**Design Principles and Mechanistic Explanation**

**Abstract:** In this essay I propose that what design principles in systems biology and systems neuroscience do is to present abstract characterizations of mechanisms, and thereby facilitate mechanistic explanation. To show this, one design principle in systems neuroscience, i.e., the multilayer perceptron, is examined. However, Braillard (2010) contends that design principles provide a sort of non-mechanistic explanation due to two related reasons: they are very general and describe non-causal dependence relationships. In response to this, I argue that, on the one hand, all mechanisms are more or less general (or abstract), and on the other, many (if not all) design principles are causal systems.

**1. Introduction**

The last three decades have witnessed a shift of focus from laws to mechanisms in the philosophy of science (Bechtel and Richardson, 1993; Machamer et al., 2000; Darden, 2006; Craver, 2007; Bechtel and Abrahamsen, 2010; Glennan, 2017). One driving force for this move was due to the recognition that scientific explanations in the special sciences, biology and neuroscience for example, do not always appeal to laws, nor do they proceed in the way stipulated by the Deductive-Nomological model of explanation (Hempel and Oppenheim, 1948). Instead, explanations in the special sciences are usually intimately associated with discovering, elaborating and disentangling mechanisms.

According to the proponents of mechanistic explanation, a mechanistic explanation usually starts with decomposition of a relatively complex system into smaller components, accompanied by localization of key operations and organizations among these components responsible for bringing about the phenomenon to be explained (Bechtel and Richardson, 1993). The procedure of decomposition and localization is often followed by *recomposing* the components into an organized whole when understanding the dynamic aspects of a complex system (Bechtel and Abrahamsen, 2010, 322). As a result, four key parts define a mechanism: the phenomenon to be explained, components, operations and organizations.[[1]](#footnote-1) Depending on the research interest, a component of a mechanism can be treated as a smaller mechanism and thus can be further decomposed into sub-components, and the original mechanism itself can also be situated as a component within an even larger mechanism—therefore, mechanisms usually form hierarchies.

In systems biology and systems neuroscience, scientists usually construct *design principles* to explain a phenomenon of interest. Design principles are very simple yet quite abstract building blocks (or patterns) that can be found in different systems. At first glance, they might look different from those paradigmatic cases of mechanisms often encountered and discussed in the philosophical literature, and this is due to the fact that design principles tend to be very abstract in form and are usually intimately coupled with the practice of mathematical modeling. When looking closely, however, one starts to realize that they are actually abstract characterizations of mechanisms and what they provide are mechanistic explanations. Other authors have already expressed similar ideas, albeit often in an implicit manner. For example, Green et al. (2018) argue that network approaches based on design principles not only support but also extend traditional mechanistic strategies (see also Green et al. (2015)). Brigandt (2013) calls for an extended philosophical understanding of mechanisms in light of the practice in systems biology, where traditional mechanistic strategies and network-based dynamic modeling approaches are indispensable parts in providing mechanistic explanation. Also, Matthiessen (2017) argues that “we can understand the mathematical modelling techniques of systems biologists as part of a broader practice of constructing and evaluating mechanism schemas” (2017, 1).

In this article, I follow in these authors’ footsteps but go a step further, proposing explicitly that what design principles in systems biology and systems neuroscience do is to present abstract characterizations of mechanisms, and thereby provide mechanistic explanation. To a first approximation, my argument for this proposal is that, when constructing design principles, scientists are engaged in the business of decomposing a complex system into component parts, identifying the properties of these parts, investigating the (typically causal) interactions between them, and probing the particular ways these parts are organized. Inevitably, these design principles constructed are very abstract and often take the form of network models made up of only a few nodes (denoting key components) and arrows (denoting key interactions). Moreover, to facilitate understanding the dynamic aspects of a mechanism, dynamic modeling techniques are routinely incorporated into the construction of the design principle. In short, according to my proposal, constructing a design principle just amounts to proposing an abstract characterization of mechanisms so as to provide mechanistic explanation. Understood in this way, I think we can make a big step forward towards completing the mission of extending the traditional mechanistic framework, for now the conception of mechanisms is broadened in the way that abstract design principles are viewed as a special type of mechanisms.[[2]](#footnote-2)

However, some might disagree with my view. Braillard (2010), for example, contends that design principles only describe *non-causal* *constraints* (or non-causal dependence relationships) between structures and functions, resulting in an entirely new type of explanation, the so-called *design explanation*.[[3]](#footnote-3) His reasons are twofold. First, mechanistic explanations are less *abstract* (or general) than design-principle-based explanations, because mechanistic explanations are concerned with articulating *specific* causal details of a mechanism while the latter merely concentrate on discovering general principles abstracted from causal details. Second, and relatedly, mechanistic explanations are usually *causal* explanations whereas design-principle-based explanations are *non-causal* because they only describe general dependence relationships between functions and structures. This dispute is addressed in the final section before conclusion.

The essay unfolds as follows. Section 2 takes a brief look at systems biology and systems neuroscience, giving you an idea of what a design principle looks like. A design principle from systems neuroscience, i.e., the multilayer perceptron, is introduced. Section 3 introduces an extended conception of mechanistic explanation proposed by some authors in the mechanistic camp, which features dynamical thinking. Sections 4 substantiates the extended conception by arguing that constructing a design principle just amounts to proposing an abstract characterizations of mechanisms so as to provide mechanistic explanation. Section 5 then addresses Braillard (2010)’s contention.

**2. Design Principles**

Systems biology (or systems neuroscience),[[4]](#footnote-4) a collection of research programs featuring mathematical and computational modeling as well as interdisciplinary collaboration developed during the last a few decades (Kitano, 2002; O’Malley and Dupré, 2005; Boogerd et al., 2007), has recently attracted a great deal of attention from philosophers of science (Braillard, 2010; Fagan, 2012; Boogerd et al., 2013; Brigandt, 2013; Levy and Bechtel, 2013; Green et al., 2015; MacLeod and Nersessian, 2015; Green, 2017; Matthiessen, 2017). One reason for this increased attention is related to the generalizable patterns or principles discovered by scientists using mathematical and computational tools. These generalizable principles, called *design principles* or *network motifs*, refer to “Patterns that occur in the real network significantly more often than in randomized networks” (Milo et al., 2002; Shen-Orr et al., 2002; Alon, 2007a, 27).[[5]](#footnote-5) They are patterns of organization that are obtained by abstracting from complex interaction networks in living (or non-living) systems and thus are usually abstractly specified; they are believed to be very simple and abstract building blocks for living (or non-living) systems that can be employed to provide explanation for a given phenomenon or behavior that occurs across a whole range of different circumstances. In short, they are simple, abstract patterns, instantiated in different circumstances.

One such a design principle in systems neuroscience is the multilayer network called *multilayer perceptron* (Hertz et al., 1991; Bray, 1995),[[6]](#footnote-6) a network that contains several layers, is made up of various components (i.e., neurons or proteins) and their connections, and whose components are capable of integrating the weighted sum of their inputs to obtain signals and transmitting the signals to the next layer.[[7]](#footnote-7) One case of such a multilayer perceptron is the *protein kinase cascade* in signal transduction, which serves as an information processing pathway in many eukaryotic organisms (Wiley et al., 2003; Kolch et al., 2005). As the name indicates, the cascade is constructed from kinases, a specific type of protein that catalyzes the phosphorylation of particular target proteins.[[8]](#footnote-8) One simplified model of the protein kinase cascade is shown below:



Figure 1. A simplified model of the protein kinase cascade. This figure comes from Alon (2007a, 108).

