ZFN Philosophical Problems in Science

ZFN Zagadnienia Filozoficzne w Nauce

Copernicus Center

Copernicus Center PRESS

Philosophical Problems in Science

Zagadnienia Filozoficzne w Nauce

Philosophical Problems in Science

Zagadnienia Filozoficzne w Nauce

LXXIV ■ 2023

Book reviews

Recenzje

The nature and norms of scientific explanation: Some preliminaries

Abel Peña & Cory Wright

California State University Long Beach,
Department of Philosophy

There are at least two deep and related debates about explanation. Firstly, there is a debate about its nature. What are explanations? How do they appear? What features do they have? It is a traditional metaphysical or descriptive debate, with much the same structure as debates over the nature of knowledge, causation, levels, reduction, or other phenomena of interest to philosophers of science. Hence, just as we can ask what knowledge or causation consists in, so too can we ask the same question about explanation; and we can inquire about the nature of scientific explanation more narrowly or even explanation in particular scientific disciplines. Secondly, there is a debate about the norms of explanation. What distinguishes a better explanation from a lesser one? What are their virtues, what do they aim at, and what are the conditions of their success? It is a traditional axiological or prescriptive debate, and concerns the evaluation of explanatory goodness. Hence, just as one might instead debate how best to reason abductively, or whether conductive and inductive inferences can both be good in the same way, so too do philosophers have serious work to do in articulating the norms of explanation and generating criteria for distinguishing them.

There is reason to believe that these two debates about explanation are ordered by a dependency or priority relation, and so not equivalently deep; for one cannot separate out the good explanations from the bad if he does not know what explanations are in the first

place. Consequently, it seems that a resolution to the second debate over the norms of explanation depends on the availability of a working resolution to the first debate over its nature. For example, there are many terrific discussions of the normative force of model-based idealizations, and, in particular, whether some further corrective step of 'alethic repair' or de-idealization is required. In these discussions, it is widely assumed without argument that models just are explanations. However, some philosophers distinguish between explanations and what they call 'explanatory texts'. An implication of this distinction is that models are instead scientific representations of explanations rather than explanations *per se*, such that discussions of the normative force of model-based idealizations would be discussions of the norms of representation rather than of explanation. For other philosophers, what this reveals is just that the explanation–explanatory-text distinction rests on a mistake. But does it? Determining the answer involves resolving the first debate about the nature of explanation; an analysis of our concepts is called for before we can turn to normative or practical topics.

Some philosophers demur. In a spirited attack on 'verbal' metaphysics, Chwistek (1932) lambasted the belief that "through honest and free discourse it is possible to reach the essence of the concepts hidden behind the words uttered in common language, such as good, love, etc." (Chwistek, 2017, p.2). Chwistek's skeptical attack expressed a kind of deflationary attitude toward conceptual analysis that strikes at the very heart of the method—an attitude which has since wended its way into the literature on scientific explanation. One sees this attitude among those who believe that, at bottom, science aims instead to solve problems (Laudan, 1977; Elliott, 2021; Levenstein and et al., 2024). For instance, Koertge writes, "suppose we could all agree on one or more [conceptions] of explanation. What would we

do with them? I am in general dubious [. . .] about the value of asking 'What is x?' questions [. . .]", and "I suggest that we reverse the order of investigation. We should begin by asking what problems a good theory about scientific explanation might reasonably be expected to solve" (Koertge, 1992, pp.85–86). Many pluralists and pragmatists, like Mantzavinos (2016, p.14) and Kitcher (2023, p.60) respectively, concur; they insist that the traditional project of conceptual analysis is regressive or sterile: there's nothing to say about what all and only the scientific explanations have in common, and they don't yield action-oriented recipes for improvement of our lot.

These lines of thought often prove to be self-defeating. Even if we can retrain our attention onto the nature of problems instead of explanations, the same questions recur *mutatis mutandis*. What is a problem? What would count as a solution? If something is a scientific problem, what features does it have? And if something has those features, is it a scientific problem? The deflationary attitude might be redeployed elsewhere; but we should expect the same point to hold true for any substituend that isn't just analytically brute: scientific practices, explanatory games, methodological constraints, etc. Some might want to turn instead toward normative debates. What makes one problem more interesting, or more tractable, or more useful for making progress? But here, again, the debates are not equifundamental. Whatever the goodness or badness of problems (the utility of practices, games, etc.) consists in, one cannot separate out the good ones from the bad if she does not know what problems are in the first place.

Conceptions of explanation provide explicative answers to questions about what explanations are. For example, according to the epistemic conception of explanation (EC), scientific explanations are complexes of representations of entities or phenomena in the physical world. It takes these representations to aim at increased

knowledge about the entities in the physical world, and it takes the norms of explanation to be the norms of knowledge (Wright and van Eck, 2018, p.998). This conception is often associated—but often too closely—with the groundbreaking work of Hempel (1965) as well as the so-called 'San Diego School' of explanation from Kitcher, Churchland, Perini, Bechtel, Burnston, and others besides. As Scriven (1962, p.224) characterized it, "[a scientific explanation] is a topically unified communication, the content of which imparts understanding of some scientific phenomena". According to the ontic conception (OC), however, explanations are instead complexes of the physical entities so represented, which are located among the other spatiotemporal parts of reality and which do not aim at representational norms of goodness. As Forge (1998, p.76) wrote, "on [OC], an explanation is actually a state of affairs in the world". Following in this vein, Craver (2007, p.27) wrote, '[ontic] explanations are not texts; they are full-bodied things. They are facts, not representations. [. . .] There is no question of [ontic] explanations being right or wrong, or good or bad. They just are'. Likewise, Jenkins (2008, p.64) wrote that explanations conceived ontically are 'worldly things', that is, "not the sort of things that are true or false, but rather the sorts of things that take place or obtain, such as facts or events". This conception is often associated with the work of Salmon (1984), as well as the so-called 'Pittsburgh School' of Salmon, Woodward, Craver, Andersen, and others.

Bokulich (2016, p.263) introduced a helpful distinction between views about what explanations are ('conceptions') versus how explanations work ('accounts'). She puts this distinction to work by correctly noting that Salmon endorsed both the ontic conception and the causal account; and, she might have added, Hempel endorsed both the epistemic conception and the nomological account. She is also

right to note that one can reject the ontic conception while accepting the causal account, just as one can accept the epistemic conception while rejecting the nomological account.

By using and enforcing the distinction between conceptions and accounts, we stand to gain a more sophisticated interpretation about the literature. For instance, Salmon (1984, p.301) distinguished EC and OC from what he called the 'modal conception', according to which "explanations explain by showing that what did happen had to happen, from which it follows that no incompatible alternative could have happened". But the modal 'conception' does not readily specify what explanations are. Indeed, one could accept EC, or could accept OC, and build in these modal commitments about how explanations show off counterfactual necessity. That is, using Bokulich's distinction, one could accept these other conceptions while endorsing the modal account. Similarly, one can accept EC while endorsing the erotetic account, which scientific explanations answer certain kinds of why- and how-questions. To take another instance, Faye (1999; 2007) describes what he calls the 'pragmatic-rhetorical conception' (PC), according to which explanations are informationally relevant answers that are advanced in the problem context of a rhetorical situation whenever the speaker intends to solve the problem. But again, deploying Bokulich's distinction, we can now see that PC is either a conception of explanation, albeit one that's derivative of EC, or else an erotetic account of how these types of representations work in communicative situations, and so not a genuine competitor.

The aim of this special issue of *Philosophical Problems in Science/Zagadnienia Filozoficzne w Nauce* (ZFN) is to survey whether or not a consensus is at hand in these debates and to help settle what it can. The overarching foci are twofold: (i) the nature of scientific explanation, with special attention to the debate between OC and EC,

and (ii) the norms of scientific explanation, with special attention to so-called 'ontic' (or better, 'alethic') norms like truth and referential success and epistemic norms like intelligibility and idealized understanding. It called for advocates of various conceptions to articulate the current state of these debates. Researchers and scholars from around the globe—including Poland, Canada, Korea, The Netherlands, the United States, Greece, Austria, and Belgium—contributed. The special issue also attempts to provide an opening for new work on the norms of explanation, such as truth or model-based accuracy, information compression, abstraction, and generalization.

The first paper in this issue, Panagiotis Karadimas's 'Explanation, representation, and information' argues that EC can encompass almost all scientific explanations by conceptualizing them as relations between representations and thus renders OC gratuitous. He arrives at this conclusion by first demonstrating that abstract explanations do not ultimately make up a distinct category apart from non-abstract ones, and thus EC and OC are not differentially applicable. Karadimas then develops some new objections uniquely faced by OC; these include the dominant role of information transfer (rather than direct observation) in scientific explanation and the fact that the ontic conception doesn't accommodate explanations that involve false propositions. It is concluded, then, that the applicability of the EC is significantly preferable in virtue of its vastly broader scope.

In 'Dimensions of explanation' Eric Hochstein rejects the threefold division of explanation into exclusive communicative, representational, and ontic aspects. Instead, his paper argues that a scientific explanation can always be analyzed along each of these dimensions. After describing his proposal, Hochstein describes how to dispatch some potential problems. The result is a means for resolving disputes involving mechanistic explanations.

The topic of mechanistic explanation also serves as the backdrop of both Jinyeong Gim's 'The ontic-epistemic debates of explanation revisited: The three-dimensional approach' and Michał Oleksowicz's 'Ontic or epistemic conception of explanation: A misleading distinction?'. Both authors concur with Karadimas that mechanistic explanations are more likely to be epistemic than ontic. Both Gim and Oleksowicz begin with a survey of how the debates seem to have changed, moving from a discussion of Hempel and Salmon's works in the last century to the current state of the debates over explanation. Salmon (1989) attributed OC to Coffa. But was Coffa instead reacting to the thought of Scheffler (1963) on inscriptionalism? Did Coffa's anti-Kantian attachment to Bolzano, and his study of Russellian propositional complexes, influence his understanding of OC, and therefore Salmon's? These deeper historical lines have yet to be excavated.

Gim arrives at a three-part classification, comprising a relational dimension of explanatoriness, a conceptual dimension of the nature of explanation, and a normative dimension evaluating the goodness of explanations. What Gim calls 'dimensions' are different than what Hochstein intends, however. His first dimension ('explanatoriness') is analyzed in terms of form, force, and relevance. This might be mapped into Bokulich's discussion of accounts of how explanations work. The second and third dimensions concern the proper conception of the nature of explanation and its normativity, respectively—basically, what we described as the first and second debates at the outset. By exploring each of these in depth, Gim aims to show that mechanistic explanation need not be ontic and can be epistemic in each dimension.

Mark Povich argues in 'A conventionalist account of distinctively mathematical explanation' that this kind of explanation averts a strong objection to the ontic status of other non-conventionalist accounts of distinctively mathematical explanation (DME). This conception

treats the explananda and explanantia of (DMEs), which are mathematical facts, as ontic items and the explanatory relation between these items as likewise ontic. Povich's article begins with an exemplar of a DME, expounds upon recent ontic conceptions of DME, develops a conventionalist account of DME and lastly anticipates some proposed challenges. For example, an open question concerns the required conception of facts. On an inflated compositional or Tractarian conception of facts, DMEs seem genuinely ontic; but conventionalism is a harder sell. On a deflated propositional or Fregean conception, the ontic nature of DMEs may be called into question. By coupling conventionalism and DMEs, Povich aims to show that a path forward for the ontic conception of explanation remains open, if only in the mathematical domain.

As Povich's paper shows, the first debate between conceptions like OC and EC has continued to evolve into different areas. One suggestion is that many discussions have recently moved in the direction of normative analyses of 'ontic and epistemic constraints' that explanations must satisfy in order to count as good scientific explanations. van Eck (2015; 2018; 2021) has argued that appealing to ontic constraints just unwittingly concedes the debate between EC and OC. Michał Oleksowicz's 'Ontic or epistemic conception of explanation: A misleading distinction?' engages with this evolution: he attempts to provide a summary of the first debate between OC and EC that can make sense of transitioning interest to the second debate: what's been called the 'normative turn'. Oleksowicz, following Illari and others, contends that the debate has shifted to a consideration of norms and constraints on scientific explanation that differentially benefit various conceptions and accounts. For the New Mechanists, the idea of mech-

anisms as complex causal systems in the world and of mechanistic explanations as tools for discovering those complex systems are both critical to an understanding of natural phenomena.

Understanding is regulative norm emphasized by many advocates of EC, and resurfaces in Federica Malfatti's review of McCain's *Understanding How Science Explains the World*. Malfatti teases out two competing views of scientific explanation discussed in the book. The first view suggests that explanation mirrors the facts, depicting dependency relations actually holding in the real world; the other view, however, ties explanation to the contingencies of evidential support and evidential standards. Malfatti wrestles with a possible reconciliation of these positions, and emphasizes the role that scientific realism can play in making such cases clearer and stronger.

Many advocates of both EC and OC are friendly to realism, but are then pressed to say something about idealizations. In 'Can fiction and veritism go hand in hand?' Antoine Brandelet considers this familiar tension in the use of models as scientific explanations. On one hand, the thesis that truth is a necessary condition of explanation (veritism) belies acceptance of model-based explanations as integral to explanatory knowledge. On the other hand, models can be highly idealized—even to the point of gross simplification and distortion. Brandelet co-opts a fictionalist strategy from responses to the representation problem in modeling to mount a defense of veritism. He argues that this fictional approach ultimately helps to clarify the distinction between OC and EC, asserting that the former does not deny the importance of such fictional processes as idealizations but, rather, emphasizes the referents of explanatory texts and representations. The debate between the two conceptions, then, is over the additional question of whether or not those referents just are explanations.

Relatedly, Kristian Campbell González Barman's review of *Models and Idealizations in Science: Artifactual and Fictional Approaches*, edited by Cassini & Redmond, provides brief overviews of the chapters, including those on the topic of de-idealization by Carrillo & Knuuttila, Cassini, and others, and on fictionalism by Frigg & Nguyen, Salis, García-Carpintero, and others. However, González Barman takes the additional step of relating several of these to ontic and epistemic concerns. Barman particularly highlights some of the strategies and problems involved with respect to idealizations and de-idealizations in scientific modeling, as well as various fictional interpretations, to draw pertinent lessons for proponents on both sides of the debate.

In 'Mechanisms 'all the way down'?', Ioan Muntean reviews of *Mechanisms in Physics and Beyond*, edited by Falkenburg & Schiemann. The New Mechanists have primarily confined their views to the life sciences, and the various chapters in this book are among the first attempts to apply this doctrine to the lower-level sciences. Muntean critically analyzes arguments adduced for and against implementing mechanistic explanations in physics, where nomological, mathematical, and non-causal explanations play a much more common role. The review is rounded out by an analysis of contributions dealing with an explanatory framework of levels, the cognitive process of interpreting mechanisms, and the relation of physical systems to computations.

Overall, the papers in this special issue jointly demonstrate the fruitfulness of these debates, and they also lay some additional groundwork for developing the theoretical issues even further. It is important to acknowledge the many referees that generously lent their expertise to provide feedback and recommendations. Finally, thanks to Ning Shao; special appreciation goes to the journal's editors and staff for providing guidance and assistance, and for their dedication to intellectual discussion in philosophy of science.

Bibliography

Bokulich, A., 2016. Fiction as a vehicle for truth: Moving beyond the ontic conception. *The Monist*, 99(3), pp.260–279. https://doi.org/10.1093/monist/onw004.

Chwistek, L., 1932. Tragedia werbalnej metafizyki (Z powodu książki Dra Ingardena: Das literarische Kunstwerk). *Kwartalnik Filozoficzny*, 10, pp.46–76.

Chwistek, L., 2017. The tragedy of verbal metaphysics: On Dr. Ingarden's Das literarische Kunstwerk (A. Trybus, Trans.). *Journal for the History of Analytical Philosophy*, 5(1), pp.1–20. https://doi.org/10.15173/jhap.v5i1.2957.

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

van Eck, D., 2015. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Axiomathes*, 25, pp.5–22. https://doi.org/10.1007/s10516-014-9243-x.

van Eck, D. and Wright, C., 2021. Mechanist idealisation in systems biology. *Synthese*, 199(1), pp.1555–1575. https://doi.org/10.1007/s11229-020-02816-8.

Elliott, S., 2021. Research problems. *The British Journal for the Philosophy of Science*, 72(4), pp.1013–1037. https://doi.org/10.1093/bjps/axz052.

Faye, J., 1999. Explanation explained. *Synthese*, 120(1), pp.61–75. https://doi.org/10.1023/A:1005258504182.

Faye, J., 2007. The Pragmatic-Rhetorical Theory of Explanation. In: J. Persson and Petri Ylikoski, eds. *Rethinking Explanation*, *Boston Studies in the Philosophy of Science*, 252. Dordrecht: Springer Netherlands, pp.43–68. https://doi.org/10.1007/978-1-4020-5581-2_4.

Forge, J., 1998. Explanation and mechanism: reflections on the ontic conception of explanation. In: D. Anapolitanos, A. Baltas and S. Tsinorema, eds. *Philosophy and the Many Faces of Science*, *CPS publications in philosophy of science*. Lanham, Md: Rowman & Littlefield, pp.76–92.

Hempel, C.G., 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Jenkins, C.S., 2008. Romeo, René, and the reasons why: What explanation is. *Proceedings of the Aristotelian Society*, 108(1), pp.61–84. https://doi.org/10.1111/j.1467-9264.2008.00236.x.

Kitcher, P., 2023. *What's the Use of Philosophy?* Oxford, New York: Oxford University Press.

Koertge, N., 1992. Explanation and its problems. *The British Journal for the Philosophy of Science*, 43(1), pp.85–98. https://doi.org/10.1093/bjps/43.1.85.

Laudan, L., 1977. *Progress and Its Problems: Toward a Theory of Scientific Growth*. Berkeley; Los Angeles; London: University of California Press.

Levenstein, D. and et al., 2024. The problem-ladenness of theory [in press]. *Computation Brain and Behavior*.

Mantzavinos, C., 2016. *Explanatory Pluralism*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316440599.

Salmon, W.C., 1984. Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, (2), pp.293–305. https://doi.org/10.1086/psaprocbienmeetp.1984.2.192510.

Salmon, W.C., 1989. Four decades of scientific explanation. In: P. Kitcher and W.C. Salmon, eds. *Scientific Explanation*, *Minnesota studies in the philosophy of science*, 13. Minneapolis: University of Minnesota Press, pp.3–219. Available at: <http://conservancy.umn.edu/handle/11299/185700> [visited on 4 October 2023].

Scheffler, I., 1963. *The Anatomy of Inquiry: Philosophical Studies in the Theory of Science*, *Borzoi Books in the Philosophy of Science*. New York: Alfred. A. Knopf.

Scriven, M., 1962. Explanations, predictions, and laws. In: H. Feigl and G. Maxwell, eds. *Scientific explanation, space, and time. Preliminaries*, *Minnesota Studies in the Philosophy of Science*, 3. Minneapolis: University of Minnesota Press, pp.170–230. Available at: <http://conservancy.umn.edu/handle/11299/184631> [visited on 8 February 2024].

Wright, C. and van Eck, D., 2018. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Ergo, an Open Access Journal of Philosophy*, 5(38). https://doi.org/10.3998/ergo.12405314.0005.038.

Articles

Artykuły

Explanation, representation and information

Panagiotis Karadimas

National and Kapodistrian University of Athens;
Hellenic Air Force Academy

Abstract

The ontic conception of explanation is predicated on the proposition that "explanation is a relation between real objects in the world" and hence, according to this approach, scientific explanation cannot take place absent such a premise. Despite the fact that critics have emphasized several drawbacks of the ontic conception, as for example its inability to address the so-called "abstract explanations", the debate is not settled and the ontic view can claim to capture cases of explanation that are non-abstract, such as causal relations between events. However, by eliminating the distinction between abstract and non-abstract explanations, it follows that ontic and epistemic proposals can no longer contend to capture different cases of explanation and either all are captured by the ontic view or all are captured by the epistemic view. On closer inspection, it turns out that the ontic view deals with events that fall outside the scientists' scope of observation and that it does not accommodate common instances of explanation such as explanations from false propositions and hence it cannot establish itself as the dominant philosophical stance with respect to explanation. On the contrary, the epistemic conception does account for almost all episodes of explanation and can be described as a relation between representations, whereby the explanans transmit information to the explanandum and that this information can come, dependent on context,

in the form of any of the available theories of explanation (law-like, unificatory, causal and non-causal). The range of application of the ontic view thus is severely restricted to trivial cases of explanation that come through direct observation of the events involved in an explanation and explanation is to be mostly conceived epistemically.

1. Introduction

The main ontic thesis is that explanations are not representations (texts, diagrams et.al.) and that they are in fact relations between real objects in the world (Glennan, 2005; Craver, 2007). Hence what scientific explanation amounts to, the argument goes, is causally relating the objects in question and showcasing how they can exhibit the causal patterns of the world (Salmon, 1984; 1989; 1998). Even though this philosophical stance is not always clearly formulated (Wright, 2015) and furthermore it has been shown that much of the causal-mechanistic explanations that supposedly vindicate the ontic view, can be reconstructed without adhering to such an ontic conception of explanation (Wright, 2012), a great deal of the current literature does rely on the main tenets of the ontic conception. And although the relevant arguments may differ slightly as for the lengths to which each account is willing to go when calling for ontological commitments, all of them eventually posit that explanation is a relation between real objects in the world (Machamer, Darden and Craver, 2000; Craver,

2005; Craver and Bechtel, 2006; Winning, 2020). Wright and van Eck have offered the strongest—at least to the best of my knowledge—objection to the ontic conception. Their argument in part entails the view that the ontic proposal does not account for "generalized explanations", namely explanations that are given by representations which abstract away from details of the world and that such instances of explanation can be conceived only epistemically (Wright and van Eck, 2018, p.1019).

Abstracting away from details of the real world that are considered unimportant to an explanation, is an issue that has attracted the attention of scholars and, one way or another, it amounts to the assumption that scientists often intentionally omit irrelevant details of the world when pursuing explanations. While Wright and van Eck are right to mention that such explanations as the so-called "abstract" ones cast doubt on the ontic approach to explanation, this philosophical stance seems to allow for some reconciliation of the ontic and the epistemic approaches to explanation. That is, it may be argued that "abstract" explanations and ontic explanations explain in different circumstances or that the epistemic view is mostly normative while the ontic is the one that in fact gives explanations, and so, they are not mutually exclusive philosophical and epistemic points of view. Sheredos for example seems to argue along these lines (Sheredos, 2016; 2019).

However, it does seem to me that instances of explanation that are now considered to be "abstract" are not in fact a special case of explanation, or a distinct category thereof, but rather constitute part of the standard way scientific explanations are offered. To appreciate this, we first need to re-conceptualize "abstract explanations" and to consider them no more that way, but rather, as one of the ways "optimization process" occurs. During optimization, scientists rule out

parameters that convey information that can render an explanation irrelevant. These can be trivial details of the world or even theoretical postulates that are irrelevant in context. Optimization processes are, as Strevens has long argued, a central ingredient in scientific explanations and given the contextual nature of why-questions, no explanation can be given without an optimization process for if that happens then the information that is transmitted can be irrelevant.

In view of such a novel description of "abstract explanations" as mere optimization processes, we can go one step further and consider them in relation to scientific representations. There seems to be a central feature of scientific representations that is not part of the optimization process, even though it may lead to informationally equivalent outcomes. That is, representations often miss aspects of the observable portion of the world and so they are constructed as giving an empirical picture of it but *not* a real one. This differs from optimization processes (and hence from abstract explanations) for it is not that scientists choose to exclude the irrelevant elements from the discussion (as it happens during optimization), it is that a set of observable events of the representation's target system do not appear in the representation for reasons not directly linked to the explanation, but which are mainly related to the very nature of scientific representations, namely that they, most of the times, represent aspects of the system they target to and not every element of it. Despite the fact that the absence of observable events is documented, these representations are nevertheless used as explanations. It is therefore necessary to be mindful of this subtlety, for if we do not do so, then explanations that come from representations that miss elements of the observable world can be described as "abstractions" even though they are inherently constructed that way in the first place, regardless of whether they are applied to explanation or not. Of course, when a set of representations

appears as a possible answer to a why-question, then it may undergo an optimization process so that its most relevant parts will be used; but the ones that are eventually used are not necessarily the more "abstract" ones, even though when the more abstract features are used this is an optimization process.

If the above are on the right track, the ontic approach to explanation is further undercut and some epistemological concerns can be expressed on whether explanation is to appeal to realist claims about the world. Philosophers have introduced the notion of "denotation" which we can use to contemplate over the distinction between representations of real world entities and representations of empirical states of affairs ("real and empirical representations" onwards), or even of hypothetical constructions that are used and represented in science. In each case, the denoting symbols of the representational schemata refer to their denotata (Russell, 1956; Elgin, 2010; Salis and Frigg, 2020) and there are examples where scientific explanation occurs but neither the explanans nor the explanandum can claim to hold the status of being propositions that represent real objects in the world for the denoting symbols do not correspond to all parts of the target system, but only to a portion of it. They can both therefore be credibly considered as empirical representations of the world, namely as representing parts of the observable world, albeit without being committed to bearing a one-to-one representation of the mind-independent reality they target to. Even in case a formal proof showing that a representation X of an event or phenomenon Y is so accurate and precise that renders Y "real", then, again, we need not commit to the reality of such an entity insofar as we do not observe it directly. But since most of the times scientists observe a representational schema that denotes the existence of its target system and they do not directly observe the system in question, it seems that they work with representations and

not with real world entities. If that is so, it seems that what matters when it comes to explanation is the *information* transmitted by the explanatory relation to the explanndum and not whether such explanations are real and not even if the respective representations represent real entities. To illustrate this, one can examine how death-rates are explained in medicine. Both the fatality rate of a disease, which can play the role of the explanans in these contexts, as well as the reasons a patient or a group of individuals died, which can be the explanandum, are in fact empirical representations and not real ones. They are both documented to miss observable parts of the world, though they are still used for explanatory purposes due to the information each representation carries.

