Phenomenological Laws and Mechanistic Explanations

*Forthcoming in *Philosophy of Science*. Please cite published version when possible.

Gabriel Siegel
Department of Philosophy
Washington University in St. Louis
gabriel.siegel@wustl.edu

Carl F. Craver
Department of Philosophy
Washington University in St. Louis
ccraver@wustl.edu

Abstract

In light of recent criticisms by Woodward (2017) and Rescorla (2018), we examine the relationship between mechanistic explanation and phenomenological laws. We disambiguate several uses of the phrase "phenomenological law" and show how a mechanistic theory of explanation sorts them into those that are and are not explanatory. We also distinguish the problem of phenomenological laws from arguments about the explanatory power of purely phenomenal models, showing that Woodward and Rescorla conflate these problems. Finally, we argue that the temptation to pit mechanistic and interventionist theories of explanation against one another occludes important and scientifically relevant research questions.

Acknowledgments

We thank Stuart Glennan, David Kaplan, Arnon Levy, Mark Povich, James Woodward and two anonymous referees from *Philosophy of Science*.

Introduction. Mechanistic explanation has been a major preoccupation of recent work in philosophy of science (e.g., Machamer, Darden and Craver 2000; Craver 2006; 2007; Bechtel and Abrahamsen 2005; Glennan 2002; 2017). Central to this development has been an emphasis on explanations in which the explanandum phenomenon is the behavior or property of a mechanism as a whole and the explanans involves the organized activities of the mechanism's components. A primary driver of this work has been to build a model of explanation that adequately describes explanatory practices in the special sciences. Such theorists built on the well-known criticisms of the covering law (Hempel 1965) and unificationist (Kitcher 1989) models of explanation (by, e.g., Scriven 1975 and Salmon 1984; 1994), and built upon positive steps away from classical models of reduction by, e.g., Wimsatt (1972; 1974), Lycan (1990), and Simon (1969). They sought to characterize mechanistic explanation in terms that could be given clear philosophical expositions (e.g., causation, organization, component), escaping once-prevalent suspicion that the term "mechanism" could be given no general and useful formulation (e.g., Schaffner 1993).

One common criticism of mechanistic approaches is that they deny that so-called phenomenological laws, such as the ideal gas law, have explanatory power (Woodward 2017; Rescorla 2018).[1] The argument is simple:

> 1. Mechanists hold that for a model to count as explanatory of a phenomenon P, that model must describe the mechanism for P.
>
> 2. Mechanists hold that phenomenological laws describe regularities and do not describe the mechanisms by virtue of which those regularities hold.
>
> C. Mechanists hold that phenomenological laws are not explanatory; for example, one cannot explain a helium balloon's expansion by citing the ideal gas law and the fact that one raised the temperature of the gas inside the balloon.

According to this argument, the mechanistic view declares illegitimate a mundane and unproblematic bit of everyday scientific explanatory practice. Given that phenomenological explanations are ubiquitous in the special sciences, this counts heavily against the prized descriptive adequacy of mechanistic views.

We will argue that this argument rests on a few simple, but subtle, confusions and equivocations. Some of these confusions are exegetical; as we'll argue, Premise 1 is false. But the tendency to believe Premise 1 traces to some deeper confusions about explanation. In particular,

[1] For related criticisms, also see Weiskopf (2011), Woodward (2013), Shapiro (2020), and Ross (forthcoming).

these critics conflate two philosophical problems: the problem of phenomenological laws and the problem of phenomenal models. This conflation, we believe, traces ultimately to a failure to digest an insight of mechanical philosophers since Salmon: that there are both etiological and constitutive aspects of explanation, each with different norms and associated practices.

The failure to mark this distinction is often a blind-spot for interventionist theories of explanation. Perhaps most fundamentally, the errors packaged in this argument derive from a mistakenly adversarial casting of the intellectual terrain: many advocates of mechanistic explanation (e.g., Craver 2007; Kaplan 2011), understood as we defend it here, did not offer their views as competitors to interventionist theories of causal explanation but rather as extensions and companions to that view.[2] Most centrally, these mechanists took themselves to be filling in the interventionist blind-spot for constitutive explanations. Once this blind-spot is addressed, a progressive research field opens in a space once crowded out by apparent conflict.

In Section 1, we consider some different ways of thinking about *phenomenological laws*, each with different implications for their status as explanatory. We use this to lay out a plausible formulation of the critics' target argument, and then to respond to it. In Section 2, we consider the altogether distinct problem of *phenomenal models*, which arises only in the context of understanding constitutive explanations. We explain why mechanists object to phenomenal models as explanations and argue that this is a reasonable view. In Section 3, we return to Woodward (2017) and Rescorla's (2018) shared critique and use the conceptual solvents introduced in Sections 1 and 2 to dissolve their arguments. This is important for allowing key and unresolved issues on scientific explanation to take form. In Section 4, we consider two among many: Developing a theory of constitutive explanatory relevance and providing an analysis of what it means to find the appropriate level of explanation. We emphasize that a collaborative view of the relationship between interventionist and mechanistic views opens such progressive research questions. Otherwise, these questions are foreclosed by the adversarial approach our arguments are designed to dampen.

[2] We focus on mechanists such as Craver and Kaplan because the critics explicitly name them as targets. Perhaps this argument, however, finds a better target in other expressions of the mechanistic view of explanation, such as Salmon (1984; 1998), Glennan (2005), Machamer (2004), Bogen (2005), though even these authors could distinguish the metaphysics of causation and theories of explanatory causal relevance. Importantly they might (and Glennan does) conceive the latter of these as interventionist in nature.

1. Phenomenological Laws. Before discussing the relationship between mechanistic explanation and phenomenological laws, we consider some of the various ways the term 'phenomenological laws' has been used. The different uses of the term are often connected with different traditions in the philosophy of science. Thus, the "problem" of phenomenological laws is not one problem but many. We start by describing various ways that the idea of phenomenological laws has been understood, and we discuss how, on these different understandings, the fact that the law is "phenomenological" is supposed to impact its status as an explanation.

Nancy Cartwright (1983), for example, holds that 'phenomenological' laws contrast with 'theoretical' laws. In this sense, which traces to the era of logical positivism, phenomenological laws are law statements articulated exclusively in terms describing observable features of the world. Theoretical laws, in contrast, posit theoretical entities that explain the regularities among observables. Nothing in this notion so far indicates one of the *problems* of phenomenological laws, as there is no general rule that bona fide explanations must posit theoretical entities. For example, the car stopped running (observable) because there was no gas in the tank (observable).