The working of this three-layer protein kinase cascade is as follows. The cascade is activated when two receptors bind two ligands[[9]](#footnote-9) respectively and activate the top-layer kinases in the cascade, i.e., $X\_{1}$ and $X\_{2}$. Each of $X\_{1}$ and $X\_{2}$, when activated by the receptors, can phosphorylate two of the kinases on the next layer, i.e., $Y\_{1}$ and $Y\_{2}$. Each of $Y\_{1}$ and $Y\_{2}$, when phosphorylated, can go on to phosphorylate two of the kinases on still the next layer, i.e., $Z\_{1}$ and $Z\_{2}$. When kinases $Z\_{1}$ and $Z\_{2}$ are phosphorylated, they phosphorylate two transcription factors which result in gene expression. In the meanwhile, phosphatases, a type of protein enzymes, constantly dephosphorylate these phosphorylated kinases—as a result, the protein kinase cascade forms a phosphorylation and dephosphorylation circle. Note that the components (i.e., the kinases) at each layer receive inputs from (typically all) the components at the previous layer and that it is the components that integrate these inputs (often in a non-linear fashion).[[10]](#footnote-10)

Consequently, the protein kinase cascade constitutes a multilayer perceptron that is able to integrate input signals from a number of receptors. So far, we have only gotten a rough idea of how the design principle works. To understand the dynamic aspects of this design principle, mathematical tools such as differential equations must be employed. However, for the purpose of this article, I will rest myself content with qualitative descriptions and leave those technically-oriented readers to Alon (2007a, 106-115). In Figure 1, we can see that $Y$ is phosphorylated by two different kinases, i.e., $X\_{1}$ and $X\_{2}$. $Y$ has two parts, phosphorylated $Y$, denoted $Y\_{p}$, and unphosphorylated $Y$, denoted $Y\_{o}$. The rate of change of $Y\_{p}$ depends on the difference between its phosphorylation and dephosphorylation. In many cases the phosphorylated part of $Y$, $Y\_{p}$, is an increasing function of the weighted sum of the concentrations of $X\_{1}$ and $X\_{2}$ with some weights, e.g., $w\_{1}$ and $w\_{2}$, respectively. Activation of $Y$ occurs only when the weighted sum of the two input kinases is larger than some threshold.

We can now use a straight line in a graph to represent the threshold, and the straight line cuts the $X\_{1}-X\_{2}$ plane into two areas: an area of low $Y\_{p}$ and an area of high $Y\_{p}$, as shown in Figure 2 below:



Figure 2. Two single-layer perceptrons with different sets of weights. In the shaded areas, $Y$ is activated. In the rest of the areas, $Y$ is inactivated. The ranges of activities of $X\_{1}$ and $X\_{2}$ are between 0 and 1, meaning that all the kinase molecules are inactive or are active, respectively. This figure comes from Alon (2007a, 110).

As Figure 2 shows, the straight lines, i.e., the threshold lines, set the boundaries for the activated and inactivated areas. The exact position and slope of the threshold lines rely on the specific values of the weights $w\_{1}$ and $w\_{2}$, which can be altered in various ways in the living cell, e.g., either by changing the regulatory mechanisms that govern the rates of the input kinases, or by mutations that directly modify the chemical affinity of the kinases to the phosphorylation sites of protein $Y$. When the values of the weights are small, the activated area takes a small corner of the plane and the activation demands high activities of both $X\_{1}$ and $X\_{2}$. In contrast, when the values of the weights are relatively large, the activated area takes a large proportion of the plane and the activation demands relatively low activities of $X\_{1}$ and $X\_{2}$—sometimes either $X\_{1}$ or $X\_{2}$ alone is sufficient to activate $Y$.

These single-layer perceptrons can be viewed as performing the simple task of computing how much $Y$ should be phosphorylated depending on the weighted sum of the inputs. If more perceptron layers are added, the resultant multilayer perceptron can perform far more complicated computations. Now consider the situation of two-layer perceptrons, where $Y\_{1}$ and $Y\_{2}$ are two middle-layer kinases which are phosphorylated by the first-layer kinases $X\_{1}$ and $X\_{2}$ and can phosphorylate the kinase $Z$ (note that only when $Y\_{1}$ and $Y\_{2}$ are phosphorylated and thus activated can they go on to phosphorylate other kinases). Hence, the concentration of phosphorylated $Z$ is an increasing function of the weighted sum of the two input kinases $Y\_{1}$ and $Y\_{2}$ (and suppose their weights are, e.g., $w\_{z1}$ and $w\_{z2}$, respectively).

As in the case for $Y$, we here also can use a curve in a graph to represent the activation threshold for $Z$, as shown below:



Figure 3. Two two-layer perceptrons with different sets of weights. This figure comes from Alon (2007a, 112).

When the values of the weights $w\_{z1}$ and $w\_{z2}$ are large enough and when either $Y\_{1}$ or $Y\_{2}$ is phosphorylated above its threshold, either $Y\_{1}$ or $Y\_{2}$ is sufficient to activate $Z$. Hence, the activation area of $Z$ is a union of the two activation areas of $Y\_{1}$ and $Y\_{2}$. By contrast, when the values of the weights are relatively small, both $Y\_{1}$ and $Y\_{2}$ are needed to be phosphorylated above their thresholds so as to activate $Z$. In other words, $Z$ is phosphorylated only in the area where both $Y\_{1}$ and $Y\_{2}$ are activated. Hence, the activation area of $Z$ is defined by the intersection of the two activation areas of $Y\_{1}$ and $Y\_{2}$. All in all, we see that adding an additional layer of perceptron helps give rise to a much more intricate computation.[[11]](#footnote-11)

Notice that in constructing this simplified model scientists usually make a number of assumptions (and/or idealizations), e.g., the first-order kinetics is applicable to these kinases, the total concentration of each kinase during the phosphorylation-dephosphorylation interactions is constant, and so on. So, very surprisingly, “Multilayer perceptrons allow even such simple units to perform arbitrarily complex computations” (Bray, 1995; Cf. Alon, 2007, 113).

So far, a design principle has been outlined, and, based on the principle’s dynamics, an explanation regarding why the multilayer perceptron can perform such complicated computations has also been provided.[[12]](#footnote-12) The next question is whether we can readily treat this design principle as an abstract characterization of mechanisms and view the corresponding explanation as a mechanistic explanation. We will come back to these questions in Section 4.

**3. Mechanistic Explanation Extended**

Recently, many philosophers have started to question the adequacy of those early characterizations of mechanism and mechanistic explanation (Bechtel and Abrahamsen 2010; 2013; Brigandt 2013; Levy and Bechtel 2013; 2016; 2020).[[13]](#footnote-13) The fundamental worry behind these philosophers is that the early characterizations might be able to make sense of very simple systems but not complex systems that are widespread in the living world.[[14]](#footnote-14)

More specifically, their worry boils down to the following considerations. First, the early characterizations view a mechanism as made up of a *linear* causal sequence which has clear-cut starting and termination points. However, scientists have come to realize that many mechanisms are very complicated networks with numerous interacting parts and feedforward/feedback loops. Second, these early characterizations usually focus on the structural, organizational, and/or spatial aspects of mechanisms, which leads to the ignorance that mechanisms are ultimately complex *dynamic* systems. Third, they see mechanisms as if they are ordinary, concrete objects such as tables or chairs that are enduring over time, have clear boundaries with their environments, and can be clearly identified and individuated. Nevertheless, research in science—especially in systems biology and systems neuroscience—has shown that mechanisms are not so easily individuated and that some mechanisms only enjoy transient existence.

Because of these limitations, many philosophers call for an extended philosophical account of mechanism, as well as an updated philosophical account of mechanistic explanation (Bechtel and Abrahamsen, 2010; Brigandt, 2013; Levy and Bechtel, 2016). Before turning to what an extended account of mechanisms might look like, two points must be noted. First, as Levy and Bechtel (2016) point out, a mechanism is dynamic in two interrelated senses: on the one hand, the mechanism itself is a complex, dynamic system that changes over time, and on the other, the process of constructing, articulating and evaluating a mechanistic explanation based on the mechanism in question is also a dynamic matter. Namely, mechanistic explanation should not be regarded as a static, final product but as a dynamic, iterative and integrative process. Second, although extended, the *baseline* for mechanistic explanation must be preserved (Levy and Bechtel, 2016). In other words, a notion of *minimal mechanism* (and a corresponding notion of a minimal mechanistic explanation) must be preserved, which will be discussed in the next section. Roughly speaking, a minimal mechanistic explanation is one that invokes unpacking how a system’s entities and activities are spatiotemporally organized so that a phenomenon of interest is brought about. Hence, against this baseline, other forms of explanation can be distinguished, e.g., topological explanation, distinctively mathematical explanation, etc.