Another problem the ontic view faces is that a great many scientific activities contain representations of false propositions. Thought experiments for example have been characterized as mingled representations i.e. representations of empirical and hypothetical conditions that explain in context and the same is the case with similar practices such as models and computer simulations (Karadimas, 2022). In these cases of scientific representation, the denoting symbols have no real-world denotata (at least as for the hypothetical part of the representation) and if the propositions induced by these practices are able to offer explanations, then ontological claims about explanation seem to be severely undermined for it does appear that explanation is achieved without true propositions being involved, let alone relating real objects in the world. Moreover, false propositions are not limited to mingled representations but expand to theories now considered false, but which were true in the past and which, in spite of being considered false nowadays, they can still give explanations in some

contexts. Aristotle's theory for example explains better than Galileo's one the speed of falling objects in terrestrial conditions, for the latter applies only to contexts where vacuum is created (Rovelli, 2015).

In order to capture such cases philosophically—which is paramount for they are central episodes in the sciences—one should focus on the epistemic side of explanation and not so much on the ontic one. If the distinction between abstract and non abstract explanations concerns us no more and all we have is representations that when applied to explanation may undergo optimization and if scientists observe representations that either miss portions of the world or induce false propositions, we can arrive at the conclusion that explanation is a relation between representations and not a relation between real objects in the world or even between real representations. There appear to be three main types of scientific representation: empirical, hypothetical and mingled. When the explanans are represented, the representations in question can be purely empirical (which include theories considered true as well as theories considered false) or they can be mingled. While one cannot exclude formally the possibility of being merely hypothetical ones, most of the times hypothetical constructions include elements of the empirical world and so I take mingled representations to capture this case as well. When the explanandum is represented the representations are mostly empirical (for science typically does not engage in explanations of counterfactual or non-existent states of affairs). A representation that attempts to answer a why-question carries information that needs to square well with the information elicited by the propositions that represent the explanandum. Moreover, as it will be shown below, two levels of optimization can take place during explanation: crude optimization that rules out representations that are in principle relevant but become irrelevant under a certain context, and sophisticated optimization that puts for-

ward the most relevant propositions in each case. More episodes of
scientific explanation are captured that way than via ontic approaches
and, it turns out, there is little to no room for the ontic approach to
present itself as a strong philosophical viewpoint.

I begin by showing that abstract explanations are not a distinct
category of explanation and that abstraction is only *one* of the pos-
sible optimization processes that may take place. This suggests that
there are no longer abstract and non-abstract explanations and so the
epistemic and the ontic view both are vying to capture all of them
or else they run the risk of missing all of them. I then present two
key problems the ontic view faces which the epistemic view does not:
first, the fact that scientific explanation comes mostly through repre-
sentations and not through direct observation and so the ontic view
ends up making claims that is not in a position to make for scientists
most of the times observe representations that represent aspects of the
world, not the world itself. The second hurdle is that the ontic view
does not account for explanations that come from false propositions,
either these are mingled propositions or empirical propositions of the-
ories now considered false but are nevertheless explanatory. Having
highlighted these drawbacks, I present how the epistemic approach
can claim to be the soundest philosophical stance on explanation by
unpacking how it manages to capture nearly all instances of scientific
explanation by conceptualizing them as "relations between represen-
tations". The conclusion is that the ontic view is rendered redundant
and that the best it can do is to capture trivial cases of explanation that
come through direct observation.

2. Abstraction, representation and optimization processes

Strevens has developed the notion of "optimization process" to describe the necessary procedure that needs to take place in explanations. As he puts it, a causal explanation occurs only when the factors that make a difference with respect to the explanandum are taken into account. These elements are the result of an optimization process that excludes factors that could make the explanation irrelevant (Strevens, 2011). Strevens is primarily preoccupied with causal explanations but as we will see it appears that his theory on optimization process finds applications to all sorts of explanations i.e. both causal and non-causal.[1]

As a part of non-causal explanations are often considered the so-called abstract explanations. Pinckock discusses examples of such a kind of explanations, such as the Konigsberg bridges problem, and by comparing abstract explanations to "microphysical explanations" concludes that the former often lead to better explanations of that explanandum (Pincock, 2007). In Pinkcock's analysis therefore "abstraction" amounts to assuming away petty or confounding events of the physical world while involving in it theoretical premises and simplified assumptions about the structure of reality. Lange examines the same problem and makes the case that the attempts to explain the Konigsberg bridges problem by appealing to nomological explanations will fail because covering laws are irrelevant in this context.

---

[1] Ironically enough, Strevens is often considered as an advocate of the ontic view of explanation due to his commitment to causal explanations. Here I use Strevens's terminology to argue for the opposite, i.e. that optimization is not restricted to causal explanations, but is a feature of explanation by and large, and that it helps us vindicate the epistemic view of explanation.

He puts forward a mathematical explanation by appealing to the notion of necessity, as well as to some context-sensitive facts, such as contingent facts that co-determine the explanation (Lange, 2013).

While both these cases are considered in the literature as abstract explanations, they can be both described simply as optimization processes, in the Strevens's sense without loss of explanatoriness. As it turned out, different parts are omitted in each case: events of the physical world in Pinckock's discussion and laws of the physics in Lange's exploration of the issue. This indicates that in pursuing the optimal explanation, the focus is on the factors that will establish the relevance of the explanation and not the abstraction from events of the microphysical world per se. Hence while omitting events of the micro-world can be part of an optimization process, eliminating laws and theoretical postulates can also be. There seems to be no good reason to consider them as different classes of explanation that find different philosophical conceptualization, for in both cases it is an optimization process that takes place whereby scholars try to figure out the premises that will help them achieve explanation and to simultaneously minimize the impact of variables that could obscure the relevance of the explanation, even if such explanations are not causal ones.

One could reply that even if in Lange's approach it is laws that are excluded as irrelevant, the explanation comes from another type of "abstract" explanation i.e. mathematical explanation and not from a non-abstract one, such as from a relation between events. However, such an objection does not take into account that it is not mathematical expressions on their own that explain, but rather the mathematical expressions *alongside* context-related facts—"contingent conditions" as Lange puts it (Lange, 2013, p.506)—and hence it becomes unclear whether such an explanation is abstract for it does not assume away

events of the micro-world, but instead it takes *the most relevant of them* into account in order to explain. Since no clear threshold for abstraction is on offer (Jansson and Saatsi, 2019), and it strikes me as if it can barely ever be one, then discriminating an abstract form a non-abstract explanation can be a matter of confusion. Even in case the explanation was given only by some "abstract" mathematical structures this would have not ratified abstract explanations as inherently different from the others for, as the examples drawn from Pinckock and from Lange show, it seems that, either the explanation is highly abstract or less-than-highly-abstract in both cases it eventually boils down to being involved in an optimization process that tries to find relevant answers to why-questions and not to an abstraction in its own right.[2]

Such representational constructs (either these are highly abstract or less than that) that are used to attain explanations appear not only

---

[2] This seems to speak against the notion of "idealization" as well, which is similar to abstraction and even though they are considered not exactly identical notions, when it comes to explanation they come up with the same suggestion and so the rejection of one notion refutes the other at once. That is, some philosophers claim that abstraction is when one intentionally omits unimportant details of the world, albeit without intentionally distorting aspects of the target system that is represented, while idealization is when the target system is intentionally distorted for the sake of simplicity or clarity (Godfrey-Smith, 2009; Levy, 2021). Speaking of their application to explanation, both notions suggest that explanation often needs to either abstract away or to distort the target system of interest in order to achieve explanation (Love and Nathan, 2015; Potochnik, 2017). However, as the examples discussed here suggest, representations miss aspects of their target systems regardless of their application to explanation and so when they are used to achieve explanations, both abstraction-like representations and idealization-like ones indicate the same epistemic results: whether they put forward content that assumes away parts of the observable world (as per abstraction) or whether the content they carry distorts the target system (as idealization demands), in both cases this is no more than an optimization process that pursues the most relevant piece of information in context. Since it is the notion of "abstraction" and not that of "idealization" the one that is more central to the ontic/epistemic debate, I will no longer consider idealization here.

to be both involved in optimization processes during explanations, but they moreover seem to share some common features which they carry *regardless* of their application to explanation, namely that they are imprecise representations of the parts of the world they represent. Scholars have paid close attention to the well-established fact that scientific representations are not perfect portrayals of their target systems (Frigg and Nguyen, 2017) and that the missing parts are often events in the observable world (Batterman, 2007). Scientists therefore may be fully aware that this is so and nevertheless accept it as a credible representation. Consider for example the findings from serological studies i.e. from studies that attempt to measure the antibody prevalence in a population and based on this to estimate the lethality of a disease. The theory that guides such measurements is that the levels of immune responses against a particular disease that exist in a population largely determine the disease's lethality. The levels of immunity are divided with the recorded deaths from the disease in question. The number of deaths is the numerator and the number of people with estimated immune responses is the denominator; the result of this division produces the infection-fatality-ratio (IFR) estimate. Immune responses are induced by different kinds of antibodies and T-Cells and they are tracked in the blood of randomly selected individuals. If immune responses are found to a large number of people in comparison to the number of deaths, then the IFR is low which suggests that the disease is widespread but that only a small fraction of the infected people has died which in turn indicates that the disease is mostly harmless. Conversely, if the levels of immunity are low in comparison to the number of deaths, then the IFR is high which indicates that the disease could pose risk to a larger segment of the population. IFRs for germs such as Sars-Cov-2 (which is said to be the virus that causes the coronavirus disease Covid-19), are constructed

by measuring the antibody levels in a population. During the Covid-19 pandemic therefore, scientists made use of some tests which were used to identify antibody-related immune responses. However, it is possible that the current antibody studies underestimate the immune responses in the population and thus overestimate the lethality of a disease. Serological studies do not account for the T-Cell responses that are either pre-existing or are elicited after mild or asymptomatic Covid-19 and they also are structured so that they detect only IgB and IgM antibodies. They thus do not detect IgA antibodies that are also important in fighting pathogens and are also produced during infection. Those who tackle the disease through T-Cells solely or through IgA antibodies, may not develop virus-specific IgG antibodies and so the prevalence of the disease may considered to be lower than it actually is. Moreover, even if IgG antibodies are secreted, they appear to decline rapidly and so late testing may miss some cases of these antibody responses too (Burgess, Ponsford and Gill, 2020). In spite of these downsides, scientists consider serological studies to give us a quite reliable estimate of how lethal a disease is, for even if they do not project with 100% precision the lethality by somehow underestimating the levels of immunity and slightly overestimating the infection-fatality ratio (IFR) they appear to capture a notable part of the target system, namely much of the antibody-related immune responses. Therefore the IFR of a disease (of Covid-19 in this case) is—irrespective of its possible application to explanation—a *representation* of the infected-to-dead individuals and, given the glitches that appear in identifying all immune responses, it cannot be said that via this estimate we have a precise representation of the prevalence of the disease and hence we do not have fullaccess to the world, even though it is stated that we do have a very good picture of how deadly a disease is.

So it seems that representations are in no need of a particular type of abstraction when they are used to explain, for they omit parts of the observable world they target to *in the first place* and this happens *both* when they represent some theoretical or mathematical postulates *and* when they represent events of the world and the relations between them. Hence what is at stake when they are applied to explanation is to optimize them, namely to select the *information* these representations carry that is suitable in each case.

While the discussion so far takes into account how abstraction, optimization and representation can be considered when the explanans are taken into account, but similar is the case with the explanandum. Bokulich has proposed the "eikonic view" of explanation in which the explanandum is in fact a representation of its target system and different representations of it are given in different contexts (Bokulich, 2018). While I am in agreement with much of her analysis, the way she employs the notion of "abstraction" in it by claiming that the representation of the explanandum entails a level of abstraction (Bokulich, 2018, p.803), is a point that I am taking issue with. Since explanation is inherently contextual, the representations of the explanandum will be shaped so that they will be on a par with the purposes of the explanation as it occurs in context. Despite the fact that the classic Adam/apple example shows that not only the same representational strategies, but moreover precisely the same lexis can be used to highlight different explananda in different contexts, in more complex cases of explanation it can be assumed that different representations will be needed as well. Whichever representations are used though, whether the explanandum is to be represented with a certain degree of abstraction this is a matter of context and thus, again, the aspects of it that are to be represented are mainly an optimization process in which only the suitable parts of the system are included. As in the

case of the explanans, the omitted factors can be theoretical postulates or unimportant (in context) events. For example, one can ask "why countries go bankrupt?". That question is not independent of context (hence its answer is not an abstract explanation), even though it may seem so at a glance. The explanandum in this context requires information related to the central causes of a country-level bankruptcy and so, it can be described as an optimization process not as an abstraction: it represents bankrupt countries via text. Of course, through textual representation the explanndum can be altered and represented in another context, thus seeking different information as an answer. In such a case one could ask "Why did country *X* go bankrupt?" In both examples, the target system is the historically recorded fact that countries may go bankrupt from time to time and this target system is represented in different explanatory contexts after it has undergone an optimization process so that different aspects of it are included in each case. In the former why-question the explanandum may be conceived as omitting trivial or confounding events of the world, such as the economic conditions in some countries, though in the latter it does not and is especially interested in country *X*'s bankruptcy which it turn leaves open the possibility of omitting some theoretical postulates employed in the representation of the country-level bankruptcy.

Reducing abstraction to sheer optimization will turn out to be the key in the discussion on explanation and on the ontic-epistemic debate. As stated, it implies that the distinction between abstract and non abstract explanations wanes which in turn suggests that all explanations find the same philosophical conceptualization; hence either all are captured by the ontic approach or all are captured by the epistemic one. To establish that the latter is the case, we can go on to see some inadequacies the ontic view faces which will reveal not only the drawbacks of this philosophical stance but will also further solidify

the epistemic approach to explanation. The first problem is related to what scientists observe when they try to offer explanations and, as I show in the next section, they typically observe representations of the world, not the world per se.

3. Scientists' Scope of Observation

While this is not the place to delve into the realism-antirealism debate, it is important to reflect on what scientists observe when scientific explanations take place. First off, as it is already mentioned here and as it is widely acknowledged, it appears that scientists work with scientific representations of the world, at least in non-trivial cases of explanation. Indeed, when economists explain reduced longevity and put it down to poor economic performance in the countries that reduced life span was recorded, they work with figures representing years of life lost and economic outcomes (such as GDP fall), they do not observe directly these countries and the economic activity of their citizens and their governments. It does follow therefore that what scientists observe when attempting an explanation is a range of representational schemata, not the "mind-independent" world and even if it goes without saying that scientific representations represent aspects of the world, it is also hardly questionable that scientists still have direct access to the representations, and only implicit access to the world via these representations (van Fraassen, 2008, p.254). This is important, for even if one backs down from the epistemic view of explanation and endorses the ontic view and its central claim that explanations are relations between real objects in the world, then it is unclear how scientists can have knowledge of these relations if they reject the representational (and hence the epistemic) view of explana-

tion. In other words, if diagrams, texts and other representations are not explanations and explanations occur in the world irrespective of the representational schemata, then those who pursue the ontic view need to show how scientists have access to these explanations. Unless this is shown, it could be argued that only scientific representations are within the scientists' scope of observation, not the real world.

Of course, I do not wish to eliminate formally the possibility that observation of the world takes place and simultaneously non-trivial explanations are given (as for example when planets are observed), but in most cases a representation of the world is first constructed and it is this representation and the information it carries that is then used for explanatory purposes.[3] When measurements in experiments take place for example, the experimental set up offers a pile of data some of which are included in a data-model (Giere, 2018) and then, with the aid of some theoretical postulates, the elements the data model entails are interpreted and eventually the measurement outcome is represented in a model that merges theoretical postulates and elements of the data model (Parker, 2017). If such a representation is to be used to explain, scientists have access solely to this schema.

To further appreciate this, consider moreover an explanation of covid-induced deaths. The IFR of Sars-Cov-2 is, of course, often used to explain death rates in a certain region or even worldwide in spite of the fact that it does not capture all immune responses of the human immune system. As for the explanandum in such a case of explanation i.e. deaths in medicine, we need to be mindful that they barely occur as a result of one factor and several confluent factors usually co-determine the outcome. This happens with nearly all causes of death

---

[3] However, even in such cases, the jury is still out on whether observation is direct or whether it is implicit via signals that appear on the telescope, as it happened with the observation of solar neutrinos (Shapere, 1982).

and it happened with deaths attributed to Sars-Cov-2 as well. That is, an elderly person with pneumonia may have been also diagnosed with blood clots and heart inflammation. Either for the sake of convenience or due to some established consensus, among the several contributors typically one is pointed out as a central cause of death which is not necessarily identical to the actual cause of death. Speaking of Covid-19, the criterion to determine cause of death was a positive PCR test. If one dies with a positive PCR test then is recorded as a covid death regardless of possible co-morbidities. However, this test can find dead viral fragments and it is known that dead viral fragments do not cause illness, let alone death. But the death certificates mention "Covid-19" as a cause of death even if dead viral fragments are indentified to the individual. Hence it is perfectly possible that a number of people who were reported as dying from Covid-19 were dead due to some other causes (Jefferson et al., 2020) thus inflating the actual number of deaths. Again, in this case, the representation misses aspects of its target system and in this respect both the death-certificates and the IFR are in accordance with how scientific representations mostly represent their target systems for both the IFR and the covid deaths can be characterized as empirical representations that represent aspects of the world they aim at, albeit without confronting themselves with the task of representing the "real world". It seems therefore that if one uses the IFR to explain death rates, in fact does not observe real conditions or relations between real objects but two representations and explores not the reality of the conditions each one of them describes, but whether the information induced by the explanans fits the information induced by the explanandum.

As it turns out therefore, the ontic view of explanation is seriously challenged when one considers that scientists observe representations and not the mind-independent world, and that the best part of these

scientific representations are empirical and not real. However, things can get even worse for the ontic view when other cases of explanation are taken into account in which the representations used are either theories that entail propositions that are now considered false but they nevertheless explain, or a schema in which empirical and hypothetical (namely, non-existent) state of affairs are represented, as are mingled representations, which are also, by any reasonable conception of truth, false.

4. False Propositions, Scientific Information and the Ontic View of Explanation

The possibility of false theories that are nevertheless explanatory was mentioned above and Aristotle's theory is not the only example of this sort. Newton's theory does explain planetary motion even though it is widely acknowledged to be a false theory. This is a very serious hurdle that advocates of the ontic view need to overcome, for it appears to refute the central claim of the ontic view, namely that explanations are causal relations between real objects in the world.[4] If false theories offer explanations, then it is highly questionable that explanations can be described as exhibiting explanatory relations between real objects in the world for if the objects described by the false theories do not count as real anymore then, it follows from the ontic approach itself that they do not count as explanations either. One attempt to save the ontic

---

[4] With the exception of Glass who recognizes that false propositions can explain, I am so far unaware of any attempt from proponents of the ontic view (or of similar realist-leaning accounts) to even mention, let alone to address, the issue. Glass mentions that false theories explain and does not tackle the problem either, though he does confess that realist explanatory schemata such as "inference to the best explanation" apply only when true theories are candidates for an explanation (Glass, 2021).

proposal could be to try to reconcile it with the pragmatic approach and claim that ontological commitments can be context-dependent and so objects that count as true *in context* are explanatorily related thereof. According to such an attempt, false theories are considered as representing real world entities in certain contexts and not independent of them.[5] While this could be a promising step, it becomes redundant since, when it is considered with respect to explanation, it turns out to be a project not dissimilar to the one described above, namely it relates empirical representations (the explanans and the explanandum) and the why-question is answered either with or without ontological commitments (even if they are described as context-dependent). That is, even if one commits to the reality of the IFR or at least to the aspects it represents, the explanation of the death-rates becomes no more robust than it already is if one uses the IFR as an empirical representation that carries information that explains the death rates, for in both cases it is the information that provides the explanation. The reconciliation of the ontic and the pragmatic view therefore seems to be a strategy which turns out to give results that are, at best, explanatorily equivalent to the ones that an epistemic approach could give.

This problem is exacerbated and the attempt to bring together the ontic and the pragmatic approach flounders when another class of false, but explanatory, propositions is taken into account which are the ones induced by prominent activities such as thought experiments. As I argue elsewhere, thought experiments (and likewise much of the

---

[5] Rovelli (2015) makes an interesting claim that highlights strong pragmatic features not only with respect to explanation (which is not his primary concern) but with respect to science in general. Most importantly, he does not restrict his analysis to false theories, but includes also those considered true and he argues in particular that Aristotle can be found right and wrong in the same way Einstein can be found right and wrong i.e. dependent on contextual factors.

modeling strategies such as computer simulations), can be described as mingled representations that carry information which explains events under several contexts (Karadimas, 2022). The explanatoriness of the mingled representations as well as of the representations that false theories give us is established by the content they carry which is *scientific information*. Scientific information is in principle explanatory relevant (van Fraassen, 1980; Richardson, 1995) and false propositions which carry scientific information are also in principle explanatory and while they may become irrelevant under some contexts, they cannot be excluded from explanation simply because they are not real world objects and they do not even represent relations between such objects. Moreover, one cannot discriminate them as being abstract non-causal explanations that capture different cases of explanation from singular (aka causal) ones. While mingled representations are non-causal explanations, they are not "abstract" explanations as opposed to singular (i.e. causal) ones for, they both share the standard features scientific representations have; mingled propositions denote in part a hypothetical and in part an empirical state of affairs, while empirical propositions denote empirical aspects of the world. When mingled representations are used as explanations, they convey information that is at once hypothetical and empirical and when empirical representations are used they transmit information that describes a causal explanatory relation between the target systems. Both are on a par with the epistemic approach and at odds with the ontic view.

We have eliminated the distinction between abstract and non-abstract explanations and we have moreover underscored two central inadequacies the ontic view faces which in turn results in it being unable to hold the status of being the philosophical stance that captures

all episodes of explanation. On the contrary, the epistemic view fares much better for it manages to encompass all sorts of explanation by conceptualizing them as relations between representations.

5. A Relation between Representations

Different explanatory relations in different contexts are, indeed, a possible outcome and the long debate in philosophy of science over the relevance relation is still ongoing. The standard theories are the law-like which was introduced by Hempel and Oppenheim (1948), the unification, by Friedman (1974) and Kitcher (1981) and causal explanation which has been shaped in diverse forms, such as causal/mechanistic (Salmon, 1984; Machamer, Darden and Craver, 2000), or counterfactual causality (Woodward, 2002), though whichever strand one takes into account, the gist is that explanation needs to describe the causal structure of the world. There are also, as already discussed, non-causal theories of explanation, which focus on how mathematical explanations can be given, as proposed by Lange (2013), or how the mingled propositions can explain under several contexts. All theories of explanation can be captured conceptually by appealing to the epistemic view and to its representational side.

Two main classes of representations can be used to that end; empirical representations and mingled representations. Explanatory relations bear information to the explanandum via one of these categories of representations. The information is scientific and can be law-like, unificatory, causal or non-causal. In any case, it is transmitted through a representation and it targets another representation, namely the explanandum. Each representation can be conceived as a set of propositions that carry scientific information. The denoting

symbols of the explanans is related with the denoting symbols of the explanandum.[6] The relation is not ambivalent and thus the propositions of the explanans target the ones of the explanandum and not the converse. It moreover appears that there are two stages of optimization in explanations; the first is when among several sets of scientific propositions only those that are in principle relevant are considered and then the second stage whereby among the ones in principle relevant, scholars determine the explanatory relevant ones. Call the former "crude optimization" and the latter "sophisticated optimization". Crude optimization rules out non scientific propositions and scientific theories that are irrelevant to the context, such as, for instance, physical theories from an explanation in economics and brings to the fore several sets of explanations that are in principle relevant.[7] Sophisticated optimization amounts not only to figuring out the most relevant set of those that passed the crude optimization, but also determines which relation among the propositions in the set in question can be used as an answer to the why-question. Maybe they can be all of them or some maybe required to be ruled out. The ones that are eliminated in the optimization process can be either laws of high generality, as it can happen when a causal relation is established, or irrelevant events of the empirical world, as for example it could

---

[6] Of course, as briefly mentioned above, scientific representations include a great deal of denoting symbols that are often called as "representational tools", ranging from texts to diagrams or mathematical expressions. Since the focus here is on the content these representations carry and not on the way it is transmitted, I will not discuss them further.

[7] Some current trends in modeling strategies work on the assumption that a unified mathematical picture of events in physics and economics is possible. However, even if a mathematical representation of these events can appear in a common schema, it does not follow that this can be explanatory, for the information required for an explanation in economics is radically different from an explanation required in physics and vice versa (Woodward, 2016, p.125).

happen when law-like or unificatory explanations are given. As for
the set of representations that represent the explanandum, they are
also open both to crude and to sophisticated optimization which make
specific which parts of the world require an explanation.