Next, consider the caricature view that phenomenological laws express relations among observations, our perceptions of worldly phenomena, rather than among observables, the objects of those perceptions. This caricature, perhaps endorsed by Mach (1914), really puts the "phenomenology" in "phenomenological." But, if we restrict our attention to explanations outside psychology and other sciences that do, in fact, concern themselves with relations among perceptions, it is clear why phenomenological laws in this sense would seem explanatorily suspect.[3] Namely, the psychological registration of worldly phenomena is, in most cases, epiphenomenal to their occurrence. The car would run out of gas even if nobody looked in the tank. If one holds, reasonably, that explanations (outside of psychology, economics, etc.) cite features of the world rather than our psychological representations of those features, this positivist caricature is at least prima facie suspect.

This caricature does not, however, capture the standard examples motivating the contemporary discussion. Wesley Salmon, a key defender of causal-mechanical views of explanation, considers a child who asks why the ocean waves are creeping slowly toward her sandcastle. He opines, "A very primitive explanation might consist in informing it [the child] of the regular way in which the tides advance and recede" (1998, 60). This "primitive explanation"

[3] We still acknowledge the importance of the logical positivism that followed on Mach's influential perspective.

is phenomenological in that it describes an observable regularity without revealing the laws or causes in virtue of which it holds: tides in, tides out. Unlike the positivist reading, Salmon's phenomenological laws describe relations among worldly magnitudes, rather than among perceptions; but they are epiphenomenal with respect to the explanandum. The low tide does not cause the high tide; rather, both the low tide and the high tide are distal indicators of the causes at work (i.e., the relative position of the moon and the earth). Phenomenological laws in this sense describe mere correlations. To use another of Salmon's examples, there is a phenomenological regularity between falling barometric readings and storms. Again, this phenomenological law fails to cite the common cause (air pressure) of both. The falling mercury is epiphenomenal with respect to storms; the storm would happen even if all barometers were pulverized. Phenomenological laws in this sense express correlations ultimately explained by factors unmentioned in the law statement.

If the phenomenological nature of phenomenological laws is understood positivistically or in terms of merely correlational regularities, then the factors cited in the phenomenological law are epiphenomenal, and so explanatorily irrelevant, to the explanandum phenomenon, as both mechanists and interventionists should insist. If you intervene ideally to prevent low tide, that is, leaving everything else as it is, without disrupting the system's causal structure, it will make no difference to the occurrence of the high tide. Consistent with Premise 2, certain phenomenological law statements are phenomenological because they fail to describe the operative mechanisms that explain why things appear as the statement describes them. But importantly, they also fail to describe causes or constitutively relevant factors producing the explanandum phenomenon. Instead, they describe epiphenomena, correlates, relations among indicators, all causally irrelevant to (though regularly associated with) that explanandum.

For this reason, we should revise the first premise of the target argument: the mechanist need not reject the explanatory value of phenomenological laws on the grounds that they fail to describe mechanisms; rather, they might reject the explanatory power of phenomenological laws on the grounds that they fail to describe explanatorily relevant factors. And what are those?

Many mechanists, such as Craver, Kaplan (e.g., see Craver 2007; Craver and Kaplan 2020) and Glennan (2017), follow Salmon in holding that models explain by describing the causal structure of the world, the causal nexus, i.e., the factors causally or constitutively relevant to the explanandum phenomenon. As Salmon (1984) articulated his view, explanatorily relevant features might be either parts of the causal nexus in the past light-cone of the explanandum phenomenon (the "etiological aspect" of scientific explanation) or parts of the underlying mechanism for the phenomenon (the "constitutive aspect"). A more accurate formulation of Premise 1 would therefore read:

> Premise 1*. Mechanists[4] hold that for a model to count as explanatory for a phenomenon P, that model must describe either a causally relevant antecedent to the occurrence of P or a *constitutively relevant* feature of the mechanism underlying P.

Premise 1 exhibits a blindness toward the distinction between etiological and constitutive explanation and to the fact that several mechanists since Salmon have distinguished and acknowledged both causal (etiological) and constitutive explanations.

Consider an example. Let P be the phenomenon that the kidney regulates plasma osmolality. Etiological explanations might explain P by describing kidney development or how evolution by natural selection shaped the osmoregulatory mechanisms in the kidney. Constitutive explanations, on the other hand, explain P by describing how the activities of parts of the kidney (the glomeruli, the loop of Henle, etc.) are organized, in the here and now, such that the kidney regulates ion concentrations. These are two different ways of situating P in the causal nexus: from before and from within. Mechanists can and do embrace both.

To return to the progress made in Premise 1* (over Premise 1), consider a third sense in which some laws are said to be phenomenological, namely, that they are derivative as opposed to fundamental. Mill (1906) used "phenomenological" in this way. All the phenomenological laws we have considered up to this point are derivative in this sense; they are regularities that depend for their existence on explanatory facts unmentioned in the law statement. But now focus on a specific subset of derivative phenomenological laws, namely, those that are derivative but that nonetheless contain variables that refer to explanatorily relevant factors. Such laws are "phenomenological" insofar as they are derivative, but *not* insofar as they lack causally relevant information. They are thus not representative of one of the *problems* of phenomenological laws.

This is the kind of phenomenological law epitomized by the ideal gas law: $pV = nRT$. This equation describes a generalizable relationship between pressure (p), volume (V), temperature (T), the amount of substance (n) and the ideal gas constant (R). The ideal gas law equation can be used

---

[4] As one reviewer points out, mechanists such as Levy and Bechtel might reject both Premise 1 and Premise 1*. In fact, Levy (personal communication) embraces 1*, but acknowledges that there might be, *in principle*, ways of supplying "understanding" via explanatory representations that are neither causal or constitutive (see Levy 2020). In the strictest sense, we use the term "mechanist" here to refer to theorists such as Craver, Kaplan and Glennan.

to represent relations among observable properties of actual gases. For example, it describes how pressure and volume change as a function of temperature. This law is famously derivative, a derivative consequence of more fundamental laws of statistical mechanics and the molecular theory of gases. It describes relations among magnitudes without describing the mechanism in virtue of which those relations hold. According to Premise 1 (versus Premise 1*), the mechanists should count that law as non-explanatory. But we need to slow down: non-explanatory of what?