So, what might an updated account of mechanistic explanation look like? Even though no clear image has emerged in the literature so far, a couple of authors’ recent work has already portrayed a new roadmap for future exploration: *dynamic mechanistic explanation*. Bechtel and Abrahamsen (2010, 2013) have advanced such an extended account, according to which a dynamic mechanistic explanation is achieved via the coupling of a mechanistic explanation with a computational part that usually takes the form of a set of differential equations.[[15]](#footnote-15) The computational part is integrated into the mechanistic part so as to explore “how the mechanism’s organized parts and operations are orchestrated in real time to produce dynamic phenomena” (Bechtel and Abrahamsen, 2010, 322).

Interestingly, the multilayer perceptron discussed in Section 2 has already hinted how this coupling can be achieved, and thereby vindicates their vision about how an extended mechanistic explanation might look like.[[16]](#footnote-16) Remember in Section 2 I mentioned that mathematical tools such as differential equations are usually employed to unpack the dynamic aspects of a design principle. It is clear that these equations play a pivotal role in determining, for example, how much $Y$ could be phosphorylated depending on the weighted sum of its input kinases, and in showing how a multilayer perceptron can perform a much more sophisticated computational task than a single-layer perceptron. Admittedly, without these equations, our mental rehearsal might inform us of some rough idea about how the multilayer perceptron works. Nevertheless, mental rehearsal certainly fails short of telling us how exactly the perceptron works, and, less so, for example, how the phosphorylated part of a kinase can be an increasing function of the weighted sum of its input kinases.

All in all, this attempt to extend the traditional framework pushes the bar very high, for it requires more than what the traditional framework demands. Namely, in addition to providing relevant mechanistic information about the system in question, a dynamic mechanistic explanation should also provide relevant dynamic information about the system.[[17]](#footnote-17) However, this does not entail that these two parts of information are separate or independent; rather, they are usually integrated together in a single explanation.[[18]](#footnote-18) Also, this proposed framework merely intends to extend the traditional framework rather than to invalidate the latter, for the traditional framework might still shade light on some simple explanatory scenarios where dynamic aspects are either not needed or nonessential.

**4. Design Principles as Abstract Characterizations of Mechanisms**

In this section I propose that what a design principle does is to present an abstract characterization of mechanisms, and thereby facilitate constructing mechanistic explanation. To validate my proposal, in what follows I will first elaborate on the commonly-agreed baseline requirements for mechanisms and mechanistic explanation, and then show how the protein kinase cascade case ticks all the boxes of the requirements.

There are various characterizations of mechanisms and mechanistic explanation in the literature (Bechtel and Richardson, 1993; Machamer et al., 2000; Glennan, 2002; Bechtel and Abrahamsen, 2005), which, as Tabery (2004) argues, represent not competing but complementary emphases and intellectual orientations. Although these different characterizations differ in emphases and intellectual orientations as well as varying in terminology, some authors argue that a consensus is not so hard to find. For example, Illari and Williamson proposes a “consensus conception” of mechanisms, according to which

“A mechanism for a phenomenon consists of entities and activities organized in

such a way that they are responsible for the phenomenon.” (Illari and Williamson,

2012, 120)

Also, to capture this consensus, Glennan suggests a concept of “minimal mechanism”:

“A mechanism for a phenomenon consists of entities (or parts) whose activities

and interactions are organized so as to be responsible for the phenomenon.”

(Glennan, 2017, 17)

So, though being aware that there are different characterizations of mechanisms and that each characterization might suit to different projects of interest,[[19]](#footnote-19) I hereafter take the minimal conception as the baseline for thinking about mechanisms and, based on this baseline, think that a mechanism usually involves four basic elements: (1) a phenomenon to be explained, (2) a set of entities (or components), (3) a set of activities (or interactions, operations) amongst the entities, and (4) a spatiotemporal organization of these entities and their activities.[[20]](#footnote-20) Another message conveyed by this baseline is that regardless of whichever particular account of mechanisms one supports, she might agree that something is a mechanism as long as it has these four basic elements.

Accordingly, a mechanistic explanation consists in elucidating how a phenomenon of interest is produced, underlay or maintained by a set of spatiotemporally organized entities and activities. More specifically, a mechanistic explanation typically involves uncovering what properties the relevant entities have, how these entities interact with one another due to their causally relevant properties, how these entities are organized spatially, how the interactions proceed temporally (e.g., from start to termination conditions), and eventually how all these elements of a mechanism bring about, underlie or maintain the phenomenon of interest.

Given these criteria, we are now in a good position to evaluate my proposal. The first question: is it true to say that what a design principle does is to present an abstract characterization of mechanisms? I think it is true. To begin with, in our current example, i.e., the protein kinase cascade, there is one phenomenon to be explained: why very simple networks of protein kinases can perform such complicated computational tasks? Like many other scientific questions, this is the driving force of the scientific enquiry into how this terrific phenomenon could happen at all. And practicing scientists usually believe that there must be a story behind this widespread phenomenon that could potentially make sense of the phenomenon. Second, it is an already known fact that protein kinases are involved in producing such a phenomenon, and we have already had a considerable amount of knowledge about protein kinases, including their interesting properties related to catalyzation. Third, it is also part of our existing knowledge about how protein kinases interact with one another. Finally, even though we have already obtained a great deal of knowledge about how protein kinases are organized spatiotemporally, we have not yet known how exactly they are organized spatiotemporally in a way that allows them to realize the function of performing complicated computational tasks. Obviously, in this case some not-yet-known particular organization (or organizations) is the key to understanding the phenomenon to be explained. Therefore, a specific structure (or organization), i.e., a particular cascade structure in our current case, is postulated, in the hope of accounting for the puzzling phenomenon observed. To the extent that such particularly organized protein kinases seem to be able to reproduce the phenomenon we are interested in, it leads us to believe that this particular structure might faithfully capture something causally responsible for the phenomenon of interest.

So, it seems clear that the four basic constituting elements of a mechanism are present in the construction of the particular design principle. At this point, one might wonder the difference between *an abstract characterization of mechanisms* and *a mechanism*, for sometimes I use the term *a mechanism* but other times the term *an abstract characterization of mechanisms*. This has something to do with the differing degrees of abstractness associated with mechanisms. Sometimes a design principle can be a very abstract entity, not denoting any particular type of system but a whole class of different types of systems that instantiate the same underlying design. So, when we mention *the multilayer perceptron* we are not referring to any particular type of system but a whole class of different types of systems that share the same particular design, e.g., the multilayer neuron system and the protein kinase cascade that share the same design. According to my proposal, a very abstract design principle is an abstract characterization of mechanisms and each instantiation of this very abstract design principle is a (particular type of) mechanism. To be clear, an abstract characterization of mechanisms is essentially still a mechanism (we may call it a *meta-mechanism*), albeit a very abstract one, corresponding roughly to Machamer (2000), Darden (2002) and Craver (2007b)’s notion of *a mechanism schema*, to be discussed in Section 5.1 below (there we will see that the abstractness/generality of mechanisms comes in various degrees and the differing degrees of abstractness/generality does not lend support to the claim that a very abstract/general design principle is no longer a mechanism). On the other hand, however, it should be noted that although design principles are abstract characterizations of mechanisms (and therefore are mechanisms), they are a very special type of mechanisms to be distinguished from the other types of mechanisms that do not enjoy the same level of abstraction. With this nuance in mind, in what follows I will use *a mechanism* and *an abstract characterization of mechanisms* interchangeably.