The representation-relation between the explanans and the
explanandum can be formalized and illustrated. Let an empiri-
cal representation (RE) to include a set of empirical proposi-
tions $\{RE_1, RE_2, RE_3 \ldots RE_n\}$, and likewise mingled represen-
tations $RM = \{RM_1, RM_2, RM_3 \ldots RM_n\}$. Finally let the ex-
planandum (REx) also to be a set of empirical propositions
$\{REx_1, REx_2, REx_3 \ldots REx_n\}$. The explanation occurs as shown
in Fig.1. From the parts of the world to be explained (REx, REx')
crude optimization of the explanandum rules out the parts of it that
are not of primary concern in context (REx') and puts forward the
ones that we are interested in explaining (REx). Then sophisticated
optimization makes precise which aspects of this part of the world will
be the explanandum. As shown in Fig.1, from a set of propositions
that represent aspects of this part of the world, only REx_1 and REx_3
turn out to be of particular interest and so they serve as the explanan-
dum (both crude and sophisticated optimization are represented by the
dotted upward arrows). Likewise both stages of optimization occur
when explanations to these empirical representations are pursued. The
in principle relevant answers are RE and RM, while only RE makes
it through the crude optimization process. From the propositions en-
tailed in RE sophisticated optimization rules out the ones that are
unfit for purpose in that context and makes use of two of them(RE_1
and RE_3) in order to explain REx_1 and REx_3 (both crude and so-
phisticated optimization are now represented by the dotted downward
arrows). Explanation eventually takes place when the propositions that
made it through the optimization of the explanans and the optimiza-

tion of the explanandum are related and in fact when the explanans target the explanandum (as represented by the rightward arrow).[8] Such a representation-relation schema enhances the epistemic view and cuts against the ontic conception of explanation.

In principle relevant answers: *RE, RM*

$\vdots$ *Crude Optimization*

$\downarrow$

$RE = \{RE_1, RE_2, RE_3 \ldots RE_n\}$

$\vdots$ *Sophisticated Optimization*

$\downarrow$

$RE = \{RE_1, RE_3\} \xrightarrow{\text{Explanation}} Rex = \{REx_1, REx_3\}$

$\uparrow$ *Sophisticated Optimization*

$REx = \{REx_1, REx_2, REx_3 \ldots REx_n\}$

$\uparrow$ *Crude optimization*

Parts of the world to be explained: *Rex, Rex'*
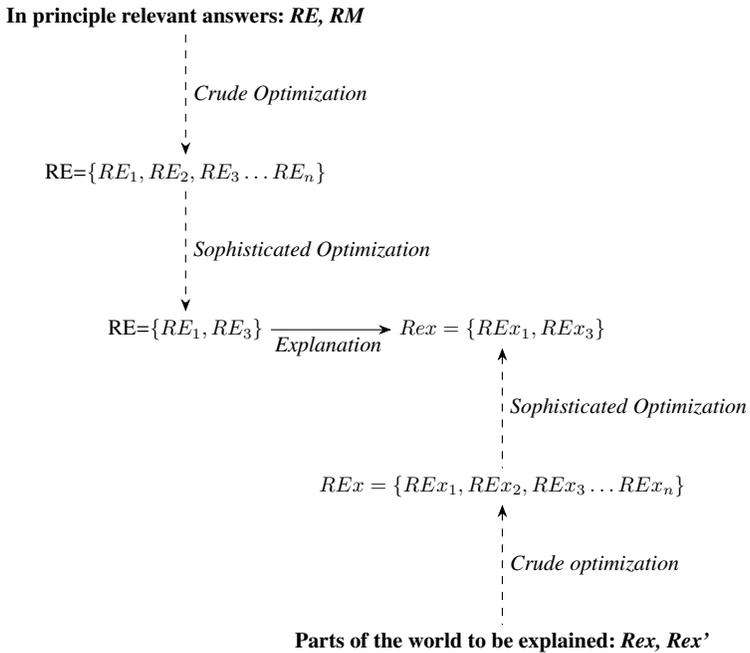
Figure 1: The representation-relation between the explanans and the explanandum.

---

[8] Note that Fig.1 is only in part used to illustrate how the relevance of each explanatory proposition is determined for this could require further elaboration that goes beyond the purposes of the current analysis; it is mostly confronted to showing that explanation is, for its best part, a relation between diverse representational schemata and that optimization processes are involved in this relationship.

To make the latter point clearer and to showcase how different stages of optimization take place in practice, we can consider a more expanded version of the IFRs' example discussed above. Suppose that experts are interested in explaining increased death rates from Sars-Cov-2 in the elderly population in the EU over the period 2020–2022 in comparison to Asian countries whereby deaths in that age-group did not soar and remained in the same ballpark to pre-pandemic levels. As "elderly population" is defined—as a matter of expert consensus—the chunk of the population that is over the age of 70. The crude optimization with respect to the explanandum therefore entails eliminating as possible explananda individuals across all age groups who died from other causes *across the globe* while sophisticated optimization rules out individuals *in the EU and in Asia* who are below 70 and who nevertheless died of Covid-19 as well as people over 70 *in the EU and in Asia* that died from other causes. There are several explanations on offer that can in principle explain such a spike in deaths from Sars-Cov-2: the IFR of Sars-Cov-2 which is orders of magnitude greater for those over 70 in comparison to those below 70 years old (Axfors and Ioannidis, 2022). Another explanation pertains to the levels of pre-existing T-Cell immunity: high levels of pre-existing cellular immunity in Asian countries were documented (Bolourian and Mojtahedi, 2021) which suggests that herd immunity was developed there even prior to the advent of Sars-Cov-2 (Le Bert et al., 2021) thus making it difficult for the virus to infect the vulnerable groups whereas in the EU the levels of pre-existing immunity were quite low and so it was easier for the virus to spread and infect people over 70. A third explanation could be that the average lifespan in the EU is higher than in Asia and so it was expectable to have more deaths in that age group. A fourth explanation could put increased deaths down to economic outcomes: it is known that one of the avails of

economic development is that healthcare facilities are better organized and thus offer high quality treatment to patients thereby saving lives. Other in principle relevant explanations can be put forward, but to keep things simple let us assume that only these are available. Crude optimization with respect to the explanans rules out the latter for EU is having, on average, a stronger economy than the majority of Asian countries and so, while such an explanation is in principle relevant, it becomes irrelevant under this context. It could also rule out the discrepancy in the average lifespan between the EU and Asia, for even if this is a matter of fact in some countries such as Afghanistan, there are Asian countries with life expectancy similar to the EU (if not higher than that), as for example Japan and Singapore, and in which the virus-related deaths did not spike (Karadimas, 2023, pp.34–35). Hence we are left with two possible explanations: Sars-Cov-2's IFR and pre-existing immunity. Sophisticated optimization will figure out which of the two carries information that best explains the explanandum or whether both, combined, offer the most relevant answer.

It appears that the ontic view struggles with such instances of explanation—which are quite common in science—for they come solely through representations and not by relating through direct observation real-objects in the world. Researchers have access to diverse representational schemata such as the IFR, the levels of pre-existing immunity and the average lifespan in several countries which all three constitute the explanans, as well as the death rates of interest, i.e. people over 70 in the EU and in Asia which are the representations that induce the explanandum, all of which (the explanans and the explanandum alike), miss, as we already discussed above, observable aspects of their target systems, and in spite of that, they are used as explanations by relating via optimization, the information each one

of them carries. An ontic approach could require directly observing all the involved variables that constitute the target systems of the explanans and the explanandum which would include observation of variables of a set of events that no human being can seriously claim to be able to observe ranging, for example, from infected individuals over 70 in the EU and in Asia, monitoring them till they die to causally relate each one's death with the virus, to the immune responses of all people in these two regions in order to determine pre-existing immunity and similarly observing everyone who dies regardless of cause so that the average lifespan will be estimated. In other words, barring the presence of a superhuman that could manage to observe all these events,[9] proponents of the ontic approach need to develop a theory showing how access to the target systems that are involved in explanations is attained without appealing to representations and so far they have not done so. Indeed, as things are, access to the "mind-independent" world is attained through a set of representations which in turn greatly weakens the ontic conception of explanation.

---

[9] Even assuming an extant and omni-observant entity from 2020–2022, since this is a non-trivial case of explanation (and most instances of scientific explanation are non-trivial) and thus it involves many variables that moreover span a wide range of locations, then this individual needs to come up with a set of representations which could summarize their observation-based findings which in turn could likely make the explanation epistemic and not ontic for explaining via representations is what lies at the heart of the epistemic approach to explanation and it seems that even such a skilled entity could end up accessing representations and explore the information each one of them carries in order to provide an explanation. Moreover, we cannot be sure that such representations include everything that occurred in the world for they, at best, would represent what the superhuman observed but it does not follow that what they observed is identical to what in fact occurred (for example some could have developed specific T-Cell responses via infection with Sars-Cov-2 and the superhuman could consider them as members of the group with pre-existing immunity since, unless antibodies are also detected, it is indistinguishable whether cellular responses came from cross-reactive immunity or via infection with the virus in question) and so even such representations are better conceptualized as empirical and not as real.

6. Conclusion: Is there room for the ontic view?

The representation-relation and the focus on the relevance of the information transmitted from the one representation to another, seems therefore to square well with much of scientific practice and with much of the philosophical theorizing on explanation, thereby bolstering the epistemic approach to explanation which seems to be rendered the prevailing philosophical view. This begs the question: is there room for the ontic view of explanation?

It can be argued that the analysis here indicates that the applicability of the ontic view is severely limited. The discussion vis-à-vis abstraction, optimization and representation and its central conclusion, namely that abstraction can be part of an optimization process and is not a distinct kind of explanation, serves as a basis for making the case that since abstract explanations are not to be considered a special case of explanation, the ontic view can no longer rely on the abstractness of a particular class of explanations in order to demonstrate its ability to capture singular ones, like the causal explanatory relations, and so it needs to capture all episodes of explanation or miss all of them at once. However, the inability of the ontic view to capture instances of explanation induced by false propositions and the fundamental epistemic problem of trying to establish explanatory relations between objects that fall outside the scientists' scope of observation, reveal problems that are avoided if explanation is considered from an epistemic point of view which implies that the ontic view is far from capturing all episodes of explanation and that it becomes highly questionable if it can conceptualize even some of them.

To be sure, it cannot be officially abolished from explanation since it is still possible to capture some cases of explanation in which directly observed objects or entities can be said to be causally related.

However, such explanations are mostly trivial examples of explanation and advocates of the ontic view bear the burden of showing that this is not the case. Even if such an account is offered, it is unclear that such an explanation will be superior, and thus more relevant, to a competing one that could capture such an episode via representations. Thus scientific explanation can be said to be in line with the epistemic conception and not that much with the ontic one.

7. Acknowledgments

Bibliography

Axfors, C. and Ioannidis, J.P.A., 2022. Infection fatality rate of COVID-19 in community-dwelling elderly populations. *European Journal of Epidemiology*, 37(3), pp.235–249. https://doi.org/10.1007/s10654-022-00853-w.

Batterman, R.W., 2007. On the specialness of special functions (the non-random effusions of the divine mathematician). *The British Journal for the Philosophy of Science*, 58(2), pp.263–286. Available at: <https://www.jstor.org/stable/30115226> [visited on 24 October 2023].

Bokulich, A., 2018. Representing and explaining: The eikonic conception of scientific explanation. *Philosophy of Science*, 85(5), pp.793–805. https://doi.org/10.1086/699693.

Bolourian, A. and Mojtahedi, Z., 2021. COVID-19 and flu pandemics follow a pattern: a possible cross-immunity in the pandemic origin and graver disease in farther regions. *Archives of Medical Research*, 52(2), pp.240–241. https://doi.org/10.1016/j.arcmed.2020.10.012.

Burgess, S., Ponsford, M.J. and Gill, D., 2020. Are we underestimating seroprevalence of SARS-CoV-2? *BMJ*, p.m3364. https://doi.org/10.1136/bmj.m3364.

Craver, C. and Bechtel, W., 2006. Mechaninsm. In: S. Sarkar and J. Pfeifer, eds. *The Philosophy of Science: An Encyclopedia*. New York: Routledge, pp.469–478.

Craver, C.F., 2005. Beyond reduction: mechanisms, multifield integration and the unity of neuroscience. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), pp.373–395. https://doi.org/10.1016/j.shpsc.2005.03.008.

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

Elgin, C.Z., 2010. Telling Instances. In: R. Frigg and M. Hunter, eds. *Beyond Mimesis and Convention: Representation in Art and Science*, *Boston Studies in the Philosophy of Science*. Dordrecht: Springer Netherlands, pp.1–17. https://doi.org/10.1007/978-90-481-3851-7_1.

van Fraassen, B.C., 1980. *The Scientific Image*, *Clarendon library of logic and philosophy*. Oxford; New York: Clarendon Press; Oxford University Press.

van Fraassen, B.C., 2008. *Scientific Representation: Paradoxes of Perspective*. 1st ed. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199278220.001.0001.

Friedman, M., 1974. Explanation and scientific understanding. *The Journal of Philosophy*, 71(1), pp.5–19. https://doi.org/10.2307/2024924.

Frigg, R. and Nguyen, J., 2017. Models and Representation. In: L. Magnani and T. Bertolotti, eds. *Springer Handbook of Model-Based Science*, *Springer Handbooks*. Cham: Springer International Publishing, pp.49–102. https://doi.org/10.1007/978-3-319-30526-4_3.

Giere, R., 2018. Models of Experiments. In: I.F. Peschard and B.C. Van Fraassen, eds. *The Experimental Side of Modeling*, *Minnesota studies in the philosophy of science*, 21. Minneapolis: University of Minnesota Press, pp.59–70.

Glass, D.H., 2021. Coherence, explanation, and hypothesis selection. *The British Journal for the Philosophy of Science*, 72(1), pp.1–26. https://doi.org/10.1093/bjps/axy063.

Glennan, S., 2005. Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. Mechanisms in biology, 36(2), pp.443–464. https://doi.org/10.1016/j.shpsc.2005.03.011.

Godfrey-Smith, P., 2009. Abstractions, Idealizations, and Evolutionary Biology. In: A. Barberousse, M. Morange and T. Pradeu, eds. *Mapping the Future of Biology: Evolving Concepts and Theories*, *Boston Studies in the Philosophy of Science*. Dordrecht: Springer Netherlands, pp.47–56. https://doi.org/10.1007/978-1-4020-9636-5_4.

Hempel, C.G. and Oppenheim, P., 1948. Studies in the logic of explanation. *Philosophy of Science*, 15(2), pp.135–175. https://doi.org/10.1086/286983.

Jansson, L. and Saatsi, J., 2019. Explanatory abstractions. *The British Journal for the Philosophy of Science*, 70(3), pp.817–844. https://doi.org/10.1093/bjps/axx016.

Jefferson, T., Heneghan, C., Spencer, E. and Brassey, J., 2020. *Are you infectious if you have a positive PCR test result for COVID-19?* Available at: <https://www.cebm.net/covid-19/infectious-positive-pcr-test-result-covid-19/> [visited on 27 October 2023].

Karadimas, P., 2022. Thought experiments and the pragmatic nature of explanation. *Foundations of Science*. https://doi.org/10.1007/s10699-022-09844-2.

Karadimas, P., 2023. *The Covid-19 Pandemic: A Public Choice View*. Vol. 42, *Studies in Public Choice*. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24967-9.

Kitcher, P., 1981. Explanatory unification. *Philosophy of Science*, 48(4), pp.507–531. https://doi.org/10.1086/289019.

Lange, M., 2013. What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science*, 64(3), pp.485–511. https://doi.org/10.1093/bjps/axs012.

Le Bert, N. et al., 2021. Highly functional virus-specific cellular immune response in asymptomatic SARS-CoV-2 infection. *Journal of Experimental Medicine*, 218(5), e20202617. https://doi.org/10.1084/jem.20202617.

Levy, A., 2021. Idealization and abstraction: refining the distinction. *Synthese*, 198(24), pp.5855–5872. https://doi.org/10.1007/s11229-018-1721-z.

Love, A.C. and Nathan, M.J., 2015. The idealization of causation in mechanistic explanation. *Philosophy of Science*, 82(5), pp.761–774. https://doi.org/10.1086/683263.

Machamer, P., Darden, L. and Craver, C.F., 2000. Thinking about mechanisms. *Philosophy of Science*, 67(1), pp.1–25. https://doi.org/10.1086/392759.

Parker, W.S., 2017. Computer simulation, measurement, and data assimilation. *The British Journal for the Philosophy of Science*, 68(1), pp.273–304. https://doi.org/10.1093/bjps/axv037.

Pincock, C., 2007. A role for mathematics in the physical sciences. *Noûs*, 41(2), pp.253–275. https://doi.org/10.1111/j.1468-0068.2007.00646.x.

Potochnik, A., 2017. *Idealization and the Aims of Science*. Chicago, IL: University of Chicago Press. Available at: <https://press.uchicago.edu/ucp/books/book/chicago/I/bo27128726.html> [visited on 27 October 2023].

Richardson, A., 1995. Explanation: Pragmatics and asymmetry. *Philosophical Studies*, 80(2), pp.109–129. https://doi.org/10.1007/BF00989758.

Rovelli, C., 2015. Aristotle's physics: A physicist's look. *Journal of the American Philosophical Association*, 1(1), pp.23–40. https://doi.org/10.1017/apa.2014.11.

Russell, B., 1956. On denoting. In: R.C. Marsh, ed. *Logic and Knowledge Essays, 1901-1950*. London: Allen & Unwin, pp.41–56.

Salis, F. and Frigg, R., 2020. Capturing the Scientific Imagination. In: A. Levy and P. Godfrey-Smith, eds. *The Scientific Imagination*. Oxford, New York: Oxford University Press, pp.17–50. https://doi.org/10.1093/oso/9780190212308.003.0002.

Salmon, W.C., 1984. Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, (2), pp.293–305. https://doi.org/10.1086/psaprocbienmeetp.1984.2.192510.

Salmon, W.C., 1989. Four decades of scientific explanation. In: P. Kitcher and W.C. Salmon, eds. *Scientific Explanation*, *Minnesota studies in the philosophy of science*, 13. Minneapolis: University of Minnesota Press, pp.3–219. Available at: <http://conservancy.umn.edu/handle/11299/185700> [visited on 4 October 2023].

Salmon, W.C., 1998. *Causality and Explanation*. 1st ed. New York: Oxford University Press. https://doi.org/10.1093/0195108647.001.0001.

Shapere, D., 1982. The concept of observation in science and philosophy. *Philosophy of Science*, 49(4), pp.485–525. https://doi.org/10.1086/289075.

Sheredos, B., 2016. Re-reconciling the epistemic and ontic views of explanation (or, why the ontic view cannot support norms of generality). *Erkenntnis*, 81(5), pp.919–949. https://doi.org/10.1007/s10670-015-9775-5.

Sheredos, B., 2019. *Re-re-reconciling the epistemic and ontic views of explanation: a reply to Wright & van Eck*. Available at: <https://dx.doi.org/10.13140/RG.2.2.14862.20805>.

Strevens, M., 2011. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

Winning, J., 2020. Mechanistic causation and constraints: Perspectival parts and powers, non-perspectival modal patterns. *The British Journal for the Philosophy of Science*, 71(4), pp.1385–1409. https://doi.org/10.1093/bjps/axy042.

Woodward, J., 2002. What is a mechanism? a counterfactual account. *Philosophy of Science*, 69(S3), S366–S377. https://doi.org/10.1086/341859.

Woodward, J., 2016. Unificationism, Explanatory Internalism, and Autonomy. In: M. Couch and J. Pfeifer, eds. *The Philosophy of Philip Kitcher*. Oxford University Press, pp.121–152. https://doi.org/10.1093/acprof:oso/9780199381357.003.0006.

Wright, C., 2012. Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science*, 2(3), pp.375–394. https://doi.org/10.1007/s13194-012-0048-8.

Wright, C., 2015. The ontic conception of scientific explanation. *Studies in History and Philosophy of Science Part A*, 54, pp.20–30. https://doi.org/10.1016/j.shpsa.2015.06.001.

Wright, C. and van Eck, D., 2018. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Ergo, an Open Access Journal of Philosophy*, 5(38). https://doi.org/10.3998/ergo.12405314.0005.038.

Dimensions of explanation

Eric Hochstein
University of Victoria

Abstract

Some argue that the term "explanation" in science is ambiguous, referring to at least three distinct concepts: a communicative concept, a representational concept, and an ontic concept. Each is defined in a different way with its own sets of norms and goals, and each of which can apply in contexts where the others do not. In this paper, I argue that such a view is false. Instead, I propose that a scientific explanation is a complex entity that can always be analyzed along a communicative dimension, a representational dimension, and an ontic dimension. But all three are always present within scientific explanations. I highlight what such an account looks like, and the potential problems it faces (namely that a single explanation can appear to have incompatible sets of norms and goals that govern it). I propose a solution to this problem and demonstrate how this account can help to dissolve current disputes in philosophy of science regarding debates between epistemic and ontic accounts of mechanistic explanations in the life sciences.

At its metaphysical core, what exactly is a scientific explanation? Is it a communicative act whereby we make some phenomenon intelligible or understandable to an audience (e.g. Nancy explains the

orbits of the planets to John)? Is it a type of theory, model, or representation (e.g. *The theory of evolution* explains why elephants have long trunks. Einstein's *General Theory of Relativity* explains why light bends when it passes a large gravitational body)? Or is it a collection of entities and relations out in the world (e.g. The car's velocity and the cut break line explains why the car crashed into the storefront)? Historically, philosophers have disagreed about which of these provides the best metaphysical account of what a scientific explanation is. For example, Peter Achinstein insists that "the illocutionary concept of explaining is fundamental and that other explaining concepts are explicable by reference to this" (1984, p.22). Conversely, Ruben insists that "explanations work only in virtue of the determinative relations that exist in the world" (Ruben, 1990, p.231).

In recent years, some have suggested that perhaps there are simply distinct concepts of "explanation" at work. Each is employed in different contexts within scientific practice, but we must be cautious not to equivocate between them. This idea has strong support. The goals and norms that seem to govern the various concepts of explanation appear to be different, and indeed can even be at odds with one another. Moreover, it seems we can often have one concept of explanation apply in a context where the others are not applicable at all. All this suggests that we may simply mean different things by the word "explanation" in different scientific contexts.

Despite the intuitive strength of such a position, in this paper I argue that this is not the case. Instead, we should think of a scientific explanation as an extremely complex and multifaceted entity. Every explanation can be analyzed along different dimensions. When each dimension of explanation is considered in isolation of the others, it provides a distorted account of what explanations are, and how they

function. But understanding how these different dimensions constitute one and the same explanation can help to clarify current disputes in the philosophy of science.

In order to make this argument, I start in section 1 with an account of why it is plausible to think that there are at least three distinct concepts or definitions of explanation: explanation defined as a *communicative* act (hereafter the "communicative" concept of explanation), explanation defined as an appropriate kind of model or theory (hereafter the "representational" concept of explanation), and explanation defined in terms of ontic structures, causes, and relations in the world that produce and sustain the phenomenon in question (hereafter the "ontic" concept of explanation). In section 2, I demonstrate why we should not consider these as distinct concepts of explanation after all, and why explanation always requires the interaction of all three elements. In section 3, I highlight why this shift from "concepts" of explanations to "dimensions" of explanation is not a trivial one, and explore how this new account can shed light on debates in the philosophy of science. Lastly, in section 4, I highlight the complications that this new account faces, and how they might be addressed.

1. Different Concepts of Scientific Explanation

An important insight that some have emphasized in recent years is that the term "explanation" may itself be ambiguous, denoting entirely distinct concepts. Waskan et al. (2014), for instance, argue that "'Explanation' appears to be ambiguous between a representational-artifact, an objective, and a doxastic sense" (Waskan et al., 2014, p.3090). Meanwhile, Craver (2014, p.29) argues that we should "disambiguate four ways of talking about explanation: as a communicative

act, as a representation or text, as a cognitive act, and as an objective structure". Likewise, Gilpin et al. (2022, p.3) argue that "the term 'explanation' is a classic example of what Marvin Minsky referred to as suitcase words: words that contain multiple meanings which are interpreted in different ways for different people in different contexts".

For the purposes of this paper, I want to focus on three different concepts of explanation, what I will call the "communicative" concept, the "ontic" concept, and the "representational" concept.[1] To begin, consider the communicative concept of explanation. This concept defines explanation as a social activity we engage in whereby we try to make some phenomenon understandable or intelligible to a given audience. As Achinstein (1983, pp.16–17) puts it, "S explains q by uttering u only if S utters u with the intention that his utterance of u render q understandable". This definition of explanation can also be seen in Craver's example of Jon, a professor trying to *explain* the action potential of the neuron to a classroom full of students (Craver, 2014, p.30). Similarly, Andrea Woody describes the Communicative concept of explanation in the following way:

> There is always, at least implicitly, a request (question) and a response, and these are typically negotiated or shared among individuals or groups. Explanations frequently pass between individuals or groups with different levels of expertise. A person requesting an explanation often is not in a good position

---

[1] This is not intended to be an exhaustive list, and other concepts of explanation may well be worth including. For instance, both Waskan et al. and Craver acknowledge a *psychological* or *cognitive* concept of explanation. In this paper, I collapse the "communicative" and "cognitive/doxastic" sense of explanation into one (since the goal of the communicative concept of explanation is taken to be providing *understanding* to an audience, which is itself a psychological/cognitive phenomenon). That being said, if one wishes to differentiate these then the argument in this paper will still go through. For the sake of brevity, however, I will focus on these three concepts of explanation.