Suppose we want to explain why Anna's balloon popped. And suppose we learn that she had been heating it with a hair dryer. The ideal gas law can now do some serious explanatory work, for in this employ, it describes a causal regularity among magnitudes (rather than epiphenomena). One can defend the explanatory value of this law from the perspective of many different theories of scientific explanation (for a review, see Woody 2013). Salmon, in contrast, rejected its explanatory value (see Salmon 1984, 136, 1998, 55). This is because he held that etiological explanations describe causes and, further, that causation involves mark transmission or exchanges of conserved quantities linking cause to effect in an unbroken chain of intersections in spacetime. In the case of gases, these marks and quantities are borne by the molecules composing the gas. The gas law abstracts away from these causal mechanisms, and so fails on his account to describe why these variables stand in this relation to one another.

But Salmon arguably need not have said this, even on the assumption of his contact-action theory of causation. He might rather have tried to develop an account of how higher-level causal relevance is possible in a world that is, fundamentally, the propagation and transmission of conserved quantities and contact action. This would require making temperature (and not just the motion of molecules) part of the past light-cone of the balloon's popping. Lewis (1986) does precisely this: the heating is explanatory in virtue of being a counterfactually relevant part of the causal history of the popping.

Contemporary mechanists such as Craver and Kaplan adopt neither Salmon's nor Lewis' approach to this problem exactly, though they are closer to Lewis' than to Salmon's. They embrace, instead, Woodward's (2003) theory of causal (and so explanatory) relevance (see Craver 2007, Chapters 3 and 6; Craver and Kaplan 2020; Kaplan and Craver 2011; Kaplan and Bechtel 2011). According to the interventionist view, the gas law is explanatory because you can manipulate the volume of the balloon by intervening ideally (in a sense that need not detain us) on the temperature of the gas. That's what Anna does with her ideal hairdryer. The ideal gas law is explanatory because it provides counterfactual dependency information, answering *what-if-things-had-been-different questions*, i.e., ‘w-questions’ (Woodward 2003). The ideal gas law indicates roughly how volume would have been different given ideal interventions on temperature, namely, in accordance with $pV = nRT$.

On Salmon’s way of fleshing out the causal interpretation of this law, the gas law statement describes superficial features and leaves out the crucial causal work done by gas molecules (see Salmon 1998, 56). But on Woodward’s and Lewis’ way of cashing it out, there is no problem: causation need not involve contact action; omissions and double preventions can be causally relevant; causes are difference-makers. Craver, Kaplan, and Glennan all adopt this interpretation themselves, so there is no disagreement about how to handle etiological explanations that appeal to phenomenological laws.

As we’ll see in the next section, Premise 1 would appear to be true of mechanistic theories of *constitutive* explanation even if, as we’ve seen, it is not true of their theories of etiological explanations. Perhaps this explains Woodward’s and Rescorla’s subtle confusion. Concerning both kinds of explanation, Premise 1* is a much more accurate summary of the mechanist approach we endorse (e.g., mechanists such as Craver, Kaplan and Glennan).

2. Phenomenal Models and Mechanistic Explanations. Because the ideal gas law is a derivative law, it expresses a relationship among magnitudes that has a deeper explanation. We are thus faced with the following question: Why is the ideal gas law (approximately) true (when it is)? When we ask this question, the law describes the explanandum phenomenon, not the explanans. It is in this kind of explanatory context, where a constitutive explanation is called for, that the problem of phenomenal models arises. This problem is altogether different from the problems of phenomenological laws. Indeed, the “problem” of phenomenal models is not so much a problem as it is an apparent truism about explanation: explanandum phenomena don’t explain themselves, or more accurately, you can’t explain something merely by describing it differently. Only some redescriptions have explanatory content.

What, exactly, is a phenomenal model? Craver describes phenomenal models as, “complete black boxes; they reveal nothing about the underlying mechanisms and so merely ‘save the phenomenon’ to be explained” (2006, 360). He says that, “the signature of a phenomenal model is that it describes the behavior of the target system without describing the ontic structures that give rise to it (2014, 38-39)."[5] More recently, Kaplan and Craver define phenomenal models as follows: "a model, M, is a phenomenal model of a mechanism if and only if it describes the inputs to, modulators of, and outputs from a mechanism without describing its relevant internal causal structure" (2020, 296).[6] A phenomenal model (or law, if that be preferred) is phenomenal in virtue of what it leaves out: it describes the behavior of the mechanism without describing anything about how the mechanism works. The root, "phenom," can lead one to mistake these two, altogether distinct concepts.

To mark the divide between purely phenomenal models and constitutive explanatory models, Kaplan and Craver (2011) impose the "model-to-mechanism mapping" (3M) requirement on constitutive explanations: for the model to go beyond a purely phenomenal redescription, it must reveal some details about the underlying mechanism. This requirement does not assert (despite what some critics allege) that an explanatory model must describe all those details. Nor does it assert that a model is more explanatory to the extent that it describes more of those details (see Craver and Kaplan 2020). Rather, it simply marks the divide between models that function only to fix the explanandum phenomenon and models that provide information about the constitutive explanation.

The problem of phenomenal models arises only in the context of constitutive explanations and does not arise for etiological explanations. The problem with phenomenal models in this context is that they, by definition, provide no information about how the activities of component entities are organized such that the relation specified in the explanandum holds. Why does the sleeping pill help you fall asleep? Because it has a "dormitive virtue". How does the hippocampus encode memories? Because it has the function by which memories are encoded. These putative constitutive explanations are hollow; they are purely phenomenal in the sense that they merely redescribe the explanandum phenomenon without telling us anything about the mechanism, the nexus of causes in virtue of which the explanandum holds true.

We must be mindful here of the distinction between etiological and constitutive explanations. If one wants to know why Anna's balloon popped, the gas law can be used as part of an etiological explanation, along with her hairdryer. But if someone asks why gases expand

[5] Also see Glennan (2002, 2005) for a related discussion of the distinction between phenomenal and mechanistic descriptions.
[6] Precisely this view is articulated in Kaplan and Hewitson's (2021) discussion of Bayesian models, a paper we discovered late in drafting this article. It is entirely complementary to the more abstract views expressed here.

when heated, the ideal gas law is explanatorily impotent. It exchanges a verbal description of the thing that we want to explain for a more precise, mathematical characterization of the same phenomenon. That is a kind of intellectual progress, but it doesn't explain why a change in temperature should yield a change in volume. In the context of constitutive explanations, phenomenal models provide no explanatory nourishment beyond that contained in the request for explanation. To take them as explanatory is to run in an explanatory circle, to explain a phenomenon by appealing to that very phenomenon.[7]

To be sure, phenomenal models are not useless *simpliciter*. We saw above that they can be deployed in etiological explanations (e.g., the ideal gas law). But even in constitutive explanations, one can make explanatory progress by characterizing the explanandum phenomenon. For David Marr (1982) the primary value of computational-level descriptions in cognitive science is that they characterize precisely the cognitive operations one wants to explain; they fix the target of the explanation. One way to start the search for a mechanism (Marr thought it was the best way) is to study the phenomenon, to master the constraints on any acceptable explanation for it, and to use that model as a spring-board for conjecturing and eliminating possible and impossible explanations for it (see Bechtel and Richardson 1993; Craver and Darden 2013; see also Shagrir 2010). Phenomenal models can also help scientists and students *conceptualize* a phenomenon and guide their reasoning about it. As Woody argues, "by providing selective attention to certain gas properties and their relations and ignoring other aspects of actual gas phenomena, the ideal gas law effectively instructs chemists in how to think about gases *as gases"* (2013, 1574).