Therefore, given what said above, we may generalize that, at least for many (if not all) design principles, what they do is to present a mechanism. Now we can move to our second question: is the explanation based on this particular design principle a mechanistic explanation? Given that it is definitely a mechanism, I think the answer to this question is straightforward: Yes, it is indeed a mechanistic explanation.[[21]](#footnote-21) More specifically, it is a mechanistic explanation because it involves unpacking the entities and their causally relevant properties, the way these entities interact with one another due to their properties, the way these entities are organized spatially, and the way the interactions temporally unfold. In a nutshell, it is a mechanistic explanation in that it shows how all these relevant elements of the mechanism underlie the phenomenon of interest, or, to put it in a different way, it demonstrates how all these things “hang together” so as to produce the phenomenon of interest.[[22]](#footnote-22)

At this point, one might wonder if this conception of mechanistic explanation is too permissive so that it is either empty for it classifies everything as a mechanistic explanation, or useless for it is far detached from actual scientific practice. My response to this worry is twofold. On the one hand, the conception is actually not as permissive as it first appears. Recall the baseline for something being a mechanism, i.e., the four basic elements. Fundamentally, at the heart of a mechanism-based explanation (i.e., a mechanistic explanation) is one key: causally relevant information about the components, interactions, and organization of a given mechanism. And this view is not new, for other authors have already championed similar ideas. For instance, Woodward (2013) argues that for something to be a mechanistic explanation, it must have difference-making and spatiotemporal information, as well as show some fine-tuned organization. In light of this key, many other forms of explanation can be readily classified as non-mechanistic explanation, e.g., mathematical, geometrical, topological, structural, spatial explanation, to name just a few.

On the other hand, this conception of mechanistic explanation is indeed much broader than many would originally envision. The recognition of the limitations of the earlier characterizations of mechanistic explanation, and the attempt to extend these earlier conceptions so as to make sense of broader scientific practice, have already demonstrated that the scope of mechanistic strategies as implemented in the life sciences is significantly underestimated. This certainly does not entail that *all* explanations in the life sciences are mechanistic explanations, but does show that constructing mechanistic explanation constitutes a very large and important portion of practice in the life sciences. More specifically, it is surely not the case that only cell biology and associated disciplines are concerned with constructing mechanistic explanation, but rather that many relatively new branches of the life sciences, e.g., systems biology and systems neuroscience, are also involved in building mechanistic explanation.

However, some might be suspicious of the scope of mechanistic explanation, contending that design-principle-based explanations are not mechanistic explanations at all. Braillard (2010) is such a dissident, to be addressed in the next section.

**5. Braillard’s Contention**

Braillard (2010) argues that design principles offer a sort of non-causal explanation differing from classic mechanistic explanation, which he terms *design explanation*. More specifically, he argues that a design principle explains not by showing how a mechanism produces, underlies or maintains a phenomenon of interest, but by showing how a function constrains a structure in the way that this particular structure rather than any alternative one *must* be present in order to fulfil that function. To use Braillard’s terminology, “while the function determines the structure, it does not cause that structure” (2010, 50-51).

So, here comes the first defining feature of the so-called design explanation: it describes a non-causal (dependence) relationship. Relatedly, there is another defining feature of design explanation: it gains an unusual degree of generality that a mechanistic explanation usually does not have. Recall that the explanation based on the protein kinase cascade only invokes very abstract information about the components and operations of the cascade. Because of this, the explanation is expected to not only apply to a few particular organisms, but also to a number of heterogeneous species and even to nonliving systems.[[23]](#footnote-23)

For these reasons, Braillard claims that design-principle-based explanations afford us an entirely different kind of explanations. I will argue in what follows that this division is misplaced. First, generality (primarily due to abstraction) is a matter of degree. Second, many—if not all—design-principle-based explanations are *bona fide* causal explanations.

*5.1. Generality*

It is evident that generality is a matter of degree,[[24]](#footnote-24) rather than an all-or-nothing matter (Levy and Bechtel, 2013; Boone and Piccinini, 2016; Craver and Kaplan, 2020). Depending on different explanatory tasks or goals, one can always build a more or less general explanation. If one views offering an explanation—be it a mechanistic explanation or a non-mechanistic explanation—as offering an explanatory model, then she can see not only that generality comes in different degrees, but also that generality usually trades off with the other representational ideals of scientific modeling such as prediction precision, realism, etc. (Levins, 1966; Orzack and Sober, 1993; Odenbaugh, 2003, 2006; Weisberg, 2006; Matthewson and Weisberg, 2009) .

Roughly speaking, generality refers to how many actual or possible physical systems to which the model applies, realism denotes the degree to which the model represents the actual physical system(s), and precision concerns how close the model’s output is to the actual physical system’s output. Thus, if one aims to build a very general explanatory model (in terms of her explanatory goals) that can be applied to many different physical systems, then she should be ready to sacrifice realism and precision. Or, if she changes her mind and merely aims to build a less general explanatory model, then she should expect a less degree to which realism or precision is sacrificed.[[25]](#footnote-25)

Therefore, if we regard providing a mechanistic explanation as providing a mechanistic (explanatory) model,[[26]](#footnote-26) then it follows that the mechanistic model can be more or less general, depending on the modeler’s explanatory goals. For example, it can be very general by omitting a lot of details about any particular physical system, so that it can be applied to a huge number of heterogeneous systems—the induced fit model of enzyme catalyzation is such a case in point; or alternatively, it can be minimally general if one merely focuses on a particular type of system (e.g., a species) or a few types of system (e.g., a few species) and prefers realism to generality and precision—the *lac* operon model of *E. coli* is such a case in point. But these explanatory models’ difference in generality does not lead one to conclude that one is a mechanistic explanation whereas another is not. The same story goes for many other kinds of models. For instance, even if two non-mechanistic models, e.g., two mathematical models, differ significantly in generality, this alone does not justify the claim that one is a mathematical explanation while another is not. In short, unless some convincing reasons are given to suggest otherwise, one should hold that generality is not a defining feature of mechanistic explanation, nor is it a defining feature of design-principle-based explanation.

This observation about the generality-realism tradeoff goes in tune with many neo-mechanists’ distinction amongst mechanistic explanations. In evaluating how complete a mechanistic explanation is, Machamer (2000), Darden (2002) and Craver (2007b) distinguish *mechanism sketches, mechanism schemas* and *complete mechanistic models,* a distinction thatmakes the realism-generality tradeoff very transparent. A mechanism sketch is “an incomplete representation of a mechanism that specifies some of the relevant entities, activities, and organizational features but leaves gaps that cannot yet be filled” (Craver and Tabery, 2019). So, a mechanism sketch usually “leaves various gaps or employs filler terms for entities and processes whose nature and functioning is unknown” (Weiskopf, 2011, 316). By contrast, a mechanism schema is a relatively more complete but “abstract description of a mechanism that can be filled with more specific descriptions of component entities and activities” (Darden, 2002, S356). For example, Watson’s central dogma of DNA transcription and translation instantiates such a schema, for—though it does not leave any gaps—it is so abstract that it does not provide any information about which particular DNA, RNA and protein are involved. Ideally, a complete mechanistic model is one that neither leaves any gaps (or uses any filler terms), nor abstracts away anything.[[27]](#footnote-27)

Admitted, this distinction is not intended to show mechanistic explanations’ differences in generality; rather, it has more to do with realism, or, as Weiskopf (2011) argues, it has something to do with representational accuracy which concerns both the size of grain and correctness of a description. Nonetheless, we can see that an ideally complete mechanistic model might correspond to a minimally general model since its specification of all the relevant details makes the model maximally specific such that it might only apply to a particular type of physical system. On the other hand, a mechanism sketch might correspond to a very general model, for its leaving out of a whole load of details makes the model applicable to a broad category of (often heterogeneous) physical systems.[[28]](#footnote-28) In between them, mechanism schemas vary in degrees of generality, depending on how specific they are. The tradeoff between realism and generality also implies that the more realistic (or specific), the less general.

Therefore, either from the perspective of the realism-generality tradeoff generally true in scientific modeling, or from the perspective of the distinction interpreted along the line of the realism-generality tradeoff, we can see that mechanistic explanations can be more or less general. Furthermore, not serving as a defining feature of either mechanistic or non-mechanistic explanation, generality, in and of itself, does not stipulate that a general explanation must be a non-mechanistic explanation.[[29]](#footnote-29)

*5.2. Causality*

Braillard (2010) claims that design-principle-based explanations are non-causal because they describe some non-causal dependence relationship between a function and a structure. However, it remains unclear how general Braillard’s claim is, for there is at least one thing quite clear, namely, the explanation regarding the protein kinase cascade discussed in Section 2 is a causal one.