> to judge the adequacy of the response, and must rely on the authority of the responder to accept the explanation as legitimate (Woody, 2015, p.81).

When thought of in this way, we can view explanation as having particular goals and norms that dictate its proper application. In Craver's example of Jon giving a lecture to a classroom of students, he tells us that if all goes right, then "the audience comes to understand how action potentials are produced." (Craver, 2014, p.30). Or consider Achinstein's definition above in which an explanation involves an utterance from a speaker *with the intention that their utterance renders some phenomenon understandable*. In this respect, we can see *understanding* or *intelligibility* as an essential goal of explanation, and the norms of good explanation can be thought of in terms of the norms of good pedagogy.

The communicative concept of explanation can be contrasted with another: the ontic concept of explanation. This concept of explanation is most commonly attributed to José Alberto Coffa (1974), but has been advocated for by many others since (Salmon, 1989; Machamer, Darden and Craver, 2000; Craver, 2006; Izadi, Anandakrishnan and Onufriev, 2014; Kaplan, 2011; Strevens, 2011; Waskan, 2011; Povich and Craver, 2018; Craver and Kaplan, 2020). Salmon describes the ontic concept of explanation as follows:

> Proponents of this conception can speak in either of two ways about the relationship between explanations and the world. First, one can say that explanations exist in the world. The explanation of some fact is whatever produced it or brought it about. [. . .] [I]t seems entirely appropriate to say such things as the gravitational attraction of the moon explains the tides, or the drop in temperature explains the bursting of the pipes. The gravitational attraction and the drop in temperature are out

> there in the physical world; they are neither linguistic entities, nor abstract entities. Second the advocate of the ontic conception can say that an explanation is something—consisting of sentences or propositions- that reports such facts (Salmon, 1989, p.86).

The idea is that a good explanation must identify the appropriate *causes, structures, or features out in the world* responsible for the occurrence of the explanandum phenomenon in order to be genuinely explanatory. To borrow an intuitive example from Craver & Kaplan (2020, pp.299–300), the explanation for why sea levels are rising is because of *global warming*. Note, the explanation of sea levels rising is not *my saying the words* "global warming" to an audience, nor is it a model or a theory of global warming by itself. What explains the sea levels rising out in the world is the actual warming of the planet. A scientific model of global warming is thought to be explanatory in the ontic sense in virtue of identifying the relevant dependencies in nature (i.e. the changing climate) which genuinely produces and thus explains the rising sea levels.

While many think that the ontic concept of explanation tends to presuppose a largely mechanistic view of explanation (i.e. that good explanations must identify *ontic mechanisms*), this need not be so and may include other kinds of entities/relations as well. Craver and Povich, for example, claim that "we adopt a more inclusive understanding of the ontic that embraces any natural regularity, e.g., statistical relevance, natural laws, or contingent compositional relations might also figure fundamentally in explanation" (Craver and Povich, 2017, p.32; see also: Povich and Craver, 2018). If one adopts this ontic definition of explanation, then there are different goals and norms that present themselves. The goals associated with this definition of

explanation are tied to things like identifying the relevant or appropriate natural regularities, laws, mechanisms, or dependencies out in the world that produce or sustain the phenomenon.

Traditional explanatory virtues such as *precision* and *depth* are often associated with this definition of explanation, as the more precise our descriptions of the ontic structures, causes, and regularities becomes, the deeper the explanation becomes. This idea is explicitly defended by Mackonis, who claims that "[explanation] H1 is *deeper* than H2 if H1 explicates a causal-nomological mechanism that produces the abductive trigger and H2 does not" (my emphasis), and that H1 is deeper than H2 if the mechanism posited by H1 is more "specific, precise, [and] fundamental" than one posited by H2 (Mackonis, 2013, p.985). Others have likewise associated the explanatory virtues of depth and precision with the ontic definition of explanation (see, Lipton, 2004; Strevens, 2011; Craver and Kaplan, 2020, p.313).

According to this definition of explanation, an essential norm of good explanation is that the more we can identify the ontic dependencies relevant to the production of the phenomenon in the world, the better the explanation becomes. As Craver (2014, p.41) argues, "the norms of scientific explanation fall out of a prior commitment on the part of scientific investigators to describe the relevant ontic structures in the world".

Lastly, let us consider the representational concept of explanation. We frequently talk of explanations in terms of scientific *models* or *theories*. For example, we might talk about how G.A. Parker's (1978) *model of dung fly copulation* explains why dung flies copulate for approximately 36 minutes (see, Rice, 2015); or how Heeger's (1992) *Normalization Model* explains cross orientation suppression (see, Chirimuuta, 2014). As Waskan et al. (2014, p.3090) argue, "On one

common manner of speaking (e.g., 'There is an explanation for the odd trait on page 25') the noun 'explanation' does seem to refer to a set of representational-artifacts".

When dealing with this concept of explanation, we again have different norms and goals that govern our explanatory practices. The goal of explanation under this definition involves our ability to represent phenomena in the appropriate sort of way to gain relevant insights into it. These insights are traditionally cashed out in terms of our ability to predict when the phenomenon occurs, describe its behaviour, identify constraints on the phenomenon, or the patterns/principles it obeys (e.g., Batterman, 2000; 2002; Batterman and Rice, 2014; Chirimuuta, 2014). This definition of explanation is often associated with the traditional explanatory virtues of *breadth* and *unification*. Breadth and unification are often characterized as the ability of an explanation to account for a wide range of disparate phenomena. The more phenomena that can be explained by the model or theory, the greater its explanatory breadth and ability to unify (see, Mackonis, 2013; Lombrozo, 2016; Mantzavinos, 2016). For instance, Mantzavinos characterizes unification in terms of explanations that seek "laws and principles of high generality with the aim of constructing a coherent world picture and fitting particular facts within this framework" (Mantzavinos, 2016, p.6, footnote). Thus, under this definition of explanation, a model that describes a set of patterns or principles that we can subsume a greater range of phenomena under, or which identifies specific constraints that apply to a greater range of phenomena, will be better explanations than models that do not. For instance, a continuum model is taken to be a good explanation of the flow behaviour of liquids given that it can successfully predict and describe the behavioural patterns of a wide range of different fluids using the same formalism (see, Batterman and Rice, 2014; Izadi, Anandakrishnan and Onufriev,

2014; Bokulich, 2018). Different kinds of predictions, constraints on phenomena, patterns that phenomena adhere to, and principles that govern phenomena, require different kinds of models or theories to be properly identified.

All this suggests that the distinct concepts of explanation ("communicative", "ontic", and "representational") involve different goals and norms that govern good explanation. Moreover, it also seems to be the case that one concept of explanation can apply in situations where the others do not. For instance, let us return to our previous example of rising sea levels. Global warming explains the rising sea level even if no one understands that the globe is warming, or whether we have a scientific model or theory which can adequately represent it. In this case, it seems like the ontic definition of explanation applies even though the communicative definition, and the representational definition, do not. Conversely, we might claim that Darwin's Theory of Evolution is a *good* explanation of why elephants have long trunks even if the audience I try to explain this to doesn't understand it (if, for instance, I am explaining the theory to a group of toddlers), and even if we do not yet know what the underlying ontic dependencies are that influence selection and genetic transmission. Lastly, we might do pedagogical research about the best ways to explain scientific principles to non-experts, and discover that certain kinds of techniques (e.g. frequent quizzes, having class discussions, one-on-one tutoring sessions, etc) will allow us to better explain. In which case, the communicative concept of explanation would apply, but the representational and ontic concepts do not (as I can discuss good explanatory techniques without invoking any particular scientific model or theory, or identifying any particular ontic structures and causes in nature).

To sum up, there are good reasons to think that the word "explanation" is ambiguous, referring to at least three distinct concepts

that are invoked in science. Each brings with them distinct goals for explanation, and distinct norms that govern good vs bad explanation. Moreover, it seems that each definition of explanation can apply in situations where the others do not. Working scientists thus need to be careful not to equivocate between the different concepts. While this sounds intuitive and plausible, I propose that this is, in fact, the wrong story to tell. Explanation is far more complex that it may initially seem, and these different concepts of explanation are in fact far more interconnected and interdependent then they may appear. In the section that follows, I want to propose an alternative way of thinking about scientific explanation and motivate this new way of approaching it.

2. One Explanation with Multiple Dimensions

Instead of assuming that we are working with three distinct *concepts* of explanation (each defined in a different way), I propose that a single explanation always has a communicative *component*, a representational *component*, and an ontic *component*. In other words, these are not distinct concepts of explanation as much as they are different evaluative dimensions along which we can analyze a single explanation. This point is not a trivial one and has the potential to help reframe current disputes within the philosophy of science. In order to demonstrate this, let us begin with the hypothesis that these *are* intended to be entirely distinct concepts of explanation, and then explore how each concept of explanation in fact requires or incorporates the others.

Let's begin with the ontic concept. To say that we can have an ontic concept of explanation apply independent of a communicative and representational one would be misleading. To illustrate, consider

the phenomenon of phase transitions in statistical mechanics. When we boil a pot of water, it transitions from a liquid state to a gaseous one. Yet our ability to explain this phenomenon is notoriously difficult since we can only model this sort of transition if we mathematically treat the pot of water as being infinitely large, allowing the molecules an infinite degree of freedom (Batterman, 2002). Now, we know that ontically the transition from liquid to gas is somehow accomplished by the interaction of the finite molecules that make up the water, and the pot containing it. Yet, it would seem unhelpful to suggest that we've *always* had a good scientific explanation of phase transitions by merely *pointing to* a pot of boiling water. Here, we are denoting the ontic structures and causes out in nature that produce the phenomenon. Yet, we are no better at *scientifically explaining* the phenomenon than we were before. This is precisely why William Bechtel (2008, p.18) notes that "the problem with this ontic view is that mechanisms do not explain themselves".

This has led several theorists to point out that those who claim to be working strictly within the ontic concept of explanation do not, and cannot, eliminate the communicative and representational dimensions from how they talk about explanations (Bechtel, 2008, p.18; Wright, 2015; Bokulich, 2018). For instance, Bechtel argues that:

> Even the advocates of the ontic perspective are unable to avoid invoking epistemic notions, although they try to minimize them. Machamer et al. sometimes refer to "giving a description of the mechanism" (p. 3) and "revealing ... productive relations" (p. 21), and Salmon uses such words as exhibiting. But these terms understate the cognitive labor involved (Bechtel, 2008, p.18).

Or consider Cory Wright (2015, p.27), who notes that "Salmon offered a putative example of ontic explanation that centers on an

epistemic agent who makes discoveries, and so is thereby positioned to elaborately communicate a causal story". Many who endorse the ontic definition of explanation have granted that there is an essential representational and communicative component to explanation that is intertwined with the ontic one. Mark Povich, for example, points out that:

> Craver's ([2014]) most recent formulation of the ontic conception backs away from the metaphysical claim that explanations are ontic structures in the world and focuses on demarcatory and normative constraints on explanation. Craver ([2014]) writes that according to the ontic conception, 'in order to satisfy these two objectives [of explanatory demarcation and explanatory normativity], one must look beyond representational structures to the ontic structures in the world' (Povich and Craver, 2018, p.129).

Similarly, Waskan (2011, p.4) points out that "ontic theories might take many forms, so long as what they propose is that explanations (primarily) reveal something about objective states of affairs. [...] Explanations are, on this view, representations—objective facts are not in the business of revealing. Specifically, they are descriptions". This point has likewise been emphasized by Phyllis Illari, who argues that advocates of ontic explanation acknowledge the essential role that representation and modeling plays in scientific explanation, and that the issue instead is one of normative constraints. In other words, that "ontic explanation is essential for marking several crucial normative dimensions by which scientific explanations are and ought to be evaluated" (Illari, 2013, p.243).[2] This provides evidence that there is

---

[2] This shift from ontic structures being explanations in and of themselves, to the idea that identifying the appropriate ontic structures provides essential normative constraints on explanation may lead some to conclude that we've weakened the ontic

not a distinct *ontic concept* of explanation, there is instead an ontic *dimension* to a scientific explanation that comes along with both representational and communicative dimensions.

To further highlight this idea, let us turn to the representational concept of explanation. It can often be the case that various theories or representations are treated as good explanations even when they are not understandable to a given audience (which runs counter to the communicative definition of a good explanation), or when no ontic dependencies are being identified by the representation in question (which runs counter to the ontic definition of a good explanation).

---

account of explanation to the point of triviality. If the claim merely reduces to some form of epistemological realism (i.e. that there are objective features of the world that we can learn about), then the claim is not particularly informative. Even those who explicitly reject the ontic definition of explanation would be willing to grant such a claim. Such a criticism would be uncharitable, however. The ontic concept of explanation does not merely insist that our models must describe some objective world, but that a good explanation must identify the *appropriate* ontic dependencies in the world necessary to account for the explanandum phenomenon. Specifically, those responsible for the production and sustaining of the phenomenon. This claim is directly at odds with those who insist that highly idealized models which do not identify any such ontic dependencies can still count as genuine explanations (e.g. Batterman, 2002; Batterman and Rice, 2014; Chirimuuta, 2014; Rice, 2015; Rice and Rohwer, 2020). Put another way, the ontic definition need not be committed to the idea that the ontic dependencies in the world are somehow self-identifying or self-explaining. It instead need only be committed to the idea that there are relations that exist between events/dependencies in the world such that some set produces and sustains the other, and that appropriate scientific explanations are those that correctly highlight such relations and dependencies. This idea has always been implicit in the ontic definition of explanation, with Salmon (1989, p.86) telling us that "the advocate of the ontic conception can say that an explanation is something—consisting of sentences or propositions- that reports such [ontic] facts", and that "one can properly say either that the explanandum-fact is explained by the explanans-facts or that the explanans-statements explain the explanandum-statement". Although it is worth noting that debates about whether this weaker interpretation of the ontic view is still problematic are ongoing. For details, see: (Illari, 2013; Wright and van Eck, 2018).

This seems to provide compelling evidence for the idea that there are distinct concepts of explanation at play. Let us examine the intuitive strength of these claims before responding to them.

Suppose a theoretical neuroscientist uses a computational model to provide a scientific explanation of the behaviour of a particular neural circuit. If the audience of the explanation happens to be a group of kindergarten children, then the computational model in question will fail to provide them with any sort of understanding of neural circuits. Thus, it would fail to be a good explanation under the communicative definition of explanation. However, such a failure on the part of the children to understand the model does not undermine the idea that the model is still considered a good explanation by the scientific community, suggesting that a different concept of "explanation" is at work.

We must be cautious with such a conclusion, however. This is because when dealing with *scientific* explanations, the audience who must understand the model or theory is not *any* audience. Put another way, the computational model is a good scientific explanation because it provides *working scientists* with an understanding of the neural circuit. Indeed, studies have shown that "high-science participants were less likely to regard a passage of text as an explanation when told that the representation, because of qualitative barriers, lacks intelligibility" (Braverman et al., 2012, p.1372). In other words, if a model or theory cannot provide an explanation *in accordance with the communicative definition* to working scientists, then scientists themselves do not consider the model or theory to be genuinely explanatory. The importance that understanding-to-a-scientific-audience has for the representational concept of explanation can be seen in the way that scientists explain to non-scientists. As Woody (2015, p.81) notes, "explanatory discourse often involves the communication of exemplary

explanations among members of a given scientific community, and one aim of explanatory practice seems to be training novice practitioners to recognize typical explanatory patterns." And so to suggest that a scientific model or theory is explanatory independent of whether it provides understanding to an audience is deeply misleading, and one cannot tease apart a representational concept of explanation from a communicative one in this regard.

But what about the relationship between the representational concept and the ontic concept? There seem to be many examples in which a model is considered explanatory despite not identifying (or indeed deliberately misrepresenting) the ontic variables in nature responsible for the production of the explanandum phenomenon. Optimality models in evolutionary biology, for example, deliberately ignore or distort causal features in the evolutionary process in order to mathematically determine what sorts of traits would be locally optimal for an organism to have, thus explaining why such traits were selected for (Potochnik, 2010; Woods and Rosales, 2010; Rice, 2015). Likewise, continuum models in physics are used to explain the flow behaviour of liquids despite deliberately saying nothing as to the molecular components and causal processes that produce such behaviours (Batterman and Rice, 2014; Izadi, Anandakrishnan and Onufriev, 2014; Bokulich, 2018, p.803). In these instances, it appears like there is no ontic component to the explanation, suggesting a distinct concept of explanation.

Yet these cases are not as straightforward as they initially seem either. For instance, take optimality models. While it may be true that an optimality model does not identify the ontic structures and causes in nature responsible for the production of phenotypic traits, it is not true that a commitment to what those structures and causes are is not an essential part of the explanation. This is because the application

of optimality models *only* works when scientists already know essential ontic structures and processes at work in the production of the phenomenon. Angela Potochnik, for example, notes the following regarding the use of optimality models in evolutionary biology:

> Because optimality models use highly simplified assumptions as placeholders for complex dynamics, their successful use depends upon evolutionary dynamics that the models themselves do not explicitly represent. In other words, optimality models are epistemically dependent on unrepresented dynamics. Information about these unrepresented dynamics helps establish whether an optimality model's simplifying assumptions are problematic, and thus how successful the model is (Potochnik, 2010, p.226).

And so the ontic dimension of the explanation is essential here. The model in question cannot function appropriately without the relevant ontic structures and processes being identified and understood by scientists, even if they are not explicitly stated in the model or theory itself. While a focus on the *representational dimension* of explanation does not highlight the ontic features of the explanation, they are a necessary part of the explanation itself, since they are required for the representations to function appropriately.

For another example, consider continuum models. As Bokulich (2018, p.803) rightly points out, "at large scales, continuum representations of water and the Navier-Stokes equations are typically more relevant" than other kinds of scientific models or representations for characterizing flow behaviour. The Navier-Stokes equations predict how fluids will behave even though they do not identify any of the underlying molecular components of the fluid. Instead, they treat the

fluid as a single continuous entity. Such models would appear to explain why fluids behave as they do despite saying nothing as to the ontic variables responsible for it.

Just like the case with models in evolutionary biology however, such models in physics are likewise *still dependent on scientists being commitment to ontic variables that are not included in the model itself*. In this case, the ontic properties of the molecules that compose the fluid must be correctly understood by working scientists for proper application of the Navier-Stokes equations. Mark Povich, for example, points out that the Navier-Stokes equations cannot be used to account for the behaviour of fluids like liquid crystals. This is because "their often rod-shaped particles result in directional preference and lack of symmetry. Liquid crystals thus cannot be accurately modelled using the unmodified Navier-Stokes equations. The addition of a stress tensor or coupling with a Q-tensor system is required to take into account the anisotropy of liquid crystals" (Povich, 2018, p.124). Note that while the Navier-stokes equations themselves do not identify the molecular properties of the fluid, scientists must be committed to various ontic properties of the molecules that make up the fluid in order to tell if and when the equations will work, or if and when they will need to be modified.

Some, like Potochnik (2010), acknowledge the importance that these ontic commitments play in constructing and applying these idealized representations, but insists that it is the idealized model that is explanatory independent of these commitments. This is because while these ontic commitments are necessary to construct and apply something like an optimality model, the model itself can only identify the relevant evolutionary patterns or principles when these ontic structures and causes are excluded from the representation itself. By trying to add such details to the representation, it becomes worse at

identifying the patterns and principles we seek to represent. Thus, the model is explanatory independent of those ontic details. Similar arguments are also made in the context of highly idealized models in physics (see Batterman, 2002; Batterman and Rice, 2014).

The problem with this style of response is that it equates an explanation with a scientific model or theory. In this case, since the model or theory does not identify the relevant ontic variables, then neither does the *explanation*. However, a scientific explanation should not be conflated with a model or theory. A scientific representation is an integral *part* of an explanation, but the explanation itself includes elements that go beyond what is included within any single representation.

We have already seen evidence of this regarding the role that understanding plays in explanation. A model or theory may, for instance, correctly identify the pattern we care about, but if working scientists cannot understand the model or theory (i.e. if it is abstracted away from the communicative dimension of explanation), then scientists will not consider it explanatory. This already suggests that what constitutes an explanation goes *beyond* what is explicitly stated by any particular model or theory. Just as a communicative dimension is essential for any particular representation to count as a *part* of the explanation, so too is a commitment to essential ontic features of the world (even if they are not stated by the model or theory itself). A commitment to what the actual ontic variables in nature are plays an essential role in the creation of both optimality and continuous flow models, their boundary conditions, their applications, and what inferences we are licensed to draw from them (see Hochstein, 2019). In this regard, these ontic commitments are intertwined with the representational content of the models to instantiate the explanation. To pull them a part is to strip the representation of its ability to carry

explanatory content. In other words, we cannot cleanly demarcate a *representational* concept of explanation from an *ontic* concept of explanation.

Lastly, let us turn to the communicative concept of explanation. While we can talk about an act of explaining abstracted away from any particular representation or ontic commitments, the representational and ontic components of explanation are always present and ineliminable in any particular instance of scientific explanation as a communicative act. As Illari notes, when talking about scientific explanation (even as a communicative act), one cannot ignore the ontic dimension of explanation given that "explanations cannot ignore worldly things" (Illari, 2013, p.251).

To illustrate, recall that the primary goal of the communicative definition of explanation is to provide *understanding* in an audience. However, our understanding of a phenomena cannot be divorced from our ontic commitments about essential structural and causal features in the world since part of what it is to understand is to correctly account for such features. Waskan, for example, points out that "understanding" is frequently treated as a "success verb much like (a sense of) 'see'—that is to say, in order to do it you must be successful at it. We might say, for instance, that whereas alchemists *felt* that they understood combustion, chemists do *genuinely* understand combustion" (Waskan, 2011, emphasis in the text). Bechtel (2008, p.14), Strevens (2011, p.3), and Illari (2013, p.245) make similar claims. Likewise, both Elgin (2004) and Potochnik (2015) argue that truth is an essential *threshold* concept when it comes to understanding. In other words, understanding requires that one correctly accounts for at least *some* of the ontic dependencies out in the world that are directly responsible for the production of the phenomenon. As Potochnik (2015, p.73) puts it, "a claim must be 'true enough' in

order to be epistemically acceptable; that is, any divergence from the truth must be negligible, or safely neglected". While both Elgin and Potochnik argue that the amount of truth required for an account to provide genuine understanding will vary based on our pragmatic needs, it highlights how the ontic dimension of explanation cannot be conceptually *divorced* from any account of understanding, making it an eliminable part of the communicative concept of explanation.

A very different kind of argument for the essential connection between the communicative, ontic, and representational definitions of explanation can also be found in the way that the various goals associated with scientific explanation are unavoidably intertwined and dependent on one another for their explanatory power (Hochstein, 2017). Consider that the primary goal associated with the ontic concept of explanation is the identification of relevant ontic dependencies existing in the world. Meanwhile, one of the primary goals associated with the representational concept of explanation is the ability to represent certain behavioural patterns that a given phenomenon adheres to. But what justifies the explanatory status of such goals? Put another way, *why* is the identification of ontic dependencies explanatory? What *makes it* explanatory? The same with the identification of behavioural patterns, or with prediction. Why consider such goals relevant to explanation? Typically, theorists will justify the explanatory status of one explanatory goal *by appealing to the other goals.*

For instance, we often justify the explanatory status of identifying ontic dependencies by appealing to the fact that knowing such dependencies allows us to better control and manipulate the phenomenon, which in turn allows us to better *predict* and *understand* it. Thus, we justify the explanatory goal of the ontic concept of explanation by appealing to the fact that it allows us to better satisfy the goals of the communicative or representational concepts of explanation. Similarly,

we might ask why predicting the occurrence of the phenomenon in a range of cases counts as explanatory. In such cases, it is common to justify the explanatory status of prediction on the grounds that "predictions help us check whether our accounts of the world have any veracity" (Douglas, 2009, p.453). Here we justify the explanatory status of a goal identified by the representational concept of explanation by appealing to how it helps us to better attain a goal identified by the ontic concept. This shows that the various goals of explanation are in fact inter-defined and inter-dependent Building on this idea, I propose that instead of having distinct concepts of explanation, we have different evaluative dimensions of an explanation that are similarly inter-defined and inter-dependent.

All of this provides substantial evidence for the idea that there is always an unavoidable ontic, communicative, and representational dimension to one and the same explanation. While we can talk about these different evaluative dimensions abstracted away from one another, we must be cautious not to assume that these dimensions denote entirely distinct concepts of explanation that are autonomous and that apply in different contexts.

3. A Distinction Without a Difference?

Before proceeding further, it is important to highlight why this alternative view of scientific explanation is worth articulating. After all, are "dimensions" of explanation really that different from "concepts" of explanation? Am I merely drawing a distinction without a difference? Why is it important to demarcate these two positions? There are important implications to the view being defended here.

To assume that there are simply distinct concepts of explanation leads to the deeply problematic assumption that the goals and norms of a particular concept of explanation, in the appropriate context, is always sufficient by itself to account for a good scientific explanation. However, this would be to distort how explanations work in science, and why they are successful. To illustrate, consider Mantzavinos's example of neoclassical economics. While Mantzavinos does not distinguish explanations in terms of communicative, representational, or ontic concepts, he does imply such a distinction when insisting that models from neoclassical economics count as genuine scientific explanation in virtue of meeting the criteria associated with the representational concept, but not those associated with the ontic one. Specifically, he notes that neoclassical models are considered good scientific explanations within the field of economics because they unify a range of economic phenomena under a particular set of principles (Mantzavinos, 2016, p.12). However, he notes that such models *do not need* to identify relevant ontic structures or causes in nature to count as explanatory, since in the context of economics, this explanatory goal is simply not what is adopted by working scientists.