Return now to our target argument, this time with constitutive rather than etiological explanations in mind. Again, Premise 1 is not so much incorrect as imprecise. Premise 1* is better, though here we focus on its second disjunct, i.e., on constitutive explanations. Premise 2 is also correct; but we should be clear that we are talking about *merely* phenomenal models as Craver and Kaplan define them: they describe the explanandum phenomenon but, by definition, fail to include details about the mechanism by virtue of which the phenomenon holds. So, it is clear that we should reject the conclusion C. If all parties agree with the apparent truism that merely re-describing an explanandum phenomenon does not suffice to explain that phenomenon (and we

[7] Stuart Glennan (personal communication) points out that this point can be put differently by distinguishing the explanandum phenomenon rather than types of explanation: are we explaining the balloon's popping (an event) or are we explaining why gasses expand when heated (a regularity). We express it in terms of kinds of explanations because regularities can also both be given etiological and constitutive explanations.

have seen little opposition to that thought),[8] then the problem of phenomenal models, described by mechanists, cannot be used to support C. While mechanists hold that phenomenological laws, such as the ideal gas law, are explanatorily impotent in the context of constitutive explanations, this does not entail that they are non-explanatory *simipliciter* (e.g., as etiological explanations). In this way, C is transparently false.

3. Phenomenal Models in the Cognitive Neurosciences: The Mechanist's Response. Woodward (2017) and Rescorla (2018) especially criticize the mechanistic framework as applied to the cognitive neurosciences, a domain Craver and Kaplan emphasize in their development of the view. If the framework fails there, this is a serious problem. However, as we now explain, Woodward and Rescorla's criticisms do not reflect an appreciation of the ideas in the previous two sections. Neither acknowledges the distinction between etiological and constitutive explanation and, as a result, each embraces the confusing ambiguity of Premise 1. Coupled with an inaccurately oppositional framing of the dialectical situation, these missteps lead to fundamental confusions about scientific explanation in these sciences and beyond.

3.1 Rescorla offers his version of the argument to support his interventionist analysis of explanation in psychology: "scientific psychology causally explains mental and behavioral outcomes by specifying how those outcomes would have been different had an intervention altered various factors, including relevant psychological states" (2018: 1920).

Rescorla uses Bayesian models of perception to exemplify the power of his approach. In general terms, the perceptual system represents distal features like shape, color, size, spatial location, etc., on the basis of proximal sensory stimulation (e.g., light on the retina). These models are Bayesian in the sense that they rely in their estimations on the prior probabilities of these features and on the likelihood of the retinal stimulation given the presence or absence of these features. We suppose with Rescorla that these models accurately describe the relationship between these proximal stimuli (inputs) and the downstream perceptual representations (outputs).

[8] One reviewer reads Silverstein, Chemero and Bechtel as providing opposition to this claim. Anyone who supports that view must distinguish the kinds of redescription involved in the case of dormitive virtues from the explanatory kinds of redescription.

Suppose we are giving an etiological explanation of why John perceives a green circle. The Bayesian model tells us how the perceptual estimate would have been different given different proximal stimulation, different background priors, and different likelihood values (e.g., if the subject had been presented with different patterns of lighting in the past) (2018, 1926). Hence, under the interventionist criteria, Bayesian models play an explanatory role because they allow us to answer a set of w-questions: how would the estimate differ if the pattern of light on the retina were different? How would it differ if one's life experience had resulted in different priors? The ability to answer such w-questions, as we discussed in Section 2, is the mark of the Bayesian model's explanatory power.

So far, mechanists and interventionists can agree, given that mechanists adopt the same model of causal relevance as Rescorla (as do Craver, Kaplan and Glennan). But Rescorla goes further and argues that Bayesian models are "highly non-mechanistic" (2018, 1935) and thus are ruled illegitimate by mechanistic theories of explanation. Exemplifying our target argument, he begins with the ambiguous Premise 1, asserting that mechanists require all explanations (rather than just constitutive explanations) to describe underlying mechanisms. His Premise 2 is that Bayesian models are phenomenal models; they describe regularities and do not describe the neural mechanisms in virtue of which those regularities hold. According to Rescorla, mechanists should falsely view Bayesian models as non-explanatory; he concludes C.

Rescorla drives home his argument about Bayesian models by comparing them to the ideal gas law (IGL):

> A recurring worry facing the mechanistic conception of scientific explanation is that many successful scientific explanations seem non-mechanistic. Consider the ideal gas law: pV = nRT, [IGL]…From a mechanistic perspective, [IGL] is not explanatory. A genuinely mechanistic explanation must instead deploy statistical mechanics, which describes the gas as a collection of tiny interacting particles. I acknowledge that statistical mechanics can augment the explanatory power of …[IGL]. It does not follow (and is not true) that [IGL] itself is unexplanatory. (Rescorla 2018, 1914).

In short, Rescorla's argument is a near perfect instantiation of our target argument, embodying all the elementary yet subtle errors we have now exposed.

First, ask the crucial starting question: What is being explained? Is IGL supposed to explain the popping of Anna's balloon? Or is to supposed to explain why gases expand when heated? If the explanandum is the former, that is an etiological explanation, and IGL has considerable explanatory power. It is false that Craver, Kaplan and Glennan, for example, believe otherwise (see Premise 1*). If the explanandum is the latter, then we are dealing with a constitutive explanation, IGL is explanatorily impotent, and mechanists are correct in so judging.