It goes without saying that a design-principle-based explanation can be very abstract in that it leaves out of a lot of causal details. Yet, the question is whether a design-principle-based explanation can be abstract to such an extent that it no longer contains any relevant causal details.[[30]](#footnote-30) My answer to this question has two parts, one related to the protein kinase cascade discussed in this essay, and one related to design principles in general.

First, we can ask ourselves whether the protein kinase cascade is abstract in a way that it leaves out all the relevant causal details—so, it is no longer a mechanistic explanation. The answer is clearly negative. True, the protein kinase cascade is very abstract for it omits a great deal of information about the components and operations involved in the system. For example, it only roughly says that the network has ligands, receptors and kinases, but does not say clearly what these components are, nor does it say something about these components’ properties or internal structures. Further, it describes the interactions involved in the system also in a very abstract fashion, for it only says that a ligand binds a receptor, a receptor activates a kinase, a kinase phosphorylates another kinase, etc., but does not elaborate clearly how these interactions proceed, what forms of transformational change are involved, etc. So, unsurprisingly, the explanation’s abstractness makes it apply not only to some particular (type of) living systems, but to many heterogeneous living systems and even to nonliving systems (e.g., artificial neural systems). Hence, the explanation gains a considerable degree of generality.

However, it is also true that the explanation is a causal one. Most clearly and undeniably, the interactions involved in the explanation, e.g., binding, activation, phosphorylation, dephosphorylation, are commonly-seen causal interactions. Even though these interactions can be encoded and represented by mathematical equations (e.g., differential equations), the mathematical equations are univocally causally interpreted, i.e., the mathematical equations represent, e.g., the rate at which kinase $Y$ will be phosphorylated given the concentration of kinase $X$. Furthermore, the organization of these interaction as a whole should be viewed as a causal structure, for it abstractly represents a “pattern of causal connections” among its elements (Levy and Bechtel, 2013, 259). It is the causal connectivity in the structure, the particular way in which the elements are causally connected, that accounts for the complex dynamic behaviors that the mechanism exhibits. Of course, the dynamic aspects of the mechanism cannot be easily accessed without the help of the coupled differential equations. But the employment of these mathematical tools does not render the interactions and thus the whole organization non-causal.

The claim that the explanation is causal gets extra support from the involvement of *intervention* implicit in the explanation. Recall that in Figure 2 we said that when the values of the weights ($w\_{1}$ and $w\_{2}$) are small, the activation demands high activities of both $X\_{1}$ and $X\_{2}$; and when the values of the weights are relatively large, the activation demands low activities of $X\_{1}$ and $X\_{2}$. This means that we can—either actually or potentially—intervene on the system in various ways, each of which might bring about different outcomes of interest. Or, we may argue that one of the purposes of building such an explanation is to conduct experiments on it—e.g., we change the connectivity, parameters, etc., so as to see how the phenomenon is affected. In a nutshell, the explanation—as being causal—helps us answer Woodward’s *what-if-things-had-been-different* questions (Woodward, 2003). For instance, if we fix the values of $w\_{1}$ and $w\_{2}$, we can wiggle the values of $X\_{1}$ and $X\_{2}$ in numerous ways and then see which particular combinations of them can drive the system through the threshold of activation. Alternatively, if we keep the values of $X\_{1}$ and $X\_{2}$ constant, we might wiggle the values of $w\_{1}$ and $w\_{2}$ (which requires more radical ‘surgery’ on the system) in various ways and then see how different combinations of them drive the system through the threshold. These interventions are not merely speculations; rather, they can be empirically implemented and tested by putting them either in computer simulations or in real experiments. And notice that it is basic practice that scientists integrate mathematical modeling approaches with experimentally-informed methods to investigate the dynamics of complex systems like the protein kinase cascade (Kitano, 2002; Alon, 2007b).

So, we can rest assured that at least the explanation based on the protein kinase cascade is a causal explanation. Given this, a further question arises: whether only this particular explanation is causal, or some other explanations based on some other types of design principles are causal, or, more radically, all explanations based on design principles are causal. Since defending the last option is well beyond the scope of this essay and is also not part of my goal, let me concentrate on the second option.[[31]](#footnote-31) Suffice it to briefly discuss one more design principle discussed in the systems biology literature: the type-1 incoherent feed-forward loop (I1-FFL).[[32]](#footnote-32)

P

Q

R

Figure 4. The type-1 incoherent feed-forward loop. Arrows denote activation while $⊣$ denotes repression.

In the I1-FFL, transcription factor $P$ activates transcription factor $Q$. Both $P$ and $Q$ bind the regulatory region of the target gene $R$ and collectively regulate $R$. Yet, $P$ activates $R$ whereas$ Q $represses $R$. So, the two paths have opposite effects on $R$. Remarkably, this design principle has the function of significantly speeding the response time of the transcription network (Mangan and Alon, 2003; Mangan et al., 2006; Alon, 2007b).[[33]](#footnote-33) For an explanation of how this design principle’s dynamics help them fulfil the speeding function, see Alon (2007a, 27–70).

Prima facie, Alon’s explanation based on the design principle is causal, for the explanation is achieved by describing the causal interactions among various components. Going one step further, explaining why the design principle has the speeding function consists in elaborating the dynamics involved in the interactions among those components. No doubt, differential equations are appealed in uncovering the dynamics. Nevertheless, the differential equations are meant to capture the very causal interactions involved in the design principle, e.g., the rate at which the concentration of $R$’s product is changed as $R$ is constantly activated by $P$ and repressed by $Q$. Hence, as in the case of the protein kinase cascade, the I1-FFL describes a pattern of causal interactions among its components. Also, it is the causal connectivity in the structure, the particular way in which the components are causally connected, that accounts for the complex dynamic behaviors that the principle displays. Accordingly, Alon’s explanation also helps answer Woodward’s *what-if-things-had-been-different* questions, for there are numerous ways (either actual or potential) to intervene on the system and to experimentally test the results after intervention. So, this is a causal explanation.

Therefore, the protein kinase cascade, plus the I1-FFL, should provide us some grounds to conclude that many design-principle-based explanations are causal. And this is compatible with the possibility that there might be some design-principle-based non-causal explanations. Yet, the fact that at least many well-studied design principles provide causal explanations has already repudiated Braillard (2010)’s claim that we need an entirely new category for design-principle-based explanations because they are non-causal and non-mechanistic.

**6. Conclusion**

In this article I proposed that what design principles deployed in systems biology and systems neuroscience do is to present abstract characterizations of mechanisms, albeit often in an abstract manner, and thereby facilitate mechanistic explanation. To show this, a design principle drawn from systems neuroscience, i.e., the protein kinase cascade, was discussed. Although this proposal might sound reasonable for many, others disagree. Hence, Braillard’s contention that design principles provide non-mechanistic explanation was examined. After invalidating the two pillars of Braillard’s argument, i.e., generality and non-causalness, I showed that Braillard’s contention was untenable.

**References**

Abrahamsen, Adele, and William Bechtel. 1991. *Connectionism and the Mind: An Introduction to Parallel Processing in Networks.* Basil Blackwell.

Alon, Uri. 2007a. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC.

———. 2007b. “Network Motifs: Theory and Experimental Approaches.” *Nature Reviews Genetics* 8 (6): 450–61.

Andersen, Holly. 2014a. “A Field Guide to Mechanisms: Part I.” *Philosophy Compass* 9 (4): 274–83.

———. 2014b. “A Field Guide to Mechanisms: Part II.” *Philosophy Compass* 9 (4): 284–93.

Baker, Alan. 2005. “Are There Genuine Mathematical Explanations of Physical Phenomena?” *Mind* 114 (454): 223–38.

Bechtel, William, and Adele Abrahamsen. 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 (2): 421–41.

———. 2010. “Dynamic Mechanistic Explanation: Computational Modeling of Circadian Rhythms as an Exemplar for Cognitive Science.” *Studies in History and Philosophy of Science Part A* 41 (3): 321–33.

Bechtel, William, and Adele A. Abrahamsen. 2013. “Thinking Dynamically about Biological Mechanisms: Networks of Coupled Oscillators.” *Foundations of Science* 18 (4): 707–23.