For example, he tells us that "microeconomics textbooks never pay any attention to 'causal processes, causal interactions, and causal laws', in the analysis of any type of market, be it the competitive market, a monopoly or an oligopoly", and that such models do not identify the sort of ontic dependencies in the world that we can intervene on or manipulate because "it operates at a level of abstraction that makes it extremely difficult to test the theory empirically" (Mantzavinos, 2016, p.13). He ultimately concludes that the "decision calculi of neoclassical economic theory are clearly argument patterns that can be only accommodated by the unification model of explanation" (Mantzavinos, 2016, p.13). For our purposes, we can reasonably interpret this

as saying that for economists, it is the representational concept of explanation that matters (since it is the unification of various market phenomena under a single model or theory that makes it explanatory), not the ontic concept (since identifying ontic structures, causes, and variables that can be manipulated, are not taken to be explanatorily relevant in this context).

But we have good reasons to challenge Mantzavinos on this point. Whether such neoclassical models identify genuine principles and regularities in market forces, or merely notes contingent correlations that do not genuinely unify market phenomena at all, cannot be determined *unless we appeal to ontic dependencies and causal variables in the world*. Neoclassical economics has built into it certain implicit ontic commitments about the structures and dependencies in the world that must be true in order for the model to genuinely provide a unifying account. If these sorts of ontic dependencies are false, then the model's ability to unify successfully, and thus meet its own standards of good explanation, are undermined. Similarly, if the model does not identify ontic variables we can intervene upon, we have no way of testing or determining whether the neoclassical model is actually identifying unifying market forces since there is no way to manipulate relevant features of the market to see whether its underlying principles apply or not. This is precisely why *economists themselves* have questioned whether such models should be taken to provide genuine scientific explanations.

For a rather stark example, consider Alfred Eichner's now famous argument for why neoclassical economics fails to count as a science, and to provide scientific explanations. He notes that neoclassical economics is built on numerous implicit ontic assumptions about the underlying structures and dependencies in the world which it uses as the foundation for its unifying models, and that many such ontic

assumptions have proven false (for discussion of such incorrect ontic comments, see: Eichner, 1983, p.511). In virtue of getting these ontic dependencies wrong, Eichner argues that the explanatory value of such models is undercut given that they fail to account for the market forces that genuinely exist.[3] In other words, the model fails to unify the actual range of phenomena in the world it is attempting to explain. As he puts it:

> Economists as a group have adopted the view that formal, or mathematical, proofs are sufficient to establish the validity of a theory rather than just being necessary. [. . .] Thus it is common for the "theorists" to set up their models in such a way that the postulated behavior runs counter to all that is known about actual economic systems without this fact impugning either the argument or the economists' reputation. So sharp a distinction between theory and empirical research is unknown in the natural and biological sciences, and for good reason. It leads to an outpouring of useless theories that waste the time and energy of empirical researchers (1983, p.517).

Similarly, the lack of intervention that such models provide is likewise taken to undermine the scientific and explanatory status of such models. This is because "the theory must be shown to make a difference to society, when translated into one or more public policies, that will lead to certain clearly distinguishable results. The policies must then be adopted and the predicted effect confirmed. This is the praxis test of a social science theory" (Eichner, 1983, p.510). Based on these very reasons, Eichner concludes that neoclassical economics fails to be scientific, and its models fail to provide genuine scientific explanations. Note that the issue here is not whether economists think

---

[3] For others who likewise emphasize this point, see: (Hall et al., 2001; Hausman, 2008; Keen, 2022).

that unifying market forces within a model or theory is explanatory. The issue is that the ability of the model to successfully meet this criterion *requires* that the norms and goals of the ontic dimension of explanation be adopted. Mantzavinos suggests that economists themselves don't consider the goals of the ontic dimension relevant to explanation, but this is untrue. Even Milton Friedman himself grants that keeping track of these implicit ontic commitments are "extremely valuable in suggesting leads to follow in accounting for divergences between predicted and observed results; that is, in constructing new hypotheses or revising old ones" (Friedman, 1953, p.31n). Put simply, *economists themselves* have emphasized the importance of the ontic dimension of explanation when evaluating whether neoclassical models are successful at unifying phenomena under an appropriate scientific representation.

To assume that there is a representational concept of explanation that is distinct from the ontic concept and the communicative concept, and that the application of this one concept is sufficient for explanation in the right contexts, ignores that each "concept" in isolation does not, and cannot, do the explanatory work that it is expected to do unless we smuggle in the other concepts of explanation with it. It is important to distinguish "concepts" of explanation from "dimensions" of explanation since one implies that we can satisfy one concept of explanation without having to say anything about the other concepts, and that *good* scientific explanation is determined entirely by the particular concept of explanation we employ. If we understand these as different dimensions of the same explanation, on the other hand, then we can better understand the interplay between the different dimensions, and why we do not have sufficient resources for determining a good explanation when focusing only on one dimension in isolation of the others.

By changing our perspective from concepts of explanations to different evaluative dimensions of an explanation, we can also better clarify some confusions that have been at the heart of certain debates in the philosophy of science. Take the ongoing debate amongst mechanists in the context of biology and neuroscience. These theorists argue that explanation in the life sciences, especially biology and neuroscience, primarily proceeds by identifying the physical mechanisms responsible for the target phenomenon (often characterized in terms of organized parts and operations that produce regular change). However, some argue that mechanisms are out in the world, and that explanations merely describe or reveal such mechanisms (Machamer, Darden and Craver, 2000; Craver, 2006; Kaplan, 2011; Strevens, 2011; Craver and Darden, 2013). Meanwhile others argue that mechanisms are interpretations of the world (by way of mechanistic *models*), and that we should think of mechanistic explanation as cognitive and representational products we create to understand the world (Chirimuuta, 2014; Bechtel, 2015; Wright, 2015; Austin, 2017). The first group argues that mechanistic explanations are *ontic*. Meanwhile, the second group argues that mechanistic explanations are *epistemic*. So how ought we to understand mechanistic explanation?

Here the dispute rests on which dimension of explanation that theorists focus on at the expense of the others. For instance, Bechtel (2015) notes that causal variables in the world responsible for a cognitive phenomenon go well beyond the structures and processes occurring within the brain, and include events in the distant past, as well as all kinds of environmental structures and processes that are not typically treated as part of the biological system. In this respect, he argues that the boundary of a cognitive mechanism is often not determined by the world itself, but by the interests of working scientists who focus on only a subset of the causal variables responsible

for a given cognitive process. Since every mechanistic explanation idealizes numerous causal variables in the world and are dependent on the interests of working scientists, this suggests that we are not merely describing the ontic joints of mechanisms in the world but interpreting the world in mechanistic terms. In his words,

> It is the scientists who impose boundaries around entities and activities in nature and impose a time scale on which their functioning is characterized. For different explanatory purposes researchers may draw these boundaries in different locations or at different time points. These choices, though, while not simply responsive to pre-existing boundaries, are not entirely arbitrary. [. . .] The networks of entities found in nature commonly exhibit small-world organization as well as being scale-free. This entails that while real-world networks are highly inter-connected, there are clusters within them that are semi-independent of the rest and productively posited to be the mechanisms responsible for specific phenomena (Bechtel, 2015, p.85).

Conversely, those who defend the ontic account note that even in such cases, there *are* "real-world networks" that cluster in ways that are semi-independent and productively posited to be mechanisms. Thus even if such models are highly idealized, they still carry explanatory content in virtue of revealing genuine ontic structures, organizations, and processes in nature that play an essential part in the production of the phenomenon.

How then should we settle such disputes? I propose that the framework presented in this paper can help untangle the confusion here. Those who advocate for *ontic* accounts of mechanistic explanation focus on the ontic dimension of explanation at the cost of the representational or communicative dimension. Meanwhile, those who argue for epistemic explanation focus on the representational and

communicative dimensions at the cost of the ontic dimension. Yet if
we consider that explanation always has a communicative, representa-
tional, and ontic dimension to it, then the two accounts don't so much
represent conflicting accounts of mechanistic explanation, as much
as highlight why we should care about, and pay attention to, different
dimensions of the same explanation.[4]

Illari (2013) similarly argues that the contrast between ontic and
epistemic views of mechanistic explanation ignores the inseparable
and intertwined nature of the various goals and norms that make up
the two positions. For instance, she argues that many of the traditional

---

[4] This account of explanation may also help to disentangle some confusions regarding
the dispute between explanatory monists, and explanatory pluralists, in the philosophy
of science. Explanatory monists argue that there is a single set of criteria for determin-
ing what counts as a good scientific explanation that applies across fields and contexts.
Meanwhile, explanatory pluralists argue that there is a plurality of types of explana-
tions in science, and that there is no single set of criteria that unities them all (e.g.
the sorts of criteria that determine a good explanation in the context of physics may
be distinct and unrelated from the sorts of criteria that determine a good explanation
in the context of biology). We may be able to help clarify some of these issues by
noting that if we focus on a single dimension of explanation, like the representational
dimension, then we seem to find a plurality of types of models that can and do count
as explanatory (e.g. statistical, mechanistic, topological, dynamical, etc). However,
if we focus more generally on the essential set of norms that all three dimensions
impose on each other, we may be able to find a general monistic set of criteria that
all explanations require. For instance, a good explanation must be understandable
to the scientific community. It must represent the phenomenon in a way that either
unifies phenomena under a set of principles, identifies constraints on its behaviour, or
makes predictions. Since a model that fails to allow us to predict what the phenomenon
will do, provides no constraints on what the phenomenon cannot do, and provides
no information as to what the various instances of the phenomenon have in common,
may strip the explanation of its ability to say anything of use to scientists. Lastly, an
explanation must identify essential ontic dependencies and regularities in nature, since
otherwise we cannot tell if our representation is genuinely identifying principles the
phenomenon obeys, or constraints on its behaviour, as opposed to mere contingent
correlations. And so explanations may turn out to be pluralistic in one regard (say, in
terms of model types), but monistic in others (say, the set of general constraints that
guide communicative, representational, and ontic practices).

virtues of explanation, such as unification, simplicity, elegance, and intelligibility, in fact depend on the *interaction* of the various goals and norms that make up both ontic and epistemic accounts of explanation. As she puts it:

> This means that what looks to us intelligible, simple and uni-fied is not a static feature of human psychology, but is affected by our empirical engagement. Newtonian action at a distance used to seem quite impossible to us; so did quantum me-chanical indeterminism, and non-locality. Physicists sincerely describe quantum mechanical equations as elegant, a claim that can generate hilarity in those unused to working with such theories. If empirical engagement continually forms what we find intelligible, simple and unified, then epistemic constraints on explanation, even the more 'psychologistic' constraints, are deeply entangled with ontic ones (2013, p.254).

In this regard, the ontic and epistemic camps are not providing distinct or incompatible views of mechanistic explanation, so much as emphasizing different essential features that a given explanation must have.

4. Implications and Concerns

On the surface, such a view may appear either obvious ("of course explanation involves all three elements!") or naïve ("It's too simplistic to assume that we can easily merge these accounts"). And so it is worth exploring the complexities and difficulties that such an account now presents us with. As noted previously, different dimensions of explanation emphasize different norms for evaluating good and bad explanations. Problems begin to arise, however, when we discover

that these norms can often conflict with one another. For instance, if we focus on the ontic dimension of explanation in isolation, it seems like the more we can identify and describe the ontic dependencies in the world, the better our explanation becomes. It is this intuition that underlies Kaplan's claim that it is a "highly plausible assumption that the more accurate and detailed the model is for a target system or phenomenon the better it explains that phenomenon, all other things being equal" (Kaplan, 2011, p.347). Yet this can run counter to some of the other norms of explanation associated with the other dimensions. For instance, adding too many accurate details to a model or theory can make it more difficult to understand, and thus makes it a worse explanation according to the norms associated with the communicative dimension of explanation.

Indeed, conflicts between the various explanatory goals of the different dimensions can be seen explicitly in scientific practice itself. Consider explanations of the action potential of the neuron in neuroscience. The Hodgkin & Huxley model of the action potential, developed in the early 1950s, mathematically characterized electrochemical features of the action potential (specifically: the ion flow of sodium and potassium channels). However, it did not provide an account of the ontic variables responsible for the production of these features. In this regard, it satisfied the explanatory goals of the *representational* dimension (by identifying constraints or patterns that the phenomenon adheres to), but not of the *ontic* dimension (by failing to denote the relevant ontic variables that actually produce these constraints and patterns). This led Hodgkin & Huxley to make inconsistent claims regarding whether their model is explanatory or not. In their original paper, they claim that their model provides "a sufficient explanation of the wide range of phenomena that have been fitted by solutions of the equations" (Hodgkin and Huxley, 1952, p.541).

Yet they also claim in the same paper that their model is in fact *not* a good explanation of the action potential, since it merely provides an "empirical description" of the phenomenon and not an account of what produces it (Hodgkin and Huxley, 1952, p.541; for the discussion, see also: Bogen, 2005; Craver, 2006).

So how then do we determine whether a model (like the Hodgkin & Huxley model) counts as a good scientific explanation or not given that the norms that govern explanation are inconsistent or contradictory across evaluative dimensions? This is a serious concern. One potential solution, recently proposed by Rice & Rohwer (2020), is to claim that a scientific model or theory need only satisfy *a sufficient number* of the norms that govern all the various dimensions of explanation to count as genuinely explanatory in a given context, as opposed to satisfying *all* of them. They claim that "many single models are sufficient to explain because the information they provide will satisfy a sufficient subset of the conditions included in the cluster concept" (Rice and Rohwer, 2020, p.1043).

The problem with this solution is that it makes the common mistake of conflating *explanation* with *representation* (i.e. the idea that explanation is always constituted by a particular model). However, as I have argued, features beyond the representational texts are also constitutive parts of an explanation. For instance, scientists have certain implicit commitments to ontic dependencies that are essential to how they construct, interpret, and draw inferences from their models despite such dependencies not being explicitly stated in the model itself. Without these commitments, the model cannot function. In this respect, the background beliefs and commitments of scientists are a component part of an explanation in addition to the particular mod-

els and theories being used (see: Bokulich, 2018; Hochstein, 2019). Similarly, how we convey models and representations to others is also an essential component to explanation:

> The history of science, meanwhile, provides rich empirical support for the claim that production of explanations serves to constitute, rather than merely communicate, "intelligibility" for a scientific discipline. Precisely because explanatory discourse inculcates particular patterns of reasoning, it functions to sculpt and subsequently perpetuate communal norms of intelligibility. In effect, explanations encode the aims and values of particular scientific communities, telling practitioners what they should want to know about the world and how they should reason to get there (Woody, 2015, p.81).

This means that a single explanation may be constituted by collections of *representations*, *social practices*, *psychological processes*, and *accurate commitments to the relevant ontic features of the world*. Likewise, determining whether an explanation is good or bad will involve an evaluation of *all* these components.

At first glance, this would appear to make the evaluation of scientific explanations a herculean affair! This idea is not without precedent however. Mantzavinos (2016) proposes an account of explanation that makes a similar claim. Mantzavinos proposes that instead of talking about explanations as entities, we should understand them as collections of inferential practices governed by the norms of the scientific community, which he calls "explanatory games". Most notably, he tells us that determining "which rules of representation guide the explanatory activities is fundamentally important for the quality of explanations generated during the game" (Mantzavinos, 2016, p.42). Here, the *representational* dimension of explanation is emphasized. However, he also notes that:

> No explanatory game can take place in a metaphysical vacuum.
> The metaphysical assumptions act as constraints on the genera-
> tion of the other rules, and belong therefore to the constitutive
> rules. The structure of the game is predicated on prior assump-
> tions concerning the way the world is and by what means it
> is explainable in principle. These rules can be implicit or ex-
> plicit and they can vary from stone-age metaphysics to highly
> refined metaphysical assumptions (2016, p.41).

Here, the *ontic* dimension is emphasized. Lastly, he tells us that explanatory practices are "a process of social interaction unfolding within the given institutional rules." (Mantzavinos, 2016, p.68). In this way, the *communicative* dimension of explanation is highlighted. For Mantzavinos, we evaluate whether an explanation is good or bad by evaluating whether all of these inferences are appropriate and in accord with the inferential rules of the explanatory game.

The problem, of course, is that we must now evaluate a huge number of variables (communicative, representational, and ontic) in order to determine whether any explanation is good or bad, which is no small feat. Yet I propose that a great deal of how we evaluate explanations in science already implicitly does this. For instance, knowing which parts of a model count as idealizations and which don't require that working scientists already have ontic commitments regarding the variables in the world that the model is deviating from, and the wrong sorts of ontic commitments are often taken to be signs of a problematic explanation. As Anya Plutynski (2013, p.472) points out:

> Good modelers are careful to be clear about when an assump-
> tion used to construct a model is deliberate simplification or
> simply false, and when it is a hypothesis supported by evi-

dence. Unfortunately, what starts as deliberate simplification may often be confused with actual hypothesis and latter reified into theory.

Thus we often evaluate good explanations in terms of whether one's ontic commitments are appropriate and being used correctly in the construction of our models. Similarly, the way that knowledge is communicated throughout the scientific community is also evaluated when determining good explanation. As Woody (2015, p.85) points out,

> Explanations must be minimally intelligible to be accepted by individuals. But explanations, especially in educational contexts, condition reasoning, and thus sculpt social norms of intelligibility. Explanations intelligible to members of one community are frequently opaque to outsiders, even other scientists.

Therefore the appropriate means of communication can be essential to whether something is adopted by the scientific community as a good explanation. Likewise, various constraints on the communicative practices we have (such as passing peer review in order to be published in scientific journals, or accepted at conferences and workshops) are also considered essential to whether something should be accepted or rejected as a good *scientific* explanation.

Like evaluating any complex system, we may have to trust that some parts are in better working order than others or look for faults only when something has gone astray. Instead of attempting to evaluate all parts of the explanation simultaneously, we focus our attention on certain features of it depending on our interests and needs, or evaluate the particular features of the explanation that we think may be problematic.

By analogy, imagine evaluating whether a car is in good working order. We might have particular concerns about a given type of car ("I hear it gets bad gas millage!", "The engine is known to frequently stall!"), and use this to gauge if the particular car we want to purchase has these problems. If not, we may conclude it is in good working order. Yet if someone comes to us and points out that the break line is cut, or that the spark plugs need replacing, or that the trunk is welded shut, then those concerns will point our attention to parts of the car we assumed were in working order, and now causes us to doubt whether the car is worth purchasing. But of course, of the thousands of things that could potentially be wrong with a car, we do not and cannot evaluate all of them simultaneously. Instead, we focus our attention on some features we consider to be most important or worrisome, and assume the rest of the car is in good working order *until given reason to doubt it*. This is because there are certain standards we assume the car must have met to be sold on the lot.

The same will apply to scientific explanations. We might question some aspects of an explanation, while assuming that others are in working order in virtue of scientific communal standards. For instance, if a paper has been published in a respected peer reviewed journal, then we might assume that it is intelligible to the relevant scientific audience. Of course, if we discover that a paper was accepted due to a case of fraud, then it would immediately cast into doubt that the paper has in fact met the appropriate standards of the communicative dimension. Similarly, if scientists working in various labs report being able to control and manipulate episodes of depression by changing the concentrations of various gut bacteria, then we may be justified in assuming that the gut microbiome is part of what explains those episodes of depression. However, if we learn that their ability to control and manipulate depressive episodes in this way was greatly

overstated, or far more limited than they let on, then it will call our attention to the ontic features of the explanation that we assumed were in good working order.

Scientists often have little reason or incentive to identify all component features of an explanation during their everyday activities. To ask whether a particular model is a good explanation for example, is not to suggest that the communicative or ontic dimensions to that explanation *do not apply*, or that *we mean something different by explanation*. It is merely the case that the other dimensions are either obvious to those involved, not points of contention, or not their primary focus. Sometimes it is the ontic component of the explanation that we may disagree about, at other times the best way to convey models to an audience. We must not confuse the fact that we can analyze explanations along different dimensions with the idea that these different dimensions correspond to distinct concepts of explanation, or denote entirely metaphysically distinct explanatory entities.

Conclusion

Traditionally, philosophers have adopted either a representational definition, communicative definition, or ontic definition of scientific explanation. This has given way more recently to a pluralistic view which argues that each concept of explanation may be scientifically appropriate in particular contexts and for particular purposes. In this paper, I have argued for an alternative view which better highlights the complex ways in which representational, communicative, and ontic features of a scientific explanation interact and depend on one another. Instead of distinct explanatory concepts, we have different evaluative dimensions along which we can analyze any instance of a scientific

explanation. The upshot of this view is that it not only better accounts for why scientific explanations work when they do, but can also help to clarify disputes within philosophy of biology and neuroscience between ontic and epistemic views of mechanistic explanation.

This new account is not without its own set of challenges, however. Evaluating whether an explanation is good or bad is substantially more difficult under this new framework. Evaluations must be done in a piecemeal way, and we may have to trust that some features of an explanation are better justified than others until given reason to think otherwise. While this makes the evaluation of explanations more difficult, it is in-line with how scientific practice in fact works.

Bibliography

Achinstein, P., 1984. The pragmatic character of explanation. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1984(2), pp.274–292. https://doi.org/10.1086/psaprocbienmeetp.1984.2. 192509.

Achinstein, P., 1983. *The Nature of Explanation*. New York ; Oxford: Oxford University Press. Available at: <http://archive.org/details/natureofexplanat0000achi> [visited on 29 September 2023].

Austin, C.J., 2017. The philosophy of biology. *Analysis*, 77(2), pp.412–432. https://doi.org/10.1093/analys/anx032.

Batterman, R.W., 2000. Multiple realizability and universality. *The British Journal for the Philosophy of Science*, 51(1), pp.115–145. https://doi.org/10.1093/bjps/51.1.115.

Batterman, R.W., 2002. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, *Oxford studies in philosophy of science*. Oxford: Oxford University Press.

Batterman, R.W. and Rice, C.C., 2014. Minimal model explanations. *Philosophy of Science*, 81(3), pp.349–376. https://doi.org/10.1086/676677.

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

Bechtel, W., 2015. Can mechanistic explanation be reconciled with scale-free constitution and dynamics? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 53, pp.84–93. https://doi.org/10.1016/j.shpsc.2015.03.006.

Bogen, J., 2005. Regularities and causality; generalizations and causal explanations. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), pp.397–420. https://doi.org/10.1016/j.shpsc.2005.03.009.

Bokulich, A., 2018. Representing and explaining: The eikonic conception of scientific explanation. *Philosophy of Science*, 85(5), pp.793–805. https://doi.org/10.1086/699693.

Braverman, M. et al., 2012. Intelligibility is necessary for scientific explanation, but accuracy may not be. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34), pp.1368–1373. Available at: <https://escholarship.org/uc/item/32m3q2sp> [visited on 29 September 2023].

Chirimuuta, M., 2014. Minimal models and canonical neural computations: the distinctness of computational explanation in neuroscience. *Synthese*, 191(2), pp.127–153. https://doi.org/10.1007/s11229-013-0369-y.

Coffa, J.A., 1974. Hempel's ambiguity. *Synthese*, 28(2), pp.141–163. https://doi.org/10.1007/BF00485232.

Craver, C.F., 2006. When mechanistic models explain. *Synthese*, 153(3), pp.355–376. https://doi.org/10.1007/s11229-006-9097-x.

Craver, C.F., 2014. The Ontic Account of Scientific Explanation. In: M.I. Kaiser, O.R. Scholz, D. Plenge and A. Hüttemann, eds. *Explanation in the Special Sciences: The Case of Biology and History*. Vol. 367, *Synthese Library*. Dordrecht: Springer Netherlands, pp.27–52. https://doi.org/10.1007/978-94-007-7563-3_2.

Craver, C.F. and Darden, L., 2013. *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago, IL: University of Chicago Press. Available at: <https://press.uchicago.edu/ucp/books/book/chicago/I/bo16123713.html> [visited on 29 September 2023].

Craver, C.F. and Kaplan, D.M., 2020. Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, 71(1), pp.287–319. https://doi.org/10.1093/bjps/axy015.

Craver, C.F. and Povich, M., 2017. The directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A*, 63, pp.31–38. https://doi.org/10.1016/j.shpsa.2017.04.005.

Douglas, H.E., 2009. Reintroducing prediction to explanation. *Philosophy of Science*, 76(4), pp.444–463. https://doi.org/10.1086/648111.

Eichner, A.S., ed., 1983. *Why Economics is Not Yet a Science*. Armonk, N.Y: M.E. Sharpe.

Elgin, C.Z., 2004. True enough. *Philosophical Issues*, 14(1), pp.113–131. https://doi.org/10.1111/j.1533-6077.2004.00023.x.

Friedman, M., 1953. *Essays in Positive Economics*. Chicago, IL: University of Chicago Press.

Gilpin, L.H. et al., 2022. *"Explanation" is Not a Technical Term: The Problem of Ambiguity in XAI*. arXiv. https://doi.org/10.48550/arXiv.2207.00007.