These simple mechanist responses apply, mutatis mutandis, to the Bayesian models. What is being explained? Are we trying to explain why an organism makes a particular perceptual judgement in response to a stimulus? In that case, the model tells us that we should look to the proximal stimulus, to their cumulative experience as stored in their priors, to their assessments of the likelihoods, etc. That is an etiological explanation. The Bayesian model (we presume) expresses a true causal generalization about the explanandum, and mechanists see this as a legitimate causal explanation. According to mechanists like Craver and Kaplan, causal phenomenological laws can describe appropriately constrained counterfactual dependency relations between magnitudes or properties. Are we instead trying to explain why the visual system behaves approximately in accordance with Bayes' rule? If so, the Bayesian model leaves us explanatorily malnourished. In the context of constitutive explanations, phenomenological laws are by definition mere redescriptions; they do not explain why the counterfactual dependency relations they describe hold. If we are looking for a constitutive explanation, this will require details about the underlying mechanisms. Nothing in Rescorla's argument gives any reason to deny this.

The distinction between etiological and constitutive explanation matters, not just in philosophy, but in the science itself. Consider again Rescorla's discussion of the success of Bayesian models of perception. As he notes, such models accurately describe counterfactual relationships between prior probabilities and perceptual estimates. If so, these models can function as part of an etiological explanation. However, a Bayesian model does not explain why such counterfactual relationships hold. That explanation might itself be supplied etiologically (in terms of development or evolution) or constitutively, in terms of how parts of the brain and other physiological systems involved are organized to instantiate that set of counterfactual relations. Indeed, scientists are beginning to understand the mechanisms that produce perceptual estimations, as described by Bayesian models.[9] For example, the capacity for calculating prior probabilities is

[9] For a detailed example see Kaplan and Hewitson (2021). Also, Woodward (personal communication) suggests that facts about the constitutive explanation are the initial/boundary conditions that must hold for the explanatory generalization to be true. This is correct, though there are many more initial/boundary conditions than there are mediators of the input-output relation, i.e., mechanisms. Regardless, this thought perhaps provides an important

thought to involve neural tuning to common physical features of the environment. In a recent review, de Lange et al. (2018) note that "prior expectations (or priors) are likely to be learnt over relatively long timescales, leading them to become encoded in the tuning properties of our sensory cortices" (2018, 766). Furthermore, they note that future research should investigate "how all the prediction signals from different sources are ultimately combined during perceptual inference…One possibility is that all regions send their predictions to a shared 'blackboard' that resides in the primary visual cortex, facilitating the combining of different priors" (2018, 775).

Here, De Lange et al. are calling for mechanistic details about how components and their activities are organized to bring about the phenomena described by Bayesian models of perception. By including mechanisms of neuronal tuning in a model for the capacity of perceptual estimation, we begin to grasp how prior interactions with the physical environment influence the expectation of distal features. (We leave for now whether "tuning" is more than a promissory note, a neural dormitive virtue). In other words, this is at least a step toward a constitutive explanation of the phenomena described in Bayesian models of perception. And the subsequent development of these models shows scientists responding to an explanatory need Bayesian models themselves cannot satisfy.

Rescorla conflates two distinct forms of causal-mechanical explanation. Interventionism has often concerned itself exclusively with etiological explanations, and mechanists have tended to focus on constitutive explanations. To compare these accounts with respect to etiological explanations alone is inappropriate. This is why Premise 1 must be revised into Premise 1*. Since it is implausible that Rescorla thinks phenomenological laws explain themselves, it's more likely that he simply has not made use of the etiological/constitutive distinction as it should be applied in the current context.

3.2 Woodward (2017) provides a similar argument to Rescorla's. Woodward writes:

> Consider the very common use of models involving recurrent networks with auto-associative features to explain phenomena like retrieval of memories from partial cues…Such models obviously abstract away from many neural details, and in this respect are relatively non-mechanistic in Craver's sense. On my view, however, we should not conclude that they are unexplanatory for

indicator of work to be done to properly integrate mechanistic approaches with the interventionist framework and vice versa.

> this reason alone. Instead, their explanatory status depends on whether they accurately capture the dependency relations in real neural structures (2017, 88-89).

Woodward's exegetical error is analogous to Rescorla's and traces to a basic misreading of Craver and Kaplan's views. Woodward plausibly construes such models as providing etiological, as opposed to constitutive, explanations. It is also plausible that he thinks the models are true (enough) in the relevant respects of the actual systems (e.g., do neural mechanisms deploy auto-association?).[10] If the explanation is etiological, and if the descriptive and counterfactual claims in these models are true (enough), then mechanistic views, as we've discussed, should entail that these models express explanatorily relevant information.

On the other hand, if the explanation he had in mind *was* constitutive, and neural network models are still true (enough) descriptions of the components and processes involved in "retrieval of memories from partial cues," then again, the mechanist view entails that the model contains explanatory information. If, however, the models are merely accurate summaries of the inputs and outputs of retrieval processes, and the components in the model bear no (or are intended to bear no) relation to actual neural mechanisms (i.e., to their parts, activities or organizational features), then it would seem everyone should agree that, within the context of constitutive explanation, they are phenomenal models with no explanatory import. In other words, if the model doesn't provide constitutive information about the causal mechanism underpinning memory retrieval from cues, then it is merely phenomenal and so explanatorily vacuous, like a dormitive virtue.

However, if the model at least represents organizational features of the relevant memory retrieval mechanism (as a plausible reading of Woodward suggests), even if the model abstracts away from details regarding specific neural parts, then this is not a merely phenomenal model. Rather, it is a form of genuine constitutive explanation, however incomplete and unsatisfying it may be. In such a case, mechanists such as Craver and Kaplan would not dismiss the model as non-explanatory, but rather would say something about the quantity and quality of causal information the model supplies.

Woodward provides another argument that doesn't appeal to any particular example of a phenomenological law. This involves a way to demarcate "purely descriptive or phenomenological models, and, on the other hand, explanatory models" (2017, 90). In this case, Woodward discusses

[10] Woodward confirmed these two assumptions in personal correspondence.

how to demarcate genuinely explanatory from non-explanatory, i.e., merely correlational or observational, models. This is indicative of some of the problems of phenomenological laws discussed in Section 1. Woodward falsely claims that mechanists use mechanistic details as the distinguishing factor (2017, 90), i.e., that they embrace Premise 1. Woodward then provides a variety of ways in which a model might fail to be explanatory (2017, 89-94). These are roughly ways in which the interventionist criteria for explanation are not met. Woodward summarizes:

> We can explain what is explanatorily defective about such models in terms of violations of basic interventionist/dependency requirements on explanation without invoking the idea that all explanations must be mechanistic. To the extent that a model avoids the problems described under 1–5 above, and satisfies the interventionist constraints on explanation, it will count as explanatory even if it fails to be mechanistic (2017, 94).