Bechtel, William, and Robert C. Richardson. 1993. *Discovering Complexity*. Princeton: Princeton University Press.

Boogerd, Fred, Frank J. Bruggeman, Jan-Hendrik S. Hofmeyr, and Hans V. Westerhoff. 2007. *Systems Biology: Philosophical Foundations*. Elsevier.

Boogerd, Fred C., Frank J. Bruggeman, and Robert C. Richardson. 2013. “Mechanistic Explanations and Models in Molecular Systems Biology.” *Foundations of Science* 18 (4): 725–44.

Boone, Worth, and Gualtiero Piccinini. 2016. “Mechanistic Abstraction.” *Philosophy of Science* 83 (5): 686–97.

Braillard, Pierre-Alain. 2010. “Systems Biology and the Mechanistic Framework.” *History and Philosophy of the Life Sciences* 32 (1): 43–62.

Bray, Dennis. 1995. “Protein Molecules as Computational Elements in Living Cells.” *Nature* 376 (6538): 307–12.

Brigandt, Ingo. 2013. “Systems Biology and the Integration of Mechanistic Explanation and Mathematical Explanation.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4): 477–92.

Colyvan, Mark. 2001. *The Indispensability of Mathematics*. Oxford & New York: Oxford university press.

Craver, Carl. 2007b. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.

Craver, Carl F., and David M. Kaplan. 2020. “Are More Details Better? On the Norms of Completeness for Mechanistic Explanations.” *The British Journal for the Philosophy of Science* 71 (1): 287–319.

Craver, Carl, and James G. Tabery. 2019. “Mechanisms in Science.” In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/science-mechanisms/#toc.

Darden, Lindley. 2002. “Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining.” *Philosophy of Science* 69 (S3): S354–65.

———. 2006. *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge: Cambridge University Press.

El-Samad, H., J. P. Goff, and M. Khammash. 2002. “Calcium Homeostasis and Parturient Hypocalcemia: An Integral Feedback Perspective.” *Journal of Theoretical Biology* 214 (1): 17–29.

Fagan, Melinda Bonnie. 2012. “Waddington Redux: Models and Explanation in Stem Cell and Systems Biology.” *Biology & Philosophy* 27 (2): 179–213.

Fang, Wei. 2021. “Towards Mechanism 2.1: A Dynamic Causal Approach.” *Philosophy of Science* 88 (5): 796–809. https://doi.org/10.1086/715081.

Gardner, Matt W., and S. R. Dorling. 1998. “Artificial Neural Networks (the Multilayer Perceptron)—a Review of Applications in the Atmospheric Sciences.” *Atmospheric Environment* 32 (14–15): 2627–36.

Glennan, Stuart. 2002. “Rethinking Mechanistic Explanation.” *Philosophy of Science* 69 (S3): S342–53.

———. 2017. *The New Mechanical Philosophy*. Oxford: Oxford University Press.

Green, Sara. 2015. “Revisiting Generality in Biology: Systems Biology and the Quest for Design Principles.” *Biology & Philosophy* 30 (5): 629–52.

———. 2017. *Philosophy of Systems Biology*. Springer.

Green, Sara, Melinda Fagan, and Johannes Jaeger. 2015. “Explanatory Integration Challenges in Evolutionary Systems Biology.” *Biological Theory* 10 (1): 18–35.

Green, Sara, Maria Şerban, Raphael Scholl, Nicholaos Jones, Ingo Brigandt, and William Bechtel. 2018. “Network Analyses in Systems Biology: New Strategies for Dealing with Biological Complexity.” *Synthese* 195 (4): 1751–77.

Hempel, Carl G., and Paul Oppenheim. 1948. “Studies in the Logic of Explanation.” *Philosophy of Science* 15 (2): 135–75.

Hertz, John A., Anders Krogh, and Richard Palmer. 1991. *Introduction to the Theory of Neural Computation*. Westview Press.

Illari, Phyllis McKay, and Jon Williamson. 2012. “What Is a Mechanism? Thinking about Mechanisms across the Sciences.” *European Journal for Philosophy of Science* 2 (1): 119–35.

Kaplan, David Michael, and Carl F. Craver. 2011. “The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective.” *Philosophy of Science* 78 (4): 601–27.

Kitano, Hiroaki. 2002. “Systems Biology: A Brief Overview.” *Science* 295 (5560): 1662–64.

Kolch, Walter, Muffy Calder, and David Gilbert. 2005. “When Kinases Meet Mathematics: The Systems Biology of MAPK Signalling.” *FEBS Letters* 579 (8): 1891–95.

Lange, Marc. 2013. “What Makes a Scientific Explanation Distinctively Mathematical?” *The British Journal for the Philosophy of Science* 64 (3): 485–511.

Levins, Richard. 1966. “The Strategy of Model Building in Population Biology.” *American Scientist* 54 (4): 421–31.

Levy, Arnon. 2013. “What Was Hodgkin and Huxley’s Achievement?” *The British Journal for the Philosophy of Science* 65 (3): 469–92. https://doi.org/10.1093/bjps/axs043.

Levy, Arnon, and William Bechtel. 2013. “Abstraction and the Organization of Mechanisms.” *Philosophy of Science* 80 (2): 241–61.

———. 2016. “Towards Mechanism 2.0: Expanding the Scope of Mechanistic Explanation.” http://philsci-archive.pitt.edu/12567/.

———. 2020. “Beyond Machine-like Mechanisms.” In *Philosophical Perspectives on the Engineering Approach in Biology*, by Sune Holm and Maria Serban. Routledge.

Lyon, Aidan. 2012. “Mathematical Explanations of Empirical Facts, and Mathematical Realism.” *Australasian Journal of Philosophy* 90 (3): 559–78.

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

MacLeod, Miles, and Nancy J. Nersessian. 2015. “Modeling Systems-Level Dynamics: Understanding without Mechanistic Explanation in Integrative Systems Biology.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 49 (February): 1–11.

Mangan, Shmoolik, and Uri Alon. 2003. “Structure and Function of the Feed-Forward Loop Network Motif.” *Proceedings of the National Academy of Sciences* 100 (21): 11980–85.

Mangan, Shmoolik, Shalev Itzkovitz, Alon Zaslaver, and Uri Alon. 2006. “The Incoherent Feed-Forward Loop Accelerates the Response-Time of the Gal System of Escherichia Coli.” *Journal of Molecular Biology* 356 (5): 1073–81.

Matthewson, John, and Michael Weisberg. 2009. “The Structure of Tradeoffs in Model Building.” *Synthese* 170 (1): 169–90.

Matthiessen, Dana. 2017. “Mechanistic Explanation in Systems Biology: Cellular Networks.” *The British Journal for the Philosophy of Science* 68 (1): 1–25.

Milo, Ron, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. 2002. “Network Motifs: Simple Building Blocks of Complex Networks.” *Science* 298 (5594): 824–27.

Odenbaugh, Jay. 2003. “Complex Systems, Trade-Offs, and Theoretical Population Biology: Richard Levin’s ‘Strategy of Model Building in Population Biology’ Revisited.” *Philosophy of Science* 70 (5): 1496–1507.

———. 2006. “The Strategy of ‘The Strategy of Model Building in Population Biology.’” *Biology and Philosophy* 21 (5): 607–21.

O’Malley, Maureen A., and John Dupré. 2005. “Fundamental Issues in Systems Biology.” *BioEssays* 27 (12): 1270–76.

Orzack, Steven Hecht, and Elliott Sober. 1993. “A Critical Assessment of Levins’s The Strategy of Model Building in Population Biology (1966).” *The Quarterly Review of Biology* 68 (4): 533–46.

Pincock, Christopher. 2015. “Abstract Explanations in Science.” *The British Journal for the Philosophy of Science* 66 (4): 857–82.

Ramchoun, Hassan, Mohammed Amine Janati Idrissi, Youssef Ghanou, and Mohamed Ettaouil. 2016. “Multilayer Perceptron: Architecture Optimization and Training.” *IJIMAI* 4 (1): 26–30.

Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: NJ: Princeton University Press.