Hall, C. et al., 2001. The need to reintegrate the natural sciences with economics: Neoclassical economics, the dominant form of economics today, has at least three fundamental flaws from the perspective of the natural sciences, but it is possible to develop a different, biophysical basis for economics that can serve as a supplement to, or a replacement for, neoclassical economics. *BioScience*, 51(8), pp.663–673. https://doi.org/10.1641/0006-3568(2001)051[0663:TNTRTN]2.0.CO;2.

Hausman, D.M., 2008. Why look under the hood? *The Philosophy of Economics: An Anthology*. Cambridge: Cambridge University Press, pp.183–187. https://doi.org/10.1017/CBO9780511819025.012.

Heeger, D.J., 1992. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9(2), pp.181–197. https://doi.org/10.1017/S0952523800009640.

Hochstein, E., 2017. Why one model is never enough: a defense of explanatory holism. *Biology & Philosophy*, 32(6), pp.1105–1125. https://doi.org/10.1007/s10539-017-9595-x.

Hochstein, E., 2019. How metaphysical commitments shape the study of psychological mechanisms. *Theory & Psychology*, 29(5), pp.579–600. https://doi.org/10.1177/0959354319860591.

Hodgkin, A.L. and Huxley, A.F., 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4), pp.500–544. https://doi.org/10.1113/jphysiol.1952.sp004764.

Illari, P., 2013. Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, 78(S2), pp.237–255. https://doi.org/10.1007/s10670-013-9511-y.

Izadi, S., Anandakrishnan, R. and Onufriev, A.V., 2014. Building water models: A different approach. *The Journal of Physical Chemistry Letters*, 5(21), pp.3863–3871. https://doi.org/10.1021/jz501780a.

Kaplan, D.M., 2011. Explanation and description in computational neuroscience. *Synthese*, 183(3), pp.339–373. https://doi.org/10.1007/s11229-011-9970-0.

Keen, S., 2022. The appallingly bad neoclassical economics of climate change. In: B.K. Gills and J. Morgan, eds. *Economics and Climate Emergency*, *Rethinking globalizations*. London; New York: Routledge, pp.79–107.

Lipton, P., 2004. *Inference to the Best Explanation*. 2nd ed., *International library of philosophy*. London; New York: Routledge/Taylor and Francis Group.

Lombrozo, T., 2016. Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, 20(10), pp.748–759. https://doi.org/10.1016/j.tics.2016.08.001.

Machamer, P., Darden, L. and Craver, C.F., 2000. Thinking about mechanisms. *Philosophy of Science*, 67(1), pp.1–25. https://doi.org/10.1086/392759.

Mackonis, A., 2013. Inference to the best explanation, coherence and other explanatory virtues. *Synthese*, 190(6), pp.975–995. https://doi.org/10.1007/s11229-011-0054-y.

Mantzavinos, C., 2016. *Explanatory Pluralism*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316440599.

Parker, G., 1978. Searching for Mates. In: J.R. Krebs and N.B. Davies, eds. *Behavioural Ecology: An Evolutionary Approach*. Oxford: Blackwell Scientific, pp.214–244.

Plutynski, A., 2013. Cancer and the goals of integration. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 44(4), pp.466–476. https://doi.org/10.1016/j.shpsc.2013.03.019.

Potochnik, A., 2010. Explanatory independence and epistemic interdependence: A case study of the optimality approach. *The British Journal for the Philosophy of Science*, 61(1), pp.213–233. https://doi.org/10.1093/bjps/axp022.

Potochnik, A., 2015. The diverse aims of science. *Studies in History and Philosophy of Science Part A*. Special Section: Formal Epistemology and the Legacy of Logical Empiricism, 53, pp.71–80. https://doi.org/10.1016/j.shpsa.2015.05.008.

Povich, M., 2018. Minimal models and the generalized ontic conception of scientific explanation. *The British Journal for the Philosophy of Science*, 69(1), pp.117–137. https://doi.org/10.1093/bjps/axw019.

Povich, M. and Craver, C.F., 2018. Because without cause: Non-causal explanations in science and mathematics (review). *Philosophical Review*, 127(3), pp.422–426. https://doi.org/10.1215/00318108-6718870.

Rice, C., 2015. Moving beyond causes: Optimality models and scientific explanation. *Noûs*, 49(3), pp.589–615. https://doi.org/10.1111/nous.12042.

Rice, C. and Rohwer, Y., 2020. How to reconcile a unified account of explanation with explanatory diversity. *Foundations of Science*, 26(4), pp.1025–1047. https://doi.org/10.1007/s10699-019-09647-y.

Ruben, D.-H., 1990. *Explaining Explanation*. 1st ed., *The problems of philosophy: their past and present*. New York: Routledge.

Salmon, W.C., 1989. Four decades of scientific explanation. In: P. Kitcher and W.C. Salmon, eds. *Scientific Explanation*, *Minnesota studies in the philosophy of science*, 13. Minneapolis: University of Minnesota Press, pp.3–219. Available at: <http://conservancy.umn.edu/handle/11299/185700> [visited on 4 October 2023].

Strevens, M., 2011. *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.

Waskan, J., 2011. Intelligibility and the CAPE: Combatting Anti-Psychologism about Explanation. Available at: <http://philsci-archive.pitt.edu/8530/> [visited on 4 October 2023].

Waskan, J., Harmon, I., Higgins, A. and Spino, J., 2014. Three senses of 'Explanation'. In: P. Bello, M. Guarini, M. McShane and B. Scassellati, eds. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, pp.3090–3095.

Woods, J. and Rosales, A., 2010. Virtuous Distortion. In: L. Magnani, W. Carnielli and C. Pizzi, eds. *Model-Based Reasoning in Science and Technology: Abduction, Logic, and Computational Discovery*, *Studies in Computational Intelligence*. Berlin, Heidelberg: Springer, pp.3–30. https://doi.org/10.1007/978-3-642-15223-8_1.

Woody, A.I., 2015. Re-orienting discussions of scientific explanation: A functional perspective. *Studies in History and Philosophy of Science Part A*, 52, pp.79–87. https://doi.org/10.1016/j.shpsa.2015.03.005.

Wright, C., 2015. The ontic conception of scientific explanation. *Studies in History and Philosophy of Science Part A*, 54, pp.20–30. https://doi.org/10.1016/j.shpsa.2015.06.001.

Wright, C. and van Eck, D., 2018. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Ergo, an Open Access Journal of Philosophy*, 5(38). https://doi.org/10.3998/ergo.12405314.0005.038.

The ontic-epistemic debates of explanation revisited: The three-dimensional approach

Jinyeong Gim

Seoul National University

Abstract

After Wesley Salmon's causal-mechanical stance on explanation in the 1980s, the ontic-epistemic debate of scientific explanations appeared to be resolved in the philosophy of science. However, since the twenty-first century, this debate has been rekindled among philosophers who focus on mechanistic explanations. Nevertheless, its issues have evolved, necessitating scrutiny of the new trends in this debate and a comparison with the original controversy between Carl Hempel and Salmon. The primary objective of this paper is to elucidate three categorical dimensions in the ontic-epistemic debates, spanning from the original to the recent controversies. Subsequently, it will explore why the conception of explanation is linked to representations, what conditions are necessary for linguistic expressions to be explanatory, and what roles norms play in explanation. Consequently, contrary to the common stereotype, it will be argued that mechanistic explanations are more likely to be epistemic rather than ontic.

Keywords

mechanistic explanation, scientific representation, explanatory norms, Wesley Salmon.

Introduction

Wesley Salmon's initial distinction between the ontic and epistemic accounts of scientific explanation stood out as one of the most prominent differences between Hempel's and Salmon's perspectives (Salmon, 1984; 1989). In broad terms, the epistemic account posits that scientific explanation involves an inferential argument aimed at understanding or predicting natural phenomena (see Hempel and Oppenheim, 1948; Hempel, 1965). On the other hand, the ontic account asserts that scientific explanation showcases the causal structures in the world that lead to the production of explanandum phenomena. Hempel's view faced numerous counterexamples, leading to the widespread acceptance of Salmon's perspective in the philosophy of science. Salmon's causal-mechanical explanation emerged as an alternative to Hempel's covering law model.[1]

Mechanistic explanation serves as a prototypical example of the ontic account of scientific explanation in line with Salmon's ideas. However, in the twenty-first century, some advocates of mechanistic explanation have posited that it is not ontic but rather epistemic (Bechtel and Abrahamsen, 2005; Wright and Bechtel, 2007; Wright, 2012; 2015). For instance, Wright and Bechtel (2007, p.51) reject the ontic perspective, stating, "Characterizing explanation as non-epistemic is problematic insofar as explanation is through-and-through an epistemic practice of making the world more intelligible." Bechtel (2008, p.18) underscores that "Explanation is fundamentally an epistemic

---

[1] The original debate between the ontic and epistemic accounts of scientific explanation was not historically limited solely to Hempel and Salmon. Ellis (1956) initiated this kind of controversy by focusing on the difference between explanation and description. Coffa (1973) also engaged in criticizing Hempel's view (see Wright and van Eck, 2018). However, this paper concentrates on Salmon's criticisms of Hempel for the purpose of philosophical clarification.

activity performed by scientists." In contrast, Craver (2007, p.21) continues to support the ontic view, aligning with Salmon's perspective: "Mechanistic explanation is associated with fitting a phenomenon into the causal structure of the world." Wright (2012, p.376), an epistemic theorist of mechanistic explanation, notes, "Perhaps because of their common interests in causality, most New Mechanists have hitched their wagon to Wesley Salmon's ontic conception of scientific explanation." Considering Salmon's original distinction, the ongoing ontic-epistemic controversy among mechanists appears somewhat peculiar.

Interestingly, the ontic-epistemic dispute remains ongoing. A decade ago, Craver (2014, p.28) argued that the ontic view plays a normative role in distinguishing good from bad explanations. He doesn't insist on defining what an explanation is but rather on identifying the essential constraints for recognizing, discovering, and using good explanations. This issue appears distinct from past ontic-epistemic debates, such as Salmon vs. Hempel, and Craver vs. Bechtel & Wright, yet Craver still emphasizes the ontic stance (see Kaplan and Craver, 2011; Craver and Darden, 2013; Craver, 2014).

In contrast, van Eck (2015) contends that epistemic norms for identifying the causal roles of mechanisms must be satisfied before any ontic norms are applied to discover relevant causal factors of phenomena. He insists on the autonomy of epistemic norms from ontic norms, particularly in neuroscience. Sheredos (2016) supports van Eck's autonomy thesis by demonstrating that general and systematic explanations cannot be simultaneously fulfilled or simply conjoined with ontic norms, as the two epistemic norms take precedence over ontic norms. While mechanists continue to use the terms 'ontic' and 'epistemic,' their current controversies differ from the original debate between Hempel and Salmon.

As noted, the ontic-epistemic debates on mechanistic explanation encompasses three distinct yet interconnected issues. While these debates have transpired over different periods and concerning various topics, each issue overlaps. Notably, philosophers who align with an ontic position on one issue may adopt an epistemic stance on another, contributing to ongoing confusion in this debate. Despite decades since the original dispute between Hempel and Salmon, new questions continue to emerge within the philosophy of science. Consequently, reviewing these debates in light of contemporary trends is imperative. Key questions persist, such as: How do individual issues interrelate? Where do proponents of the New Mechanisms find agreement or disagreement? In what ways are mechanistic explanations connected to cognitive concepts like representation and models? Should we delineate between good and bad explanations, and if so, how do we make such distinctions? Lastly, how can we differentiate between how-possibly and how-actually explanations?

Section 2 aims to scrutinize various dimensions of the ontic-epistemic debates over the past four decades. These issues can be categorized into three dimensions: (i) A relational dimension concerning *explanatoriness* (or being explanatory) between explanans and explanandum; (ii) A conceptual dimension addressing the *nature* of explanation in science; and (iii) A normative dimension evaluating the *goodness* of explanations. The first dimension of explanatoriness can be further broken down into three sub-dimensions: (i–1) What *form* does explanation take? (i–2) What imparts explanatory *force* to the explanandum phenomenon? (i–3) How do we ascertain the *relevance* of explanans to the explanandum phenomenon? Additionally, what is the overarching purpose of our pursuit of explanation? These

sub-dimensions of the first relational dimension of explanatoriness commonly pertain to the conditions that ensure the qualifications of scientific explanations.

In each dimension of the ontic-epistemic debates, it's essential to note that the distinctions are categorical, and we don' allow degrees between the two extremes to enhance clarity in our discussions. Specifically, (i) In the dimension for explanatoriness, we consider the categorical items of causal relevance vs. logical necessity. (ii) Within the dimension addressing the nature of explanation in science, the categorical items are facts or things in the world vs. representations as to actual mechanisms. (iii) In the normative dimension evaluating the quality of explanations, the categorical items are completeness & accuracy vs. intelligibility (or generality or systematicity).

Next, I will critically evaluate the widely accepted claim among proponents of the ontic view that "Mechanistic explanations are ontic." Utilizing the dimensional framework, I will delineate various phases of the ontic perspective on mechanistic explanation. Subsequently, I will analyze individual ontic theses, cautiously suggesting that mechanistic explanations do not necessarily have to be ontic.

By reviewing the ontic-epistemic debates spanning several decades and classifying them under three dimensions, I will underscore that the epistemic nature of explanation should be the focal point in the ontic-epistemic debates. Each section will delve into questions related to the three dimensions of this debate. In Section 3, focusing on the conceptual dimension of the nature of explanation, I will explore two key questions: (1) Why is the concept of representation required when comprehending the conceptual meaning of mechanistic explanation? (2) Is the relationship between the explanation and the target mechanism a binary pair or something more? Moving to Section 4, centered on the relational dimension of explanatoriness,

I will address two additional questions: (3) What stance can we adopt toward Hempel's account from the epistemic viewpoint? (4) How can we develop an epistemic account of mechanistic explanation while avoiding many of the criticisms of Hempel? Finally, Section 5 will concentrate on the normative dimension of explanation, delving into the following questions: (5) Are ontic norms like completeness actually achievable? Should we appeal to ontic norms when evaluating explanations? (6) Must we distinguish good from bad explanations? In conclusion, contrary to stereotypical thinking, the answers to these questions will reveal that mechanistic explanations may be epistemic in each dimension.

1. Three dimensions in ontic-epistemic debate

The ontic-epistemic debate of explanation traces its roots to a disagreement on the qualifications of explanation between Hempel and Salmon. Hempel puts forth specific philosophical conditions for scientific explanation, outlining four key criteria: (i) the explanandum must be a logical consequence of the explanans; (ii) the explanans must contain general laws; (iii) the explanans must have empirical content; and (iv) the sentences constituting the explanans must be true (Hempel, 1965, pp.247–249; Hempel and Oppenheim, 1948, pp.137–138). As a logical empiricist, he distinguishes the first three conditions as a set of logical conditions of adequacy, while the last is an empirical condition of adequacy. These four conditions are integral to Hempel's conception of scientific explanation, a viewpoint that Salmon outrightly rejects, labeling it as the epistemic view.

1.1 The relational dimension: Explanatoriness

Explanatoriness means the quality or property of being explanatory. Hempel's conditions provide clues of explanatoriness, including explanatory form, relevance, force, and goal. Salmon (1984, p.84) describes Hempel's claims of explanatoriness as the *inferential* version of the epistemic conception. As we will see, Salmon (1989) enumerates counterexamples against Hempel's claims and guides us into the ontic conception of explanatoriness.

1.1.1. Explanatory form: Arguments

Hempel's first condition of scientific explanation concerns the relation between the explanans and the explanandum. He introduces logical terms to stipulate their relation, deduction, and induction. Hempel (1965, pp.335–393) suggests a form of scientific explanation, *argument*. Many philosophers, including Michael Scriven, Richard Jeffrey, Bas van Fraassen, and Peter Railton, reject this inferential conception of explanation (see Salmon, 1989 for overall discussions). Salmon also denies Hempel's idea that scientific explanation is a form of argument. One of the most potent counterexamples is Bromberger's flagpole case (and Salmon's eclipse case) (Salmon, 1989, pp.46–47). According to the deductive-nomological (DN) model, we may explain the length of a shadow cast by a flagpole using specific laws of optics, geometry, and the sun's position in the sky. However, by using the same laws, the shadow's length, and the sun's position, we can also explain the height of the flagpole. Both cases satisfy Hempel's requirements of explanatory form, but the former seems explanatory, whereas the latter does not. The so-called asymmetry of explanation implies that arguments as explanatory forms are an insufficient condition of scientific explanation. In addition, because logical forms of

argument are closely related to subsequent explanatoriness, including explanatory relevance, force, and goals, many philosophers indirectly refuse Hempel's formal idea of scientific explanation.

1.1.2. Explanatory force: Universal or probabilistic laws

For a deductive (or inductive) argument to qualify as a scientific explanation, universal (or statistical) laws should be included in the premises. If the lawful propositions are empirically true, all the deductively derived propositions are also true through valid arguments. The empirical content of the explanans is the third requirement of Hempel's account. If statements have *empirical* content supported by experimental evidence, they become true, and vice versa. The laws of nature in science are generally empirically confirmed either directly or indirectly. Kepler's laws are derived directly from vast observational data about the motions of celestial bodies in the solar system. Newton's laws are indirectly justified by deducing Kepler's laws from Newton's law of gravitation. Universal or highly probabilistic laws provide explanatory forces for the explanandum phenomena in Hempel's account.

However, Salmon points out that the requirement of universal or probabilistic laws is insufficient for scientific explanation. A typical counterinstance is the barometer case (Salmon, 1989, p.47). A sharp drop in the gauge of a barometer is highly correlated with the occurrence of a storm. By assuming a law stating that whenever the barometric pressure sharply decreases, a storm will happen, we can deductively (or highly probabilistically) infer the realization of the storm. However, the barometer's reading never explains the occurrence of the storm because atmospheric conditions in a particular region are common causes of both the drop in the barometric reading and the occurrence of the storm. A lesson from this case is that causal patterns producing the explanandum phenomena must be exhibited

whenever we explain the world. Salmon (1989, p.128) states, "According to the ontic conception, explanatory knowledge is knowledge of the causal mechanisms, and mechanisms of other types perhaps, that produce the phenomena with which we are concerned."

1.1.3. Explanatory relevance: Deductive necessity or high probability

Hempel's first condition of scientific explanation is also associated with explanatory relevance between the explanans and the explanandum. As Hempel introduces two kinds of arguments, two epistemic criteria support relevant relationships between them. In deductive arguments, either nomological or statistical, logical necessity holds a specific relation between the explanans-statements and the explanandum-statement. Logical necessity relies upon the laws of deductive logic. The deductive implications of laws under initial conditions are the very explanandum. Similarly, in inductive arguments, high probabilities guarantee a highly probable possibility for the occurrence of the explanandum phenomenon. That is, if statistical laws are close to 1, then it is certain that the explanandum phenomenon happens. Explanatory relevance can be achieved within arguments through deductive necessity and high probability.

Salmon raises questions about Hempel's two epistemic conditions. The first question regarding deductive necessity is why irrelevancies are harmless to arguments but fatal to explanation. Salmon (1984, p.93) states, "In deductive logic, irrelevant premises are pointless, but they do not undermine the validity of the argument." However, explanation requires that "*Only* considerations relevant to the explanandum be contained in the explanans" (Salmon, 1984, p.94, an emphasis original). A typical counterexample is the contraceptive pill case. The fact that a man faithfully consumes his wife's contraceptive pills does

not explain *his* failure to become pregnant (Salmon, 1989, p.50). Donald Trump could have avoided getting pregnant during the past year because he took his wife's birth control pills regularly, and it is well-known that every man who regularly takes birth control pills avoids pregnancy (Salmon, 1971, p.34). Even if men take birth control pills, the pills can never work because only women have babies, not men.

Salmon points out a similar difficulty between inductive argument and explanation. According to Carnap (1950, sec.45B), the requirement of total evidence is essential in inductive logic. Salmon (1984, p.93) says, "This requirement demands the inclusion of all relevant evidence." Inductive arguments face no harm when irrelevant evidence is included in the premises. However, if irrelevant evidence overwhelms relevant factors in the explanans, it fails to explain the explanandum phenomenon. For instance, the massive consumption of vitamin C is statistically irrelevant to alleviating the severity of common colds because vitamin C does not have a directly alleged efficacy for common colds (Salmon, 1984, p.94; 1989, p.58). That is, probabilities based on relevant factors to the explanandum phenomenon are essential in inductive arguments, too.[2]

One of the main consequences in Hempel's account is the symmetry of explanation and prediction. This symmetry thesis originates from the explanatory form of explanation in Hempel's idea. Along with Newton's laws of movements of celestial objects under suitable initial conditions, we can deduce the occurrence of the

---

[2] Further, *high* probability is not a necessary condition to achieve explanatory relevance in inductive arguments (Salmon, 1989, p.49). For example, it is noted that approximately 1/4 of all victims of latent untreated syphilis develop paresis. The attack rate from syphilis to paresis is below 50%, but the low probability contributes to explaining the occurrence of paresis. In short, deductive necessity and high probability are insufficient conditions for explanatory relevance. Besides, high probability is not a necessary condition for it.

appearance of a comet. According to Hempel, the logical form of explanation as an argument applies to scientific prediction and explanation (Hempel and Oppenheim, 1948, p.138; Hempel, 1965, p.249).[3] Salmon refers to this thesis as *nomic expectability* by saying that an explanation could be described as "*An argument to the effect that the event-to-be-explained was to be expected by virtue of the explanatory facts*" (Salmon, 1984, emphases original). Salmon (1989, pp.129–130) points out a gap between explanation and prediction with the barometer example; "The sharply falling barometric reading is a satisfactory basis for predicting a storm, but contributes in no way to the explanation of the storm. The reason is, of course, the lack of a direct causal connection. For the ontic conception, [...] There must be a suitable causal relation between the explanans and the explanandum—at least as long as we steer clear of quantum mechanical phenomena."

To summarize, the first dimension of the ontic-epistemic debate is what linguistic descriptions must achieve to be explanatory. Hempel contrasts with Salmon on the form of explanation, its power, and the relation of relevance between the explanandum and its explanans. Hempel's epistemic view holds that we can secure explanatory relevance only if we make a valid argument in which the explanans is empirically true and if the explanandum is a logical consequence of the explanans. Inferential argumentation is an essential perfor-

---

[3] "The difference between the two is of a pragmatic character. If E is given, i.e. if we know that the phenomenon described by E has occurred, and a suitable set of statements C_l, C_2, ..., C_k, L_1, L_2, ..., L_k, is provided afterwards, we speak of an explanation of the phenomenon in question. If the latter statements are given, and E is derived prior to the occurrence of the phenomenon it describes, we speak of a prediction. It may be said, therefore, that an explanation is not fully adequate unless its explanans, if taken into account in time, could have served as a basis for predicting the phenomenon under consideration. Consequently, whatever will be said in this article concerning the logical characteristics of explanation or prediction will be applicable to either, even if only one of them should be mentioned" (Hempel and Oppenheim, 1948, p.138).

mance for scientific explanation. On the contrary, Salmon's ontic view holds that we can retain explanatory relevance only if we discover causal structures in the world that give rise to an explanandum phenomenon. According to Salmon, without information about the causes of phenomena under empirical inquiries, a scientific explanation cannot avoid many counterexamples that Hempel's epistemic view confronted. The first dimension concerning explanatoriness between Hempel and Salmon is a categorical classification between logical arguments and causal mechanisms.

1.2 The conceptual dimension: Nature of explanation

After the late twentieth century, Salmon's ontic view has been widely accepted or discussed in the philosophy of science. Recently, many philosophers of science who focus on mechanisms to understand biology and biochemistry are also based on Salmon's ontic view. As Salmon emphasizes exhibitions of causal structures, most mechanists believe that mechanisms result in explanandum phenomena, so discovering mechanisms is an essential scientific practice.

Understanding mechanisms is a prerequisite condition for mechanistic explanation. There is no disagreement about the main features of a mechanism (Craver and Darden, 2013, pp.15–26). In short, a description of a biological mechanism serves as an explanation for a phenomenon presumed to be the outcome brought about by the mechanism. Thus, a mechanistic explanation is also a binary relation between a mechanism and its behavior, which needs to be explained; consequently, explanatory relevance in mechanistic explanation can be achieved when we discover the mechanism that brings about the behavior of the mechanism.