But, to reiterate, mechanists such as Craver and Kaplan adopt the interventionist framework for explanatory relevance. Several mechanists can and do agree that interventionist criteria supply a key notion of causal (and so explanatory) relevance and so are useful for sorting good from bad explanations (see, e.g., Craver 2007, Chapter 3).

Any model that meets the interventionist criteria will contain counterfactual dependency information. It will thus outline part of the causal structure of the world and hence be potentially explanatory of something. But, again, of what? If the counterfactual dependency information correctly describes the relation between an effect and its cause, then such models can be used in etiological explanations of the effect. But does the model provide a constitutive explanation of why that counterfactual relationship holds? That depends on whether the model contains, in addition, details about the mechanism by which that relationship holds. If we are trying to explain a phenomenon P, and our model describes P alone, without giving any of the mechanistic details in virtue of which P occurs or any of the causes of P, then the mechanist is committed to that explanation's explanatory impotence.

3.3 How did this confusion arise? Above we hinted at a possible historical source in Salmon's idiosyncratic contact-action-based articulation of the causal-mechanical view of explanation. Although Salmon (1984) was, as far as we know, the first to distinguish etiological and constitutive aspects of scientific explanation (unlike his contemporary, David Lewis), he dedicated the vast majority of his attention to etiological explanation and, in fact, never developed a working model of constitutive explanation.

There is good reason for that: his account of causal (and so explanatory) relevance does not apply to (indeed, cannot apply to) the constitutive explanatory relation between parts and wholes. There are general reasons to be suspicious of the idea of part-whole causal interactions, as many philosophers have emphasized (see, for example, Lewis 1986; Craver and Bechtel 2007; Kim 1992). But whatever general suspicions there are, Salmon had a particularly acute difficulty (as Craver 2007 notes): his understanding of causation involved causal processes intersecting in space-time, exchanging either marks (Salmon 1984) or conserved quantities (Salmon 1992) with one another, and carrying those marks or quantities beyond the intersection point. It is at least very hard to make sense of a whole receiving conserved quantities it did not already have from its parts or to understand how a whole might come to carry a mark it did not have, though one of its parts had from the start. And given that wholes and parts are always everywhere intersecting one another in space and time, it is hard to see how they would come to intersect or continue to carry marks and conserved quantities once those intersections cease. In short: His view of causation prohibits interlevel interactions; interactions are the basis of his theory of relevance; constitutive explanation must be utterly mysterious.

Phenomenal models describe regularities without describing the billiard-ball style causal interactions that, on his account, fully constitute the explanation. So, Salmon was thus primed to assert that phenomenal models are universally explanatorily vacuous (even in etiological contexts). If one presumes incorrectly that contemporary mechanists (specifically Craver and Kaplan) follow Salmon in this regard (and indeed, some have, see Machamer 2005; Bogen 2005), one might conclude that they similarly must insist on the explanatory inadequacy of phenomenal models.

Perhaps, that is, Woodward has simply not turned the page from Salmon to contemporary mechanists, and so finds himself responding to echoes of a previous generation. Consider how Woodward's (1989) review of Salmon's *Scientific Explanation and the Causal Structure of the World* (1984) parallels his critiques of Craver thirty years later:

> It is a central claim of much theory in cognitive psychology that such accounts can provide genuine explanations even though they do not describe in detail the operation of neurophysiological or biochemical mechanisms and even though similar information-processing strategies may have interestingly different neurophysiological realizations in different subjects. None of these

> explanations seems to explain by tracing in detail continuous causal processes or underlying physical mechanisms…I think it is at least not obvious…how Salmon can avoid the conclusion that many of the above theories are pretty dubious as explanations, in virtue of their apparent failure to specify continuous causal processes (1989, 366).

Woodward argues that Salmon cannot accommodate the use of phenomenal models in cognitive psychology. Here we are using "phenomenal models" in the sense described in Section 2. Unlike Craver and Kaplan, Salmon *does* seem to hold that explanatory models must specify the underlying causal processes underpinning some phenomenon. However, Craver and Kaplan break with this idea and embrace an interventionist account of causation as a useful component of their view of mechanistic explanation. Woodward's criticisms of Salmon cannot simply be transferred, mutatis mutandis, to the work of contemporary mechanists.

In summary, Woodward (2017) and Rescorla (2018) both rely on the target argument in their criticism of mechanistic views of explanation in cognitive neuroscience. In doing so, they fail to mark the crucial distinction between etiological and constitutive explanations, perpetuating a blind-spot in interventionist theories generally concerning constitutive explanation. Given the ubiquity and import of constitutive explanations, not just in cognitive neuroscience but in science generally, this blind-spot renders interventionism incomplete as a theory of explanation in the special sciences. The tendency to cast interventionist and mechanistic theories as opponents prevents the interventionist from hearing what mechanists since Salmon have had to say about this central and neglected aspect of scientific explanation.

4. Beyond Conflict. In this section, we demonstrate the importance of recognizing that interventionism and mechanistic accounts should not be seen as competitors. To illustrate this importance, we describe two issues related to scientific explanation: the problem of constitutive explanatory relevance and the problem of finding the appropriate level of explanation. In our view, neither interventionism nor mechanism can solve these problems on their own; they require collaboration.

The adversarial framing, of mechanists pitted against interventionists, is not unique to interventionists such as Woodward and Rescorla; it has been invited by writings of certain mechanists, for example, by Machamer (2004), Bogen (2005), and (at times) Darden (2008). These mechanists describe interventionist counterfactuals as "behaviorist" input-output relations, impoverished in articulating how the input is transformed into output. Sometimes this contrast is expressed in an alliance between mechanistic theories and Anscombe's (1971) view that "causation" is a philosophical abstraction that gains content only in diverse activities such as burning, scraping, and pushing. Whatever the merits of such substantive and singular theories of causation for thinking about discovery and about the metaphysics of causation, these notions have proved to be of limited use in articulating the idea of explanatory relevance (see Craver 2007, Chapter 2).

In contrast with mechanists such as Machamer and Bogen, explicit appeal to interventionist counterfactuals lies at the causal heart of the theory of mechanism supplied by Craver (2007), Kaplan (2011), and Glennan (2017) and is the basis for their discussion of higher-level explanations (see Craver 2007, Chapter 6) and higher-level laws, such as dynamical equations (see Kaplan and Bechtel 2011). The novelty in the work of these interventionist-mechanists lies not in the *contrast* with Woodward's view but rather their efforts to supplement it with an account of constitutive explanation. Woodward's interventionism, like its counterfactual antecedents (Lewis 1986), has appeared blind to constitutive explanation generally. Seen in this light, mechanists have something to offer interventionists; they are not adversaries.[11] A theory of explanation that has nothing to say about constitutive explanation simply cannot be adequate to the cognitive neurosciences.

