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

Third, we might have committed a *lumping error* (Craver 2007, 123) on the side of the mechanism. We have individuated the mechanism type too broadly. Perhaps only some of the instances of the causal sequence type we thought to be the mechanism are in fact instances of the mechanism, and we fail to see comparative regularity because the causal sequence type we mistake for the mechanism indeed produces various different phenomena. Examples of such lumping errors are rather common in animal experimentation. Often researchers wrongly assume that there is one and

the same mechanism for a certain disease operating in humans and certain animals used as animal models. A famous example is the drug Contergan, which was used to treat morning sickness in pregnant women. Animal experimentation showed that the agent thalidomide was harmless; later it turned out that the drug was extremely harmful to the developing fetus. In this case, scientists wrongly assumed that the relevant mechanisms in humans are of the same type as the corresponding mechanisms in the non-human animals used as animal models.

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

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

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

3.5 Summary

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

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

and accidental goal contributions, and it cannot account for failure-talk concerning mechanisms of pathologies.

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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