Recall that the first dimension of explanatoriness is a criterion to find sufficient conditions that all scientific explanations must sat-

isfy. Notably, most mechanists are not interested in defending such a sufficiency thesis, but they assume that mechanistic explanation is an explanatorily relevant scientific explanation. Instead, they began to dispute the ontic-epistemic debate differently. The second dimension of the debate is about the nature of explanation relating to mechanistic inquiries. According to ontic theorists, including Craver, a biological mechanism explains its regular phenomenon (Machamer, Darden and Craver, 2000; Craver, 2007; Craver and Darden, 2013). For instance, the mechanism of protein synthesis explains how linear sequences of amino acids can be synthesized from DNA (see Darden, 2006). The mechanism of neuronal depolarization explains how neural signals are transmitted from one neuron to another. These biological mechanisms are realizers of the phenomenon. To explain the phenomenon is to discover the mechanism. This idea is widely accepted in the New Mechanism because the descriptions of the mechanism are descriptions of the causes for the phenomenal occurrence. As Salmon defends that explanations are objective features of the world, Craver (2007, p.200) also "Advocates an ontic view of explanation according to which one explains a phenomenon by showing how it is situated in the causal structure of the world." In other words, ontic theorists think that mechanistic explanations are causal structures of the world.[4]

However, discoveries of biological mechanisms are inevitably dependent on artificial experiments or observations under ideally stipulated conditions and manipulations. Additionally, mechanisms in nature are never observed directly. Whenever scientists employ

---

[4] When Salmon (1984, p.127) states that "the exhibition of causal mechanisms is an essential part of scientific explanation," the term "exhibition" does not refer to an active agent's cognitive outcome, such as an argument, but rather a passive result produced by causal processes and interactions in the physical world (see Salmon, 1984, p.132). In line with Salmon's causal thinking, ontic theorists, including Craver, believe that mechanistic explanations are the mechanisms themselves.

experiments to find causal factors of phenomena, idealizations are methodological presumptions in laboratories. For this reason, knowledge about a biological mechanism is not identical to the mechanism itself; thus, all linguistic or diagrammatic descriptions can be regarded as a representation of them (Bechtel and Abrahamsen, 2005; Bechtel, 2008). Bechtel holds that "Explaining is still an epistemic activity" and that "The mechanism in nature does not directly perform the explanatory work" (Bechtel and Abrahamsen, 2005, p.425). "Explanation is fundamentally an epistemic activity performed by scientists" (Bechtel, 2008, p.18). Similarly, Wright (2012) also emphasizes human cognitive performances to make an explanation in science by saying that "Explanations are not produced until an explainer contributes her or his epistemic and/or cognitive labor." Hence, epistemic theorists think that explanations are human cognitive processes.

Although Bechtel and his colleagues stress that explanations (and even mechanistic explanations) are products from cognitive processes, not things in the world, Craver still adheres to an ontic view:

> The task is to develop an account of scientific explanation that makes sense of the scientific project of connecting our models to structures that can be discovered through experience and objective tests. In domains of science that concern themselves with the search for causes and mechanisms, this amounts to the idea that the norms of explanation fall out of a commitment by scientists to describe as accurately and completely as possible the relevant ontic structures in the world (Craver, 2014, p.48).

Craver acknowledges that we cannot but depend on laboratory epistemic procedures by constructing various models. However, only some models as to mechanisms can show a genuine causal structure to bring about an explanandum phenomenon. For instance, Hodgkin and Huxley's model of action potential in neuroscience helps predict all phases

of electrical pulses in a neuron. However, this model cannot provide how individual ion channels within a neuron membrane produce the shape of potential, how they temporally interact, what electromagnetic activities occur, etc. Ptolemaic models of the solar system allow us to predict the motion of planets, but they cannot provide what causes the motions of planets at all. Craver (2006, p.367) argues that models can be explanatory only when they describe mechanisms in the world. Mechanistic schemata, not mechanism sketches, are worthy of explaining phenomena.

To summarize, the second debate within the New Mechanism is controversial about the conception of mechanistic explanation. Ontic theorists hold that causal structures existing in the world explain phenomena. Epistemic theorists hold that all procedures to exhibit causal structures are cognitive, so explanations are epistemic.

1.3 The normative dimension: Explanatory norms

The third dimension of the ontic-epistemic debate is about explanatory norms or constraints to evaluate explanations. Craver (2007, p.21) argues, "Good explanations in neuroscience show how phenomena are situated within the causal structure of the world." Shortly speaking, "Good explanations explain effects with causes" (Craver, 2007, p.26).[5] While Craver insists on the ontic constraints, he never argues that we should not cast our glance upon epistemic constraints. Craver (2014, p.28) says, "I do not claim that one can satisfy all of the normative criteria on explanatory models, texts, or communicative acts

---

[5] Craver (2007, p.26) enumerates five norms that good explanatory texts reveal the causal structure of the world: (E1) mere temporal sequences are not explanatory (temporal sequences); (E2) causes explain effects and not vice versa (asymmetry); (E3) causally independently effects of everyday causes do not explain one another (common cause); (E4) causally irrelevant phenomena are not explanatory (relevance); and (E5) causes need not make effects probable to explain them (improbable effects).

by focusing on ontic explanations alone. There are questions about how one ought to draw diagrams, organize lectures, and build elegant and useable models that cannot be answered by appeal to the ontic structures themselves." That is, Craver never denies the existence of epistemic norms. However, he emphasizes the 'demarcation' between good and bad explanations by ontic norms rather than epistemic (Craver, 2014, p.27).

Craver's emphasis on the ontic nature of mechanistic explanation in the conceptual dimension is closely associated with the normative dimension of explanation. Craver believes that one of the essential requirements for a successful mechanistic explanation is a successful representation of mechanistic models to target mechanisms. Craver (2006, p.360) introduces the notion of "Ideally complete descriptions of the mechanism," which are models including all the entities, properties, activities, and organizational features relevant to every dimension of the phenomenon to be explained. In addition, the target mechanisms should be responsible for the occurrences of the phenomenon to be explained in the world. Similarly, Kaplan and Craver (2011, p.611) suggest a model-to-mechanism-mapping (3M) requirement that satisfies the causal commitment with an ontic point of view:

> In successful explanatory models in cognitive and systems neuroscience, (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism.

Whereas Craver explicitly advocates ontic norms of mechanistic explanation, Bechtel and Wright do not. It is unclear whether Bechtel

and Wright conversely defend epistemic constraints, but they are just concerned about such norms: "Explaining refers to a ratiocinative practice governed by certain norms" (Wright and Bechtel, 2007, p.51). Because of such implicit commitment, Wright and Bechtel do not complain about Craver's ontic normativity. That is, among Craver, Bechtel, and Wright, there seems to be no disagreement on the normative dimension of mechanistic explanation, such as Craver's idea, "The ontic structure of the world thus makes an ineliminable contribution to our thinking about the goodness and badness of explanations" (Craver, 2014, p.41).

Instead of Bechtel and Wright, other philosophers criticize Craver's ontic view in the normative dimension (van Eck, 2015; Sheredos, 2016). First, van Eck (2015) focuses on experimentation in neurosciences: "Achieving the ontic aim, as well as the epistemic aim of understanding, associated with mechanistic explanation, therefore, is deeply *constrained* by epistemic considerations" (van Eck, 2015, emphasis added). Van Eck mentions two reasons to support epistemic constraints: (i) explaining extinct mechanisms cannot be possible under ontic constraints; (ii) identifying performance to discover causal factors of an explanandum phenomenon is governed by epistemic norms before ontic norms are applied. Thus, van Eck argues epistemic norms have more explanatory power than ontic norms. Sheredos follows van Eck's basic idea but further focuses on the normative priority of explanation between ontic and epistemic norms. He defends the priority of epistemic norms by mentioning two epistemic norms: generality and systematicity. The generality of a mechanistic explanation is determined by its scope of application. The more an explanation applies to other cases, the more general it is. If the mechanistic explanation is general, its systematicity is explicit because it demands a principle of unity or extrapolation that unifies similar cases under

a limited scope (Sheredos, 2016, p.932). He says that the two epistemic norms are prior to Kaplan and Craver's 3M norms by arguing that (i) fulfilling the norm of generality and systematicity is logically independent of fulfilling 3M norms, and (ii) ontic norms ontologically depend on generality and systematicity.

Contrary to radical defenders of the epistemic norms, Illari, Kästner & Haueis regard ontic and epistemic constraints as equally important. As noted, Bechtel and Wright (2007), who emphasize the epistemic nature of mechanistic explanation in the conceptual dimension, acknowledge that exhibiting real mechanisms is essential to examine mechanistic explanations. Bechtel and Wright are involved in this debate only at the conceptual dimension and do not seem to have yet taken a clear position on the normative dimension. Illari (2013) was the first philosopher who tried to integrate ontic and epistemic views in the normative dimension. She argues that "good mechanistic explanations must satisfy both ontic and epistemic normative constraints on what is a good explanation" (Illari, 2013, p.237). That is because "we will often not be in a position to know whether a particular mechanistic explanation is there to satisfy epistemic constraints alone, or ontic, or both" (Illari, 2013, p.254). The ontic conception concerns mechanisms themselves in the world, whereas the epistemic conception concerns our knowledge-building practices regarding the mechanisms. In more detail, ontic aims of mechanistic explanation are to "describe the entities and activities and the organization by which they produce the phenomenon or phenomena" (Illari, 2013, p.250), and epistemic aims of mechanistic explanation are to "build a model of the activities, entities and their organization that scientists can understand, model, manipulate, and communicate, so that it is suitable for the ongoing process of knowledge-gathering in the sciences" (Illari, 2013, p.250). Thus, Illari says, "the real achievement of

mechanistic (and possibly other forms of) explanation is satisfying both ontic and epistemic constraints simultaneously, to get a story constrained by all the empirical contact with the world that ingenuity can design" (Illari, 2013, p.253).

Kästner and Haueis (2021, p.1636) more clearly argue that "ontic and epistemic norms are equally important in mechanism discovery." (Kästner and Haueis, 2021, p.1637) subsume mechanism discovery and mechanistic explanation under mechanistic inquiry and show that "an adequate account of mechanistic inquiry should not only distinguish ontic and epistemic norms but also make explicit how they work together in practice." That is because both types of norms are interdependent. Kästner and Haueis (2021, p.1658) say, "practitioners must share the commitment that (1) they search for the entities and activities responsible for the phenomenon in question and (2) that their epistemic activities make intelligible how the mechanism is responsible for the phenomenon to be explained" to represent patterns in the world successfully. Illari, Kästner & Haueis agree that ontic and epistemic norms are equivalently essential for building mechanistic explanations.

To summarize, the third dimension of the ontic-epistemic debate also occurs within the New Mechanism. Craver, an ontic theorist, holds that mapping the relationship between representations and mechanisms is essential to making good explanations. Epistemic theorists hold that ontic norms are subordinate to two kinds of epistemic norms, so epistemic constraints have priority over ontic constraints. Other philosophers pursue unifying ontic and epistemic norms because both kinds of norms interact in pattern recognition of mechanisms.

2. On the nature of mechanistic explanation:
Things vs. representations

In the previous section, I chronologically enumerated three dimensions of the ontic-epistemic debates of explanation. As an ontic theorist, Salmon is engaged with all dimensions by suggesting the distinction of ontic-epistemic conceptions of explanation. Fundamentally, he asserts that scientific explanations are things in the world to acquire objective explanations. Salmon categorizes Hempel's covering law account of explanation as the inferential version of the epistemic conception, contrasting it with his causal-mechanical explanation as the ontic conception. He identifies two philosophical demarcation problems when emphasizing the ontic conception (see Salmon, 1984). One is distinguishing scientific explanations from other scientific achievements related to the first dimension of explanatoriness. The other is demarcating good from bad explanations, related to the third dimension of explanatory normativity, which will be discussed in Section 5. Salmon contends that the ontic conception of explanation provides a critical standard for addressing these two problems and advocates for it by implicitly revealing that Hempel's inferential version of the epistemic conception fails to fulfill these philosophical requirements.

From now on, I will delve into each dimension of explanation, focusing on mechanistic explanations. Salmon's ontic stance leads many philosophers of science to assume that mechanistic explanations are inherently ontic (see Craver, 2007; Kaplan and Craver, 2011; Povich, 2018). However, Bechtel and Wright argue that mechanistic explanations stem from cognitive efforts to uncover mechanistic patterns, even though they acknowledge mechanistic explanations as scientific. Given the three-dimensional distinctions of the ontic-epistemic debates, my initial exploration will assess whether mechanistic expla-

nations qualify as "things." This inquiry poses a significant challenge to ontic theorists, who persist in their belief that Bechtel and his colleagues lead us astray in the realm of mechanistic explanation (see Craver, 2014). In the first conceptual dimension, I will advocate for the epistemic stance on mechanistic explanation by demonstrating that the central ontic thesis yields counterintuitive results.

2.1 Explaining as representing mechanisms

The conceptual dimension of explanation delves into the very nature of explanation itself, addressing the fundamental question: "What is explanation?" The response typically takes the form of "Explanation is something like *X*." This line of inquiry is distinct from asking, "How is the explanandum phenomenon related to the explanans?" The latter question can be further subdivided to explore the various forms of explanation, identify what imparts explanatory power to the explanandum phenomenon, and ascertain how to determine the relevance of the explanans to the phenomenon. Salmon's causal-mechanical account of explanation provides a set of answers to these questions, shaping the initial debate between Salmon and Hempel around the nature of explanation. Despite Salmon presenting numerous counterexamples to Hempel's model, the distinction he draws between ontic-epistemic *conceptions* of explanation can be perplexing, as he seems to primarily address the question related to the ontic *account* of being explanatory rather than the ontic *nature* of explanation itself. The conceptual dimension of explanation is about the semantic meaning of explanation. When Salmon emphasizes the ontic conception of explanation, his ontic view must be an idea of the nature of explanation in the conceptual dimension. The dimensional approach to ontic-epistemic debates aids in distinguishing among various issues within the controversies. As

noted, Salmon's response to the question in the conceptual dimension of the debate is, "Explanation is a thing in the world producing the explanandum phenomenon."

Wright and Bechtel are pioneers in hightlighting the distinction between the account and the conception of explanation.[6] Wright and Bechtel (2007, p.73) argue that "Explanation is inherently an epistemic or cognitive activity" by focusing on "How investigators reason with the models and representations of mechanisms." Their view contrasts with the ontic perspective, where the "Explanans just is a mechanism or parts is taken literally" (Wright and Bechtel, 2007, p.49). There are compelling reasons to abandon the ontic view of the nature of mechanistic explanation in the conceptual dimension independent of the first relational dimension of explanation. First, mechanistic explanation relies on mechanistic *inquiries* to search for relevant components of the target mechanism and their organizational features. These inquiries encompass all scientific investigations aiming at discovering mechanisms and developing mechanistic explanations. All agential acts revealing the target mechanism's relevant parts, operations, and organizations "minimally need to be captured and codified in a structural or functional *representation* of some sort"

---

[6] Bokulich (2016, p.263) also emphasizes distinguishing the account from the conception of explanation. An account of explanation is a view of *how explanations work*. On the other hand, a conception of explanation is a view of *what explanations are*. Interestingly, Bokulich is in line with Wright and Bechtel's position when we look at the following Bokulich's mention: "Salmon endorses both the *ontic conception* of explanation and the *causal account* of explanation. One can, however, reject the ontic conception of explanation (i.e., deny that explanations are things in the world, independent of human theorizing), but endorse the view that many explanations are indeed causal (i.e., involve citing and representing the relevant subsets of causes of the phenomenon)" (ibid, *emphases original*).

(Wright and Bechtel, 2007, p.51).[7] Bechtel and Richardson (2010) suggest methodological strategies, decomposition, and localization as epistemic constraints of mechanistic inquiries.

Wright and Bechtel's epistemic view of the nature of mechanistic explanation is grounded in the idea that explaining is an agential activity of human beings. Wright and Bechtel (2007, p.51) emphasize that the mere existence of mechanisms does not suffice to construct an explanation because the presence of a phenomenon is not explained "until a cognizer contributes his or her explanatory labor." There is a limited possibility that all non-living creatures can explain natural phenomena. To explain a phenomenon, one must determine the initial and terminal conditions. Subsequently, relevant parts and their activities are identified by choosing adequate instruments or methods. Determining and choosing are decision-making activities for explaining the phenomenon. Mechanistic explanations arise from from *goal-oriented* inquiries of humans who seek mechanisms. If one embraces the epistemic nature of mechanistic explanation, it becomes apparent that the ontic conception leads to an absurd consequence: only humans, among all external materials, can explain phenomena.

Mechanisms are non-cognitive systems in the world. If one believes that a mechanism itself explains a phenomenon, one cannot help but confront another counterintuition. For instance, all explanations are potentially misleading and/or incomplete because explanation is a cognitive activity. Under the ontic view, by contrast, an erroneous explanation implies the non-existence of the mechanism. Probably, explanation errors originate from the objective reality of the target mechanism and practical activities through mechanistic inquiries. Of

---

[7] The dependence of mechanistic explanation upon epistemic activities is widely recognized (see Kästner and Haueis, 2021, p.1637). This point will be discussed in Section 5 again.

course, if there is no mechanism to produce the phenomenon, all mechanistic explanations are undoubtedly erroneous. However, the linkage between the failure of explanation and the non-existence of the mechanism also leads to an absurd consequence: no mechanisms exist in the world whenever we identify irrelevant parts or non-organized features. The synthesis of proteins exists independently of whether molecular biologists explain how the primary sequence of amino acids is produced. The action potential in a nerve cell occurs regardless of whether physiologists explain how voltage-gated ion channels control the flow of ions across a membrane. If we adopt the possibility of providing erroneous and/or incomplete explanations as to mechanisms, then the epistemic view seems more promising than the ontic view. As we will see, failures of explanation originate from misrepresentations. Wright and Bechtel (2007, p.51, fn.10) also say, "Erroneous explanation would not be possible since there would have been no mechanism for the erroneous explanation to have identified."

Nevertheless, ontic theorists still argue that mechanistic explanation must be a thing or fact in the world, not a representation. Bechtel and Wright point out a metaphysical confusion of the ontic theorists, which is called the *de re* & *de dicto* confusion. Wright and Bechtel (2007, p.51) say that "mechanisms themselves are not the sorts of things can be constituents of any explanans." Constructing explanations requires the identification of relevant parts and comprehension of organizational structures. Identifying and comprehending are cognitive activities. Mechanisms *de re* are passive or inefficient in conducting both activities. Epistemic outcomes, either linguistic or diagrammatic, are *de dicto* mechanisms employed to explain the explanandum phenomena. For this reason, Wright and Bechtel (2007, p.51) say, "Explaining refers to *ratiocinative practice* that cognizers

engage in to make the world more intelligible." Thus, the ontic asser-tion of the nature of mechanistic explanation must be understood as *metonymic*, not realistic (*ibid*).

To summarize, it is crucial to distinguish the account of explana-tory relationships between the explanandum phenomenon and ex-planans from the conception of explanation itself. Salmon primarily focuses on the former when critically analyzing Hempel's account of scientific explanation. The first relational dimension of the ontic-epistemic debate pertains to this issue, whereas the second conceptual dimension is linked to the latter. Regrettably, Salmon occasionally uses the term *conception* when differentiating his ontic theory of explanation from Hempel's inferential version of the epistemic the-ory of explanation. Using the term *account* rather than *conception* was more suitable when Salmon criticizes Hempel's covering law model. As pointed out by Wright, Bechtel, and Bokulich, the asser-tion that non-agential things outside of our mind explain something appears absurd, as explaining itself is a form of cognitive activity. Mechanistic explanations are also outcomes dependent upon human research inquiries regarding resources such as components and their organizations within a mechanism.

2.2 Mechanistic explanation and scientific representation

If a mechanistic explanation is an outcome of epistemic activities, representation arises as a central concept for understanding the nature of mechanistic explanation. The contents of mechanistic explana-tions emerge from scientific endeavors to represent mechanisms. Let me further develop an epistemic view of mechanistic explanation by considering the concept of representation. Generally, representa-tion is equated with presenting a likeness or resemblance to what is

represented. Representation as a resemblance involves a binary relation between the represeter and what is represented. I think the ontic theorists implicitly adopt this binary relationship when discussing a relation between the explanandum phenomenon and explanans. Craver and Kaplan's 3M criteria is a typical case of this binary idea. The mapping relations between a mechanistic model and a mechanism are based on the resemblance relation independent of epistemic agents or human activities.

Note that we need further reasons to adopt the epistemic position of explanation in the conceptual dimension, even though explanations inherently involve a cognitive activity. That is because the binary relation of representations allows for the non-subjective interpretation of a mechanistic explanation. To adopt an epistemic position in the conceptual dimension, wherein explaining is representing, representations should transcend a binary relation based solely on resemblance. Instead, they should be characterized by a tertiary relation, enriched by agential performances. The question then becomes: Is a representation inherently binary, or can it involve more complex relationships?

To advance the epistemic understanding of explanation in the conceptual dimension, I will draw upon the philosophical discourse on *scientific representation* in the philosophy of science. Specifically, I will introduce Bas van Fraassen's ideas to contemplate the nature of mechanistic explanation.[8] Van Fraassen distinguishes representations-

---

[8] Before doing this, I simply talk about whether linguistic representations, such as statements or propositions, are suitable forms of mechanistic explanation. A statement basically consists of a subject (or the referent) and a predicate (or its property). Interestingly, Bechtel and Abrahamsen argue that there are reasons to prefer diagrams as representations of mechanisms rather than linguistic descriptions. Diagrams help in depicting spatiotemporal information. Arrows in a diagram represent sequential orders of stages within a mechanism. Dots represent variables or individual objects. Pictorial illustrations help in visualizing spatial structures among objects. Bechtel and Abrahamsen (2005, p.429) do not deny the possible usage of linguistic descriptions,

as from representation-of. **Representation-as** is a binary relation between the represener and what is represented and is formulated by "*X* represents *Y* as *F*," where *X* is linguistic or pictorial things to represent, *Y* is the target object or referent represented, *F* is a predicate indicating the object's properties. For example, in the case of linguistic expressions, "Snow is white" represents snow as colored (van Fraassen, 2008, p.16). Similarly, the phrase "DNA is double-helical" linguistically represents DNA as having a double-helical shape. In the realm of non-linguistic expressions, a famous photograph of the Eiffel Tower, *Au Pont de l'Alma* by Doisneau, represents the Eiffel Tower in Paris. If we consider Rosald Franklin's B-form picture of DNA, we can assert that Franklin's B-form picture non-linguistically represents DNA as being double-helical. Despite the different modes of representation for DNA, whether linguistic or non-linguistic, they share a commonality in having a binary relation with DNA.

Contrary to the binary relationship of representations as a resemblance between a source and a target, van Fraassen defines representations by adding the indexical feature to scientific representation.

> What is represented, and how it is represented, is not determined by the colors, lines, or shapes, in the representing object alone. Whether or not *A* represents *B*, and whether or not it represents the represented item as *C*, depends largely, and sometimes only, *on the way in which A is being used* (van Fraassen, 2008, emphases original).

---

but Bechtel and his colleagues explicitly emphasize that diagrammatic representations are preferable forms (see Bechtel and Abrahamsen, 2005, p.432; Wright and Bechtel, 2007, p.52; Bechtel and Richardson, 2010, p.xix). However, I do not discuss the privileged form of mechanisms between linguistic and diagrammatic representations. Rather, I presume that both forms of representation have their own advantages to depict the mechanisms independent of whether we take Hempel's model of explanation. Both linguistic and diagrammatic descriptions simultaneously fill in most textbooks in biological sciences.

According to van Fraassen, 'use' assimilates the terms 'make' and 'take.' For instance, when an English person linguistically describes the shape of DNA as double-helical with the statement "DNA is double-helical," Watson and Crick take Rosalind Franklin's B-form X-ray figure as evidence to denote physical properties of DNA, including a 20-angstrom diameter, 34-angstrom height, and a 3.4 interval length between two bases. According to van Fraassen (2008, p.23), "*Use*, in the appropriate sense, must determine the selection of likenesses and unlikenesses which may, in their different ways, play a role in determining what the thing is a representation of, and how it represents that" (*emphasis original*). Based on the intentional situation, **representation-of** refers to the tertiary relation as "Someone, *Z*, uses *X* to represent *Y of F* in a context, *C*." The use of representation is relative to the intentional activities of agents.[9] In pragmatics, van Fraassen also stresses the perspectival indexicality. For instance, "An apple is red" depicts the apple as red, while a *normal-eyesighted* person represents the snow *under the sun*. A red-blind person cannot represent the apple in normal light circumstances like that sentence. A normal-eyesight person cannot represent the apple in the darkroom. From the pragmatic view of scientific representation, a representation-of results from an intentional activity. In short, "*There is no representation except in the sense that some things are used, made, or taken, to represent some things as thus or so*" (van Fraassen, 2008, emphases original).

How does van Fraassen's view of scientific representation align with the epistemic view of the nature of mechanistic explanation? In the epistemic view of the conceptual dimension, a mechanistic explanation arises from cognitive activities that involve representing

---

[9] Similarly, Ronald Giere (2006, p.60) defines scientific representations as such "*S* uses *X* to represent *W* for purposes *P*."

relevant parts and their organized features of mechanisms. Mechanistic inquiries, aiming at representing mechanisms, are permeated with cognitive decision-making and choices.

Firstly, the objective is to explain a phenomenon by characterizing it, and considering which instruments are suitable for understanding its properties. This characterization determines the aspects that an agent explains. For instance, the mechanistic model of protein synthesis varies depending on the topics of the explanandum. James Watson, a molecular biologist, focused on the genetic flow from DNA to protein. At the same time, Paul Zamenick, a biochemist, was interested in enzymatic reactions leading to ATP synthesis, a process of energy production preceding the amino acid sequence synthesis (see Judson, 2013, part 2). Other biochemists, such as Francois Jacob and Jacques Monod, explored the regulation of the β-galactosidase gene expression when lactose is present (see Pardee, Jacob and Monod, 1959). Additionally, numerous molecular biologists conducted various experiments to study specific aspects of the protein synthesis process according to their interests.