Schmickl, Thomas, and Istvan Karsai. 2018. “Integral Feedback Control Is at the Core of Task Allocation and Resilience of Insect Societies.” *Proceedings of the National Academy of Sciences* 115 (52): 13180–85.

Shen-Orr, Shai S., Ron Milo, Shmoolik Mangan, and Uri Alon. 2002. “Network Motifs in the Transcriptional Regulation Network of Escherichia Coli.” *Nature Genetics* 31 (1): 64–68.

Tabery, James G. 2004. “Synthesizing Activities and Interactions in the Concept of a Mechanism.” *Philosophy of Science* 71 (1): 1–15.

Weisberg, Michael. 2006. “Forty Years of ‘The Strategy’: Levins on Model Building and Idealization.” *Biology and Philosophy* 21 (5): 623–45.

Weiskopf, Daniel A. 2011. “Models and Mechanisms in Psychological Explanation.” *Synthese* 183 (3): 313–38.

Wiley, H. Steven, Stanislav Y. Shvartsman, and Douglas A. Lauffenburger. 2003. “Computational Modeling of the EGF-Receptor System: A Paradigm for Systems Biology.” *Trends in Cell Biology* 13 (1): 43–50.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

———. 2013. “II—James Woodward: Mechanistic Explanation: Its Scope and Limits.” In *Aristotelian Society Supplementary Volume*, 87:39–65. The Oxford University Press.

Wouters, Arno G. 2007. “Design Explanation: Determining the Constraints on What Can Be Alive.” *Erkenntnis* 67 (1): 65–80.

Wright, Cory, and Dingmar Van Eck. 2018. “Ontic Explanation Is Either Ontic or Explanatory, but Not Both.” *ERGO* 5 (38): 997–1029.

Yi, Tau-Mu, Yun Huang, Melvin I. Simon, and John Doyle. 2000. “Robust Perfect Adaptation in Bacterial Chemotaxis through Integral Feedback Control.” *Proceedings of the National Academy of Sciences* 97 (9): 4649–53.