Consider, then, some research topics that emerge once competition is abandoned for symbiosis. First, central to a theory of constitutive explanation is an account of constitutive explanatory relevance. Causal accounts of etiological explanation (including interventionists and mechanists alike) tend to identify etiological explanatory relevance with causal relevance: to be etiologically explanatorily relevant to some event is to be part of the causal history of that event (Lewis 1986), to be a causal process in the past-light cone of the event (Salmon 1984), or to be related via interventionist counterfactuals to the event (Woodward 2003). Yet when we turn our attention to constitutive explanations, in which the behavior of a whole (e.g., a gas) is explained in terms of the activities and arrangements of its parts (e.g., the molecules), one cannot straightforwardly appeal to causal relations between the explanans and the explanandum. This is because each of

[11] In fact, even mechanistic critics of Woodward (e.g., Machamer and Bogen) could, like Glennan (2017) accept interventionism as an account of explanatory relevance.

these authors either explicitly rejects the appropriateness of causal language for describing the relationship between parts and wholes or ignores the topic entirely.

For example, Lewis (1986) explicitly rejects the possibility of causal relations between parts and wholes, claiming it "will not do" to assert that speaking the sentence causes the speaking of its first half. This is so despite the fact that Lewis includes the parts of an event in his definition of the causal history of the event and asserts that to explain is to describe a portion of its causal history. As discussed briefly above, neither of Salmon's views of causation, the mark transmission theory nor the conserved quantity theory, can countenance part-whole causal relations.

Woodward's interventionist theory, as currently articulated, applies to cases in which the value of one variable is explained in terms of the value of another variable. It is not equipped to discuss the relationship between a *variable* describing a part and an input and output *regularity* characteristic of the whole. To hammer this topic into the standard form of an interventionist explanation, one might suppose that one can describe a regularity (such as completing memories from partial prompts) as a variable, or perhaps as a capacity (such as memory completion from prompts, to use Woodward's example) that can be on or off.

Now consider some possible interventionist counterfactuals relating that capacity to other variables: (i) when we ideally intervene to engage the capacity, a brain region is differentially activated in an MRI scanner; (ii) when we ideally remove brain region X, subjects can no longer engage the capacity; (iii) when we ideally stimulate brain region X, subjects report apparently complete memories. Anyone who has spent time thinking critically about findings in cognitive neuroscience knows that these counterfactuals can each be true without saying anything explanatorily relevant about the mechanisms of memory completion. (i) can be true even if the brain region is irrelevant, for example, its activation is downstream from a relevant part. (ii) can be true, but due to swelling in an area adjacent to the brain lesion. (iii) can be true due to spreading activation from brain region X to neighboring regions. Our point is not that the interventionist cannot find solutions to these problems (mechanist interventionists such as Craver have invested significant energy in that topic). Rather it is this: one cannot simply assume that the interventionist theory developed to handle etiological explanations applies without further elaboration to its constitutive cousin. There are too many true interventionist counterfactuals that are explanatorily empty.[12]

One of the major advances of the mechanistic theory of explanation is the centering of this topic in discussions of explanation in the special sciences. There are now many options for understanding constitutive relevance, some of which are interventionist in character (Craver 2007; Craver, Glennan and Povich 2021; Harinen 2018; Krickel 2018; Prychitko 2021; Ylikoski 2013), others of which are more conceptual and make use of Woodward's "requirement of independent fixability" (Baumgartner and Casini 2017), and others of which hope to do all the work required of a theory of constitutive relevance with suitably restricted non-interventionist counterfactuals (Couch 2011; Harinnen 2018). Which of these views is ultimately successful remains a matter of active discussion.

Suppose now that an interventionist solution to this problem is viable. Another important question, of some relevance to scientists, is this: is there a correct level at which a given constitutive explanation should stop? Do all constitutive explanations bottom out in fundamental physics? Or are some explanatory problems solved at intermediate levels? Every textbook in cognitive neuroscience, even those authored and edited by die-hard cellular-molecular reductionists, recognizes the importance of neuroscientific work at multiple "levels of organization." Is there a general answer to the question of how we decide on the right level for solving a given constitutive explanatory problem?

Interventionists have addressed an analogous problem for etiological explanation, such as the "problem of mental causation" (e.g., Heil and Mele 1993; Kim 1998), or biological causes, or the problem of social-level causes (Hedstrom and Ylikoski 2010). They do so by appealing to the proportionality of a cause to its effect (Yablo 1992) and to the specificity and sensitivity of the cause to the effect (Woodward 2010). To what extent can solutions developed for thinking about the correct way to formulate an etiological explanation be retooled in the interlevel domain of constitutive explanations? This remains an open question. Yet it is a question that remains invisible so long as we refuse to acknowledge the crucial distinction between etiological and constitutive forms of explanation.

[12] Also see Craver (2014) and Povich and Craver (2017) who offer a separate reason for rejecting interlevel causation.

5. Conclusion. The tendency to confuse the problem of phenomenological laws with the problem of phenomenal models, as well as the failure to distinguish etiological from constitutive causal-mechanical explanations, has led to copious confusion in debates about the nature of scientific explanation. Our aim has been to expose and undo these confusions and chart a path forward. A key upshot is that the mechanistic view can be and should be (as opposed to Salmon 1984; Machamer 2004; Bogen 2005) viewed as a helpful companion, rather than a competitor to, the interventionist framework as an aspect of a causal-mechanical explanatory worldview. By distinguishing these problems and developing the mechanistic response to each, we have shown that the criticisms of Woodward and Rescorla are premised on simple misunderstandings of the (interventionist) causal-mechanical view and its significance. Removing this misunderstanding is the first step toward a productive synthesis that promises fecund exploration in coming years.

References

Anscombe, G. E. M. 1971. *Causality and determination: An inaugural lecture*. CUP Archive.

Baumgartner, Michael, and Lorenzo Casini. 2017. An abductive theory of constitution. *Philosophy of Science*, 84, no. 2: 214-233.

Bechtel, William and Abrahamsen, Adele. 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), 421-441.

Bechtel, William and Richardson, Robert C. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research.* Princeton University Press.