1. Notice that different authors of the mechanistic camp employ slightly different terminology, e.g., Machamer et al. (2000) use activities rather than operations. However, these differences in terminology do not concern my article. [↑](#footnote-ref-1)
2. At one extreme, it might appear that some design principles are so abstract that they are no longer mechanisms, e.g., integral feedback control discussed by Braillard (2010). However, as Matthiessen (2017) argues, it remains unclear if this principle is biological at all, for it is a principle originally developed in the engineering context and later borrowed to biology. More importantly, once this abstract mathematical framework is applied to the biological context where mechanistic information is added and mechanistic interpretations of the key terms in the framework are made, it inevitably becomes a biological mechanism. [↑](#footnote-ref-2)
3. Wouters (2007) also employs the term “design explanation”, but in a different sense. For Wouters, a “design explanation” explains why an actual design of some organisms is better than some contrasting design, e.g., why fishes respire with gills rather than lungs. So, his “design explanation” differs from design principles in systems biology. [↑](#footnote-ref-3)
4. Systems biology and systems neuroscience are different disciplines. However, since both study signal transduction in neuronal networks, in what follows I treat them as interchangeably—unless otherwise noted. [↑](#footnote-ref-4)
5. Although the terms *design principles* and *network motifs* are used interchangeably throughout the essay, they are not exactly the same thing. Network motifs are almost always associated with a network with a number of nodes connected in a certain way, while design principles are more general and are not restricted to network representations. We may say network motifs are a type of design principles but not vice versa. [↑](#footnote-ref-5)
6. It is worth mentioning that Abrahamsen and Bechtel (1991) is perhaps the earliest philosophical discussion of perceptrons, where they examined in detail both single-layer and multilayer perceptrons. I thank one anonymous reviewer for letting me know this. [↑](#footnote-ref-6)
7. Perceptrons were first introduced to characterize how neurons might work to perceive patterns (hence the name). Later, the limitations of single-layer perceptrons were noticed by scientists, who also recognized that multilayer systems could overcome the limitations. However, they dismissed these multilayer systems (because there was no learning rule for them) until considerably later with the introduction of backpropagation. I thank one anonymous reviewer for letting me notice this history. [↑](#footnote-ref-7)
8. Phosphorylation is the chemical reaction in which a charged PO4 group is added to a specific site on the target protein, whereas dephosphorylation is the removal of the PO4 group. Both reactions are catalyzed by specific enzymes, e.g., kinases and phosphatases. [↑](#footnote-ref-8)
9. Receptors are proteins that detect input signals to signal transduction networks, and ligands are specific molecules, a kind of input signal, that can be detected by receptors. [↑](#footnote-ref-9)
10. I thank one anonymous reviewer for letting me emphasize this point. [↑](#footnote-ref-10)
11. Note that adding new layers to the perceptron will not necessarily increase the computational power of the perception, for we can always construct a simple perceptron without extra layers that will compute the same function. So, what really matters here is non-linearity. For this reason, here we assume that the relevant activation function involved is non-linear. [↑](#footnote-ref-11)
12. It is worth mentioning that design principles can perform many other complicated functions not discussed in this essay, e.g., discrimination (being able to tell apart a set of very similar stimuli patterns), generalization (being able to “fill in the gaps” in partial stimuli patterns), graceful degradation (deteriorating rather than crashing down upon damage to parts or connections of the perceptron), etc. For details of these functions, see Hertz et al. (1991). [↑](#footnote-ref-12)
13. Those early characterizations primarily refer to Machamer et al. (2000), Glennan (2002) and Bechtel and Abrahamsen (2005). [↑](#footnote-ref-13)
14. Bechtel and Richardson anticipated this extension in their early book *Discovering Complexity*, where they not only noticed the simplicity of some accounts of mechanisms but also concerned how to overcome the simplicity (see Bechtel and Richardson (1993)). I thank one anonymous reviewer for letting me notice this early discussion. [↑](#footnote-ref-14)
15. Fang (2021) recently has taken a further step, i.e., *a* *dynamic causal approach*, to extend the mechanistic framework, which highlights the dynamic and causal dimensions of a mechanism, and stresses the relevance of computational and causal modeling to establishing a mechanistic explanation. [↑](#footnote-ref-15)
16. Notice that Bechtel and Abrahamsen (2010) employ a different example, i.e., circadian rhythm, to illustrate their dynamic mechanistic explanation. However, their purpose and mine are the same, i.e., showing how a computational part must be incorporated into a mechanistic explanation so as to explain phenomena arising from complex dynamic systems. [↑](#footnote-ref-16)
17. Mechanistic information here means the information about the properties of a system’s components, the kinds of interactions amongst the components, and the components and their interactions’ spatiotemporal organization. Dynamic information means the information encoded in the mathematical or dynamic tools, e.g., differential equations. [↑](#footnote-ref-17)
18. Kaplan and Craver (2011) hold that the mechanistic part of a mechanism and the mathematical tools (used to represent the dynamic parts of the mechanism) stand in a one-to-one mapping relationship. However, my position remains neutral with this view. [↑](#footnote-ref-18)
19. Andersen (2014a; 2014b) argues that the different characterizations might suit to different projects of interest to philosophers and that the minimal conception might only suit to some more permissive sense of that term, e.g., ontological or causal structure sense. [↑](#footnote-ref-19)
20. Craver and Tabery (2019) also think that a mechanism typically has these four basic elements. [↑](#footnote-ref-20)
21. By referring to Carl Craver, one anonymous reviewer suggests that we view design principles as providing *how-possibly explanations*. However, whether design principles can correspond to Craver’s *how-possibly explanations* remains an open question. Since, according to Craver, *how-possibly explanations* are usually not adequate explanations because they are “only loosely constrained conjectures about the sort of mechanism that might suffice to produce the explanandum phenomenon” (2007b, 112). In Craver’s framework, at the other extreme are *how-actually explanations*, explanations that invoke real components, interactions and organizational features that actually bring about the phenomenon of interest. *How-plausibly explanations* reside in the middle of these extremes, which are “more or less consistent with the known constraints on the components, their activities, and their organization” (Ibid., 112-113). For the limitations of space, I will leave this problem for another occasion. [↑](#footnote-ref-21)
22. One anonymous reviewer points out that in constructing design principles one is not carrying out research to figure out a mechanism but is building upon it, because she has already had knowledge about the mechanism—what she now needs to do is to abstract away from the details of this particular mechanism and obtain an overall design. I agree with this view, but with a slightly different interpretation of what is really going on here, for I think when developing a design principle, researchers are indeed engaged in figuring out a yet-to-be-discovered *meta-mechanism*, i.e., an abstract characterization of mechanisms, though not a particular mechanism. This meta-mechanism, though built upon existing mechanistic knowledge about specific systems, is obtained by abstracting a whole load of details away from any specific systems. So, in a sense, this meta-mechanism has not been already present because of our prior mechanistic knowledge, but only starts to emerge when we strive for a level of abstraction that goes beyond any particular mechanism. Hence, the reviewer is right to point out that in constructing design principles one is not carrying out research to figure out *a mechanism*, because one is carrying out research to figure out *a meta-mechanism*; and the reviewer is also right to point out that this carrying out research to figure out a meta-mechanism is built upon existing mechanistic knowledge about specific mechanisms. Interpreted in this way, I think my view in the article is consistent with the reviewer’s view. [↑](#footnote-ref-22)
23. In fact, the multilayer perceptron is also studied in artificial intelligence and artificial neural networks. See Bray (1995), Gardner and Dorling (1998), Ramchoun et al. (2016). [↑](#footnote-ref-23)
24. One thing about the relationship between generality and abstraction must be noted. Undoubtedly, generality and abstraction are closely related concepts, but they are not equivalent, for an abstract model may only apply to a limited set of physical systems, while a detailed, not-so-abstract model may apply very generally. However, by omitting details and thus making the model more abstract, we usually obtain a more general model. Levy and Bechtel (2013) also discuss the relationship between realism and generality in network motifs and connectivity models (and they also highlight the generality of network motifs), though their terminology is slightly different: abstraction versus generality. For them, abstraction denotes the degree to which specific details about parts and connections are left out, which corresponds to what I mean “realism” in this essay. So, in what follows, unless otherwise noted, I will use these two terms interchangeably. [↑](#footnote-ref-24)
25. One anonymous reviewer points out that how general a design principle is depends on what one takes to be *the* design principle. For instance, a perceptron can be very general prior to training but very specific/realistic after training (due to the acquired weights). This contrast is not unlike the distinction between uninstantiated models (parameter values not assigned) and instantiated models (parameter values assigned), and it is not difficult to see that an uninstantiated model is not the same as an instantiated model (for many instantiated models can be derived from a single uninstantiated model). For the same reason, I do not think a perceptron before training is the same perceptron as the one after training. [↑](#footnote-ref-25)
26. In fact, many authors view mechanistic explanations through the lens of mechanistic models, e.g., Craver (2007b) distinguishes *how-possibly, how-plausibly and how-actually* (mechanistic) models, Weiskopf (2011) directly takes mechanistic explanations to be mechanistic models, etc. Notice that Craver (2007b) holds an *ontic-conception of explanation*, according to which models are derivative of the real explanations, which for Craver (like Salmon (1984)) are the mechanisms in the world. For a discussion of the ontic conception, see Wright and Van Eck (2018). [↑](#footnote-ref-26)
27. Notice that these authors (Machamer et al., 2000; Darden, 2006; Craver, 2007) often assume that the more concrete (or specific) a mechanistic explanation is, the better it is. So, mechanism sketch and schemata are not really explanatory for they are just way stations on the road to genuine mechanistic explanations. Craver expresses this very explicitly: “progress in building mechanistic explanations involves movement along […] the sketch-schema-mechanism axis” (2007b, 114). However, following Brigandt (2013) and Levy and Bechtel (2013), I do not think this is true, for a very abstract mechanistic explanation, e.g., the protein kinase cascade, can also be a good mechanistic explanation, and abstraction is even an indispensable part of a good mechanistic explanation. [↑](#footnote-ref-27)
28. Note that the degree of generality of a model is largely driven by the target phenomenon to be explained and the intended goal of modeling, for an explanatory model is built for the purpose of providing explanation for the target phenomenon in the first place. Therefore, if the phenomenon to be explained is quite general in the sense that it is exhibited in a wide range of actual instances, then the corresponding model for this phenomenon is also relatively general. By contrast, if the phenomenon to be explained includes many specifics of a given case, then the corresponding model for the phenomenon can be less general. I thank one anonymous reviewer for letting me notice this point. [↑](#footnote-ref-28)
29. One anonymous reviewer points out that Braillard’s conception of generality differs from what we usually associate with the term, for his conception is about *constraint-based generality*. Roughly speaking, constraint-based generality stems from the fact that different systems share the same organizational pattern because they are all under the same set of constraints. In our current situation regarding design principles, this refers to the fact that the dependence relation between a structure (or a design) and a function (or a phenomenon/behavior) arises from “constraints on the possible stable states, the possible functional relations, or the possible evolutionary trajectories of a class of systems” (Green 2015, 632). The very existence of these constraints means that a given function can only be produced by a particular structure (or a limited set of structures). On the other hand, the underlying structure (as well as the manifested function) can still be found in a whole range of different systems, so here comes generality. I admit that this is an interesting sense of generality, but do not think that this sense of generality cannot be accommodated within the *orthodoxic* framework of generality. For one thing, although constraint-based generality concerns the way generality arises (due to constraints), it is ultimately still linked to how many actual (or potential) systems it can be applied to. Essentially, we say a design principle is general because we see it is instantiated in many different systems. So, I see no difference between constraint-based generality and our usual sense of generality. For another, and more importantly, constraint-based generality does not only applies to design principles but also to typical-sense mechanisms. Constraints exist at all levels of biology, and, due to physical (e.g., thermodynamic) and biological reasons (e.g., evolutionary processes and historical contingencies), it is also the case that in reality a given function can only be realized by one (or a limited set of) mechanisms. Perhaps there is a difference in degree with respect to constraints, but this sort of difference should not justify the claim that design principles’ constraint-based generality deserves an entirely distinct category. [↑](#footnote-ref-29)
30. For a discussion of when it is both sufficient and advisable to leave out causal details while still capturing causal relations, see Levy (2013). [↑](#footnote-ref-30)
31. It remains an open question whether there are any non-causal explanations provided by design principles. Although I am noncommittal about the answer(s) to this question, I’d like to briefly discuss one example that advocates of design explanation usually treat as providing an exemplar of non-causal explanation (Braillard, 2010; Green, 2015): *integral feedback control* (IFC). IFC is an abstract mathematical principle that shows the property of robust adaption exhibited by diverse living and nonliving systems across multiple scales, e.g., thermostats, bacterial chemotaxis (Yi et al., 2000), calcium homeostasis (El-Samad et al., 2002), resilience of insect societies (Schmickl and Karsai, 2018), etc. However, as Matthiessen has argued, in the context of bacterial chemotaxis, discovering IFC only marks the beginning of constructing a mechanistic explanation and “we can understand the mathematical modelling techniques of systems biologists as part of a broader practice of constructing and evaluating mechanism schemas” (2017, 1). This is simply because—as Matthiessen argues—IFC itself is merely an *abstractum* with some abstract properties; it starts to explain a phenomenon only when it is embedded into a particular (or a particular type of) system. So, when embedded into a biological system, mechanistic details are incorporated so as to explain a biological phenomenon of interest. I am fully aware that this discussion ultimately leads to the question of whether mathematical explanation can explain physical phenomena (Colyvan, 2001; Baker, 2005; Lyon, 2012; Lange, 2013; Pincock, 2015). Nevertheless, addressing this issue is surely beyond this essay’s scope and the success (or failure) of this essay does not rely on whether this issue is addressed or not. [↑](#footnote-ref-31)
32. The FFLs actually consist of a whole set of design principles depending on how one component interacts with the others (e.g., whether it is activation or repression). For simplicity, what I show in this article is only one type of them, i.e., the type-1 incoherent FFL. [↑](#footnote-ref-32)
33. For a comprehensive explanation of how these design principles’ dynamics help them fulfil the speeding function, see Alon (2007a, 27–70). [↑](#footnote-ref-33)