Bogen, Jim. 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(2), 397-420.

Cartwright, Nancy. 1983. *How the laws of physics lie.* OUP.

Couch, Mark B. 2011. Mechanisms and constitutive relevance. *Synthese*, 183, 375-388.

Craver, Carl F. 2006. When mechanistic models explain. *Synthese*, 153(3), 355-376.

Craver, Carl F. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. OUP.

Craver, Carl F. 2014. The ontic account of scientific explanation. In *Explanation in the special sciences*, 27-52. Springer, Dordrecht.

Craver, Carl F. and Bechtel, William. 2007. Top-down causation without top-down causes. *Biology & philosophy*, 22(4), 547-563.

Craver, Carl F. and Kaplan, David M. 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 F., Glennan, Stuart and Povich, Mark. 2021. Constitutive relevance & mutual manipulability revisited. *Synthese*, *199*(3), 8807–8828.

Darden, Lindley. 2008. Thinking again about biological mechanisms. *Philosophy of Science*, *75*(5), 958–969.

De Lange, Floris P., Heilbron, Micha, & Kok, Peter. 2018. How do expectations shape perception?. *Trends in cognitive sciences*, 22(9), 764-779.

Glennan, Stuart. 2002. Rethinking mechanistic explanation. *Philosophy of science*, 69(S3), S342-S353.

Glennan, Stuart. 2005. Modeling mechanisms. *Studies in history and philosophy of science part C: studies in history and philosophy of biological and biomedical sciences*, 36(2), 443-464.

Glennan, Stuart. 2017. *The new mechanical philosophy.* OUP.

Harinen, Totte. 2018. Mutual manipulability and causal inbetweenness. *Synthese*, *195*(1), 35–54.

Hedström, Peter, and Petri Ylikoski. 2010. Causal mechanisms in the social sciences. *Annual review of sociology* 36: 49-67.

Heil, John, & Mele, Alfred. 1993. Mental causation. *Philosophy of Mind*, 214.

Hempel, Carl G. 1965. *Aspects of scientific explanation* (Vol. 1). New York: Free Press.

Kaplan, David M. 2011. Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339–373.

Kaplan, David M. and Bechtel, William. 2011. Dynamical models: An alternative or complement to mechanistic explanations? *Topics in Cognitive Science*, *3*(2), 438–444.

Kaplan, David M. and Craver, Carl F. 2011. The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, *78*(4), 601–627.

Kaplan, David M. and Hewitson, Christopher L. 2021. Modelling Bayesian computation in the brain: unification, explanation, and constraints. In *Neural Mechanisms*, 11-33. Springer, Cham.

Kim, Jaegwon. 1992. ‘Downward causation’ in emergentism and nonreductive physicalism. In *Emergence or Reduction?*, 119–138.

Kim, Jaegwon. 1998. *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT press.

Kitcher, Philip. 1989. Explanatory unification and the causal structure of the world. In *Scientific Explanation*, 410-505. University of Minnesota Press.

Krickel, Beate. 2018. Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A*, *68*, 58–67.

Levy, Arnon. 2020. Metaphor and Scientific Explanation. In *The Scientific Imagination*, 281-303. OUP.

Lewis, David K. 1986. Causal explanation. In *Philosophical Papers: Volume II*. Oxford: Oxford University Press.

Lycan, William G. (1990). The Continuity of Levels of Nature. In *Mind and Cognition: A Reader*, 77–96. Blackwell.

Mach, Ernst. 1914. *The analysis of sensations, and the relation of the physical to the psychical.* Open Court Publishing Company.

Machamer, Peter. 2004. Activities and causation: The metaphysics and epistemology of mechanisms. *International studies in the philosophy of science*, 18(1), 27-39.

Machamer, Peter, Darden, Lindley and Craver, Carl F. 2000. Thinking about mechanisms. *Philosophy of science*, 67(1), 1-25.

Marr, David. 1982. *Vision: A computational investigation into the human representation and processing of visual information.* Henry holt and co. Inc.

Mill, John S. 1906. *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*. Longmans, Green.

Povich, Mark and Craver, Carl F. 2017. Mechanistic levels, reduction, and emergence. In *The Routledge handbook of mechanisms and mechanical philosophy*, 185-197. Routledge.

Prychitko, Emily. 2021. The causal situationist account of constitutive relevance. *Synthese*, 198(2), 1829-1843.

Rescorla, Michael. 2018. An interventionist approach to psychological explanation. *Synthese*, 195(5), 1909-1940.

Ross, Lauren N. Forthcoming. Cascade versus mechanism: The diversity of causal structure in science. *British Journal for the Philosophy of Science.*

Salmon, Wesley C. 1984. *Scientific explanation and the causal structure of the world.* Princeton University Press.

Salmon, Wesley C. 1998. *Causality and explanation.* OUP.

Schaffner, K. F. 1993. *Discovery and explanation in biology and medicine.* University of Chicago press.

Scriven, Michael. 1975. Causation as explanation. *Nous*, 3–16.

Shagrir, Oron. 2010. Marr on computational-level theories. *Philosophy of science*, 77(4), 477-500.

Shapiro, Lawrence A. 2020. Mechanism or bust? Explanation in psychology. *The British Journal for the Philosophy of Science.*

Simon, Herbert A. 1969. *The Sciences of the Artificial* (3rd ed.). MIT Press.

Wimsatt, William C. 1972. Complexity and organization. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1972, 67–86.

Wimsatt, William C. 1974. Reductive explanation: A functional account. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1974*, 671–710.

Weiskopf, Daniel A. 2011. Models and mechanisms in psychological explanation. *Synthese*, 183(3), 313-338.

Woodward, James. 1989. The causal mechanical model of explanation. *Minnesota studies in the philosophy of science*, 13.

Woodward, James. 2005. *Making things happen: A theory of causal explanation*. OUP.

Woodward, James. 2013. Mechanistic explanation: Its scope and limits. In *Aristotelian Society Supplementary Volume* (Vol. 87, No. 1, pp. 39-65). Oxford, UK: Blackwell Publishing Ltd.

Woodward, James. 2017. Explanation in neurobiology. In *Explanation and integration in mind and brain science*, 70-100. OUP.

Woody, Andrea I. 2013. How is the ideal gas law explanatory?. *Science & Education*, 22(7), 1563-1580.

Yablo, Stephen. 1992. Mental causation. *The Philosophical Review*, 101, no. 2: 245-280.

Ylikoski, Petri. 2013. Causal and constitutive explanation compared. *Erkenntnis*, *78*(2), 277–297.