Watson and Francis Crick proposed a diagram illustrating the flow of genetic information from DNA to protein (see Fig. 1). Zamecnik and Mahlon Hoagland discovered a crucial intermediate molecule, transfer RNA (tRNA), which carries free amino acids into the process of synthesizing polypeptides. Furthermore, Robert Holley et al. (1965) successfully isolated tRNA and analyzed its structure (see Fig. 2). Jacob and Monod identified key components, including the regulator gene, promoter, operator, and lac operon (structural genes), interacting with each other in synthesizing messenger RNA (see Fig. 3). Hence, the characteristics of the explanandum phenomenon determine the scope, extent, and relevant entities of the mechanistic explanation.
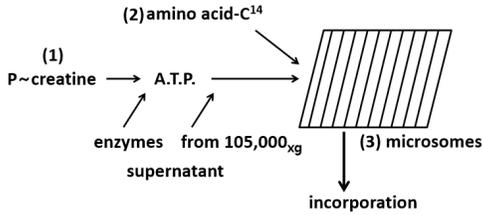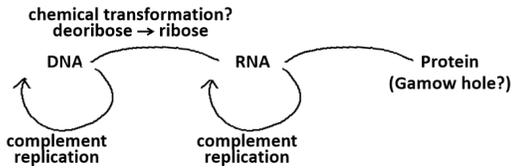
Figure 1: Zamecnik and Watson's diagrams (based on Judson, 2013, p.273)
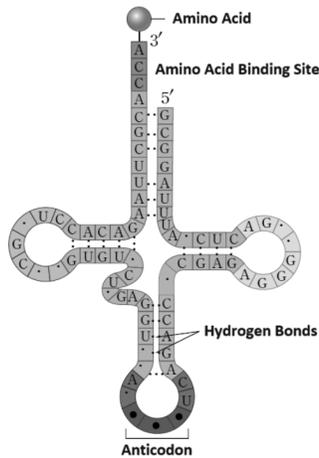

Figure 2: The two-dimensional representations of tRNA (based on Holley et al., 1965, p.1464)
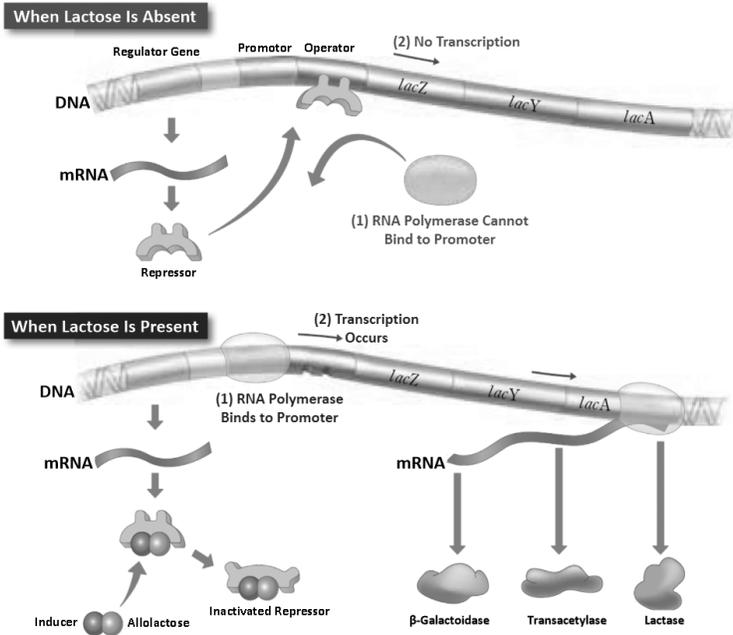
Figure 3: Models of the regulation of protein synthesis (based on Jacob and Monod, 1961, p.344)[10]

---

[10]The regulation of gene expression in prokaryotes transpires at the transcriptional step. Multiple genes with analogous functions may be conglomerated into a singular transcriptional unit, thereby co-regulating their expression. This collective entity is denoted as an *operon*, comprising a promoter, an operator region, and structural genes. The promoter serves as the site where RNA polymerase binds, initiating the transcription process. Simultaneously, the operator region, where a repressor protein is implicated in the regulation of transcription, spatially overlaps with the promoter, impeding RNA polymerase binding when the repressor protein is present. The structural genes encapsulate information pertaining to the proteins of three enzymes—β-galactosidase, transacetylase, and lactase—indispensable for lactose utilization. These three enzyme genes are collectively regulated within a singular operon, commonly known as the lactose operon. Expression of the lactose operon remains quiescent in the absence of lactose as a carbon source. Furthermore, even in the presence of lactose, expression
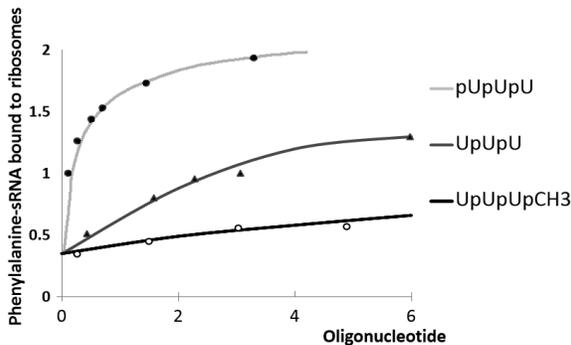
Figure 4: Effect of poly-U nucleotides with terminal phosphate on activities of synthesizing Phenylalanine (based on Rottman and Nirenberg, 1966, p.562)

After determining the phenomenon to be explained, researchers hypothesize what objects are relevant to the phenomenon. They either formulate hypotheses about the relations between candidate objects or conduct experiments to identify whether the phenomenon occurs due to chemical interactions between objects. Watson, in 1954, and Crick, in 1958, proposed the RNA template hypothesis between DNA and protein based on existing results at that time. Zamecnik conjectured that a ribonucleoprotein particle, later named a 'ribosome,' within the microsome is engaged with protein synthesis (Hoagland et al., 1958). The most famous experiment in molecular biology, the so-called PaJaMo experiment, performed by Arthur Pardee, Jacob,

---

ensues solely when a more favorable carbon source, such as glucose, is unavailable. The regulatory process unfolds sequentially: (1) In the absence of lactose, a repressor protein binds to the operator region, impeding transcription within the lactose operon. (2) Conversely, in the presence of lactose, the lactose inducer, allolactose, binds to the repressor protein, instigating a conformational alteration that precludes its binding to the operator region, thereby facilitating transcription within the lactose operon (for more details see Jacob and Monod, 1961).

and Monod (1959), implies DNA itself is not a direct template of protein synthesis. Other biologists, including Severo Ochoa, Marshall Nirenberg, and Har Gobind Khorana, contributed to explaining how genetic information is transferred from one molecule to another.[11]

Those theoretical or experimental results play a crucial role in explaining the phenomenon of protein formation in distinct ways. Watson and Crick's RNA template hypothesis represents the genetic flow from DNA to protein via messenger RNA. This hypothesis is instrumental in explaining how DNA's genetic information is transferred to proteins. Zamecnik's diagram delineates the biochemical flow for protein synthesis utilizing the formation of ATP. This diagram proves useful in explaining the energetic components required in protein synthesis (see Fig. 1). Holley's two-dimensional scheme from 1965 represents the structural shape of alanine RNA. This schematic representation is valuable in explaining how free amino acids are transported into the formation of polypeptides (see Fig. 2). Jacob and Monod's models from 1961 illustrating the regulation of protein synthesis represent the interaction between regulator, operator, and structural genes. These models are effective tools in explaining how different genes contribute to synthesizing transcribed RNA (see Fig. 3). Nirenberg's model of data shows that 5'-terminal phosphate poly-uracil sequences are more active in synthesizing phenylalanine linkages than in other sequence cases. This model is utilized to explain the structural features of mRNA required to form polypeptide linkages (see Fig. 4).

---

[11] Ochoa was a Nobel Prize winner in 1959 for their discovery of the mechanisms in the biological synthesis of ribonucleic acid. Nirenberg and Khorana were also Nobel Prize winners in 1968 for their interpretation of the genetic code and its function in protein synthesis.

It's important to note that researchers characterize the phenomenon of protein synthesis based on their individual interests. Molecular biologists, for instance, concentrate on genetic mappings from DNA to protein, while biochemists explore different facets of the phenomenon, such as energetic aspects (Zamecnik), structural analysis (Holley), enzymatic processes (Ochoa, Jacob & Monod), and reaction rates (Nirenberg). Their research interests guide the representation of relevant parts of the protein synthesis mechanism. Watson and Crick, for example, speculated on the existence of messenger RNA between DNA and protein, while other biochemists employed various experimental instruments and interpreted data models. Some outcomes represent temporal steps of the mechanism, while others focus on the structural features of specific components. These diverse results prove valuable in explaining how proteins are formed in a cell. In essence, most representations serve as worthy explanans for mechanistic explanations when used to depict the physical features of candidate parts, revealing their structural shape and considering them as relevant components. In a nutshell, "A researcher *uses* linguistic and diagrammatic *representations* of relevant parts and their organizational features of a *mechanism* to *explain* a phenomenon depending on the researcher's interests."

Certainly, it is crucial to closely identify the relevant parts and their organizational features that are intricately linked to the occurrence of the explanandum phenomenon. However, the judgments regarding whether certain parts or organizational features are pertinent to the occurrence of the explanandum phenomenon are not issues within the conceptual dimension of explanation but rather fall under the relational dimension of explanatoriness. Here, I emphasize that explaining a phenomenon is conceptually identical to representing something that produces the phenomenon, and that representations are

not binary but tertiary relations involving linguistic or non-linguistic representations, target entities or their organizations in the world, and human agents. These two emphases form the foundation for subsequent discussions about the relational and normative dimensions of the ontic-epistemic debates.

3. On the explanatoriness of mechanistic explanation: Sufficiency vs. necessity

As noted, Salmon has led us to consider mechanistic explanation as an alternative to Hempel's covering law model, presenting it as the ontic conception of explanation, in contrast to the epistemically inferential nature of the latter. In the original ontic-epistemic debate, it is well-known that Hempel's model faces numerous counterexamples, as detailed in Section 2. Consequently, many proponents of the New Mechanism, including Bechtel, argue that mechanistic explanation naturally falls into the ontic realm within the relational dimension of explanation.

However, caution is warranted because the prioritization of the ontic conception of explanation against Hempel's viewpoint is situated within the dimension of explanatoriness. I previously underscored the importance of representation in understanding the nature of explanation within the conceptual dimension. Herein, several questions arise. If we adopt an epistemic position in the conceptual dimension of the nature of explanation, are we compelled to align with Hempel's viewpoint in the dimension of explanatoriness once again? If not, does that imply a rejection of Hempel's ideas? Suppose we choose not to adhere to Hempel's inferential version of the epistemic account of explanatoriness. How can we maintain a non-ontic position of expla-

nation within the dimension of explanatoriness? Furthermore, how compatible are the two epistemic positions in the two dimensions— explanatoriness and the nature of explanation—with each other?

One of the most fatal shortcomings in Hempel's model is that his conditions for scientific explanation are insufficient in the relational dimension of explanatoriness. For instance, the flagpole or eclipse case serves as a counterexample to the condition of the explanatory form, which is an argument. The barometer case functions as a counterexample to the condition of the explanatory force, whether it be universal or probabilistic laws. The contraceptive pill case serves as a counterexample to the explanatory relevance condition, which is a logical necessity. Similarly, the vitamin C case is a counterexample to the condition of explanatory relevance based on high probability. These counterexamples demonstrate that arguments, laws of nature, logical necessity, and high probability alone are insufficient for scientific explanations. I pursue a way to adopt the epistemic position of *mechanistic* explanation in the relational dimension without falling into the swamp that Hempel's model confronted.

My primary strategy to advocate the epistemic view of mechanistic explanation does not involve searching for other ideal conditions of scientific *explanation*, as Hempel did. Instead, I will regard Hempel's covering law model as a pursuit of *necessary* conditions for scientific *representation*, assuming that mechanistic explanations are representations relative to pragmatic interests. In other words, Hempel's four conditions are necessary for the format of representations of mechanisms in the context of "An *agent* uses *linguistic representations* to depict *mechanisms* to *explain* the explanandum-phenomenon." If some descriptions of a mechanism fail to satisfy Hempel's conditions, then the linguistic representations of the mechanism are also deemed inadequate.

Let me begin with a question: Are all linguistic representations, such as statements, propositions, and descriptions, inadequate for depicting mechanisms? Bechtel and Abrahamsen (2005) seem to prioritize non-linguistic representations of mechanistic explanations. However, I do not intend to discuss the priority of representational forms here. Instead, I presume that linguistic and diagrammatic representations are equally essential modes of mechanistic explanation. Although many mechanistic explanations are diagrammatic, they never completely replace all linguistic expressions. Instead, linguistic descriptions support us in illustrating spatiotemporal processes of mechanisms, understanding detailed information, and communicating with diagrams. Most philosophers of science who advocate the New Mechanism believe that diagrammatic representations are primary modes of mechanistic models. Still, I will concentrate on linguistic representations of mechanisms.

3.1 Validity as a necessary condition of linguistic representations

As to the first condition about explanatory form, I suggest adopting Hempel's first condition for linguistic representations of mechanisms, not for qualifying or eligible provisos of scientific explanation. I never intend to advocate Hempel's model as an account of scientific explanation. I am in alliance with the mainstream to criticize Hempel's covering law model. However, I will deal with Hempel's conditions of *linguistic representations* to satisfy them. My starting point was the dimension of the nature of mechanistic explanation. If we seriously take the epistemic conception of the nature of mechanistic explanation being representations, I think Hempel's conditions help satisfy the necessary conditions for linguistic statements. In other words, all statements must be true, and an argument must be deductively valid.

In the case of protein synthesis, the ontic theorists may believe that facts, including transcription from DNA to mRNA and subsequent translation from mRNA to protein, explain how a protein is synthesized from genetic materials. However, there seems to be a wide gap between the belief that an actual mechanism exists and the observational evidence needed to justify that belief. No biologists directly observe the entire process from DNA to protein via mRNA. Strictly speaking, the mechanism of protein synthesis is unobservable. All processes of protein synthesis are inevitably decomposed and discovered separately in different laboratories. The results of experiments conducted on protein synthesis in each laboratory can be briefly summarized *verbally*. For instance, Roger Kornberg observed a specific case of transcription: an RNA polymerase synthesizes an mRNA sequence from DNA in a eukaryotic cell. Marshall Nirenberg and Heinrich Matthaei observed a particular case of translation such that a typical complex of ribosomes and additional components synthesizes a polypeptide containing only phenylalanine from a poly-uracil sequence that is artificially composed by sole uracil. We can abstractly imagine that Kornberg's mRNA and Nirenberg's poly-uracil sequence are theoretically equivalent because both are a single strand of nucleotide synthesized by an RNA polymerase in principle within a typical environment. Under this theoretical condition, the two observations can be integrated into a productively continuous process from DNA to protein via mRNA. Of course, diagrammatic representations are more intuitive to figure out detailed occurrences within both processes, but linguistic expressions are also possible as follows:

C_1. Macromolecules, including DNA, mRNA, and protein, exist in a eukaryotic cell.

C_2. Enzymes, including RNA polymerase and a complex of ribosomes, also exist in eukaryotic cells.

P_1. An RNA polymerase synthesizes a single nucleotide (mRNA) strand from DNA.

P_2. A complex of ribosomes synthesizes a primary sequence of proteins from mRNA.

C_3. mRNA's physical properties are always invariant, independent of concrete syntheses.

C_4. A transcribed mRNA is abstractly identical to a template of translation in principle.

E. The product of proteins is produced from DNA via mRNA.

The two conditions C_1 and C_2 are true on the ground that Watson & Crick's model of DNA, Wilkins & Franklin's X-ray diffraction evidence, and Kornberg & Nirenberg's experimental models. Crick's central dogma, DNA $\rightarrow$ mRNA $\rightarrow$ Protein, is filled with objects. The two premises, P_1 and P_2, are also true in Kornberg and Nirenberg's research. That is, arrows in Crick's central dogma are filled with enzymes. The additional conditions, C_3 and C_4, are essential to integrate separate transcription and translation. These additional conditions are not ontic but epistemic claims because they are based on abstractly regular patterns from empirical research findings. Without direct observations of the mechanism of protein synthesis, the explanandum-phenomenon, E, is a deductive consequence of initial conditions and premises. In other words, a deductive structure between statements is essential for linguistic representations to be useful in explaining phenomena.

Note that the mechanistic explanation above is linguistic. Premises are expressed lawfully and include empirical contents. All statements must be true either empirically or in principle. Due to two

conditions C_3 and C_4, the conclusion, E, can be derived from C_1, C_2, P_1, and P_2. Be cautious that I never argue that the mechanistic explanation satisfies Hempel's covering law model. I intend to show that mechanistic explanation can be expressed linguistically as well as diagrammatically, and that if expressed linguistically, it must suffice Hempel's condition of explanatory form. Validity is a fundamental requirement of linguistic representations to be an argument. I never deal with valid arguments as a sufficient condition of scientific explanation. My focus is not on finding sufficient conditions of scientific explanation but on finding the necessary conditions of linguistic representations. In other words, Hempel's logical condition is a *necessary instruction* of linguistic representations, not a scientific explanation. If linguistic representations are not valid, then they are absurd and non-informative in the end.

3.2 Laws as a necessary condition of understanding for explanatory force

The ontic theorists claim that universal laws (or highly probabilistic laws) are insufficient conditions for scientific explanation. I agree. However, I emphasize that laws of nature are *necessary* to understand how activities (or component operations) are engaged with entities (or component parts) and how they are spatiotemporally organized.

According to the ontic view, only things and facts are explanatory. In the case of protein synthesis, DNA, mRNA, and proteins are relevant entities. Moreover, transcription from DNA to mRNA means that the genetic information of DNA is transmitted into that of mRNA. Translation from mRNA to protein is also a factual process that determines the types of amino acids and their order based on the genetic information of mRNA. In Crick's central dogma, all pertinent objects (DNA, mRNA, protein) are discovered, and two sequential transitions

Type	Activities
Geometrical	Shaping
Mechanical	Fitting, Colliding, Pushing, Pulling, Opening, Closing
Electrical	Attracting, Repelling
Chemical	Bonding, Breaking
Energetic	Thermodynamic
Electro-magnetic	Electrically Conducting

Table 1: Types of activities.

are also identified empirically. A question arises: Can we understand, with only those objective facts, how entities act in reactions, how activities occur, and how components are organized?

Most advocates of the New Mechanism, whether they are ontic or epistemic, believe that activities (or component operations) give explanatory forces. According to Machamer, Darden, and Craver (2000, p.14), there are four types of activities: (i) geometrical-mechanical, (ii) electro-chemical, (iii) energetic, and (iv) electromagnetic. I will classify such activities in Table 1.
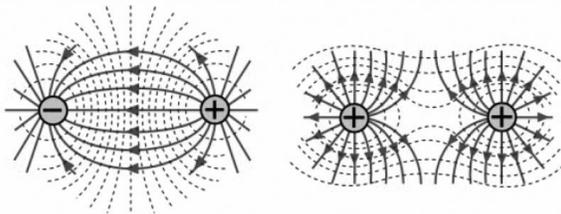

Figure 5: Attracting and repelling

As noted in Table 1, activities are referred to as metaphorical terms. However, I think that metaphorical terms are not fundamental foundations for providing explanatory forces because they must be elucidated under lower-level scientific theories. I suggest metaphorical terms gain explanatory grounds when anchored in scientific facts, laws, or theories.

For example, electrical terms are based on Coulomb's law.[12] Assume that atoms are the fundamental entity in our discussion.[13] Fig. 5 illustrates the explanatory ground of the electrical activities. The electrical field vector $\overrightarrow{E}(\overrightarrow{r})$ is represented by lines pointing in the direction of the electric field at any point; that is, the electric field vector is tangent to the electric field lines at any point. The number of lines per unit area passing through a surface perpendicular to the lines is proportional to the strength of the electric field in a given region. For a positive point charge, the lines radiate outward from a point, whereas for a negative point charge, the lines converge inward. The number of field lines leaving the positive charge equals the number of lines terminating at the negative charge. Fig. 5 illustrates the electric field lines for either two positive point charges or two negative point

---

[12] The potential energy, $\overrightarrow{E}(\overrightarrow{r})$, of a test charge Q is the Coulombic energy of interaction between Q and an arbitrary charge q separated by a distance $\overrightarrow{r}$ such that $\overrightarrow{F}(\overrightarrow{r}) = Q\overrightarrow{E}(\overrightarrow{r}) = Q\frac{q}{4\pi\varepsilon_0 \overrightarrow{r}}$, where $\overrightarrow{E}(\overrightarrow{r})$ is the electric force and ε_0 is the permittivity constant of vacuum.

[13] This assumption is based on the atomic structure. All atoms consist of a nucleus and one or more electrons. A nucleus is electrically positive, whereas electrons are negative, so electric activities occur between them. There is attraction between the nucleus and the electrons, while there is repulsion between electrons. Repulsion also occurs between the nuclei of one atom and those of another atom adjacent to it. Electric activities are starting points to formalize entities and other activities because all materials contain a nucleus and electrons.

charges. The attraction and repulsion activities, describing the electrical interaction between charges, are mathematically formulated by Gauss' law.[14]

What about chemical activities? Those activities are also based on laws and theoretical backgrounds in physics. The two activities based on Coulomb's and Gauss's laws play a fundamental role in constructing molecules with interactions between the nucleus and the electrons of atoms. The second step in constructing molecules is to examine the *chemical* activity relating to how atoms combine in a molecule. A molecule comprises more than two atoms that share electrons. Covalent bonding is the most potent chemical bond when atoms share electrons. Molecular orbital theory (MO theory) is a widely accepted theory to explain how covalent bonding occurs. Further, the Schrödinger equation can be solved analytically only for H^+_2, but the solution cannot be applied to hetero-polyatomic molecules. In MO theory, the linear combination of atomic orbitals (LCAO) is an algebraic method to calculate the overall wavefunctions of electrons within a molecule.[15]

---

[14] The direction of electric flux is from a positive point charge to a negative point charge. In other words, electric flux, Φ_E, is a measure of the electric field vectors penetrating a given surface. It is proportional to the number of field lines that pass through a given region, $\overrightarrow{A}$, oriented perpendicular to the field, which is represented by the following equation: $\oint_S \overrightarrow{E} \cdot d\overrightarrow{A} = \Phi_E = \overrightarrow{E}\,\overrightarrow{A}\cos\theta$, where a vector perpendicular to the region, $\overrightarrow{A}$, is at an angle θ concerning the field. The equation represents Gauss's law, $\nabla \cdot \overrightarrow{E} = \Phi_E = \frac{Q}{\varepsilon_0}$. Gauss's law implies that the electric flux through any closed surface is equal to the net charge inside the surface Q divided by ε_0. Gauss' law is a fundamental ground for the electrical activities.

[15] An atomic orbital is a wave function that describes the behavior of an electron within an atom. Molecular orbitals can be calculated using the available atomic orbitals within various chemical contexts. (i) Electrons supplied by the atoms are accommodated in the orbitals to achieve the lowest overall energy, adhering to the constraints of the Pauli exclusion principle, which states that no more than two electrons may occupy a single orbital, and they must be paired. (ii) If several degenerate molecular orbitals are available, electrons are added singly to each orbital before doubly occupying

Hence, laws and theoretical backgrounds provide a scientific understanding of why metaphorical terms are explanatory. I do not argue that universal laws are a sufficient condition of scientific explanation. Instead, I emphasize the necessity of laws for an intellectually ultimate understanding of the explanandum phenomenon.

3.3 Counterfactual inference as a necessary condition of explanatory relevance

Logical necessity is not a sufficient condition for explanation. Even if an argument concludes the occurrence of a phenomenon from sentences about relevant components and their organizational features is valid, the presence of a sentence about an entity being causally inefficacious to the occurrence of the phenomenon in the premises renders the argument not a mechanistic explanation. Moreover, no matter how much an inductive argument is based on high probability in the case of most prokaryotic cells, there is no guarantee that the argument will immediately apply to the case of eukaryotic cells. For this reason, Hempel's two conditions of explanatory relevance are deemed insufficient.

Assuming that explanations are representations, I emphasize that Hempel's conditions should be considered necessary but insufficient. As noted earlier, logical necessity is crucial when explaining a phenomenon through a set of linguistic descriptions. It serves as a prereq-

---

any one orbital because that minimizes electron-electron repulsions. (iii) Hund's rule implies that if two electrons occupy different degenerate orbitals, a lower energy is obtained if they do so with parallel spins. In short, the structure of a molecule is determined by the configuration of electrons within the molecules under the minimal energetic state. All these descriptions of physicochemical backgrounds are prepared with reference to Atkins and de Paula's Physical Chemistry textbooks (see Atkins, De Paula and Friedman, 2014).

uisite for the explanatory form rather than a condition guaranteeing explanatory relevance. Within Hempel's account of explanation, there are no adequate conditions for mechanistic explanation.

Alternatively, I introduce an inference to affirm the relationship of constitutive relevance between the explanandum phenomenon and descriptions of parts and organizational features: *counterfactual conditional*.[16] All phenomena to be ontologically explained depend on their lower-level mechanisms. This ontological dependence can be linguistically reformulated by a 'because' complex sentence. To ensure explanatory relevance in the relational dimension, it must be demonstrated that the statement 'The description of the phenomenon clause (P) is true because a set of descriptions of mechanism clauses (M) is true.' The truth conditions of this sentence are provided by the counterfactual condition, stating that 'P-clause because M-clauses' is true if and only if: (i) 'P-clause' is true, (ii) all 'M-clauses' are true, and (iii) 'If it were not the case that M, it would not be the case that P' is true.[17]

A sentence 'P because M' indicates the explanatory relevance of mechanistic explanation between the explanandum phenomenon and

---

[16] In most cases of biological mechanisms, descriptions of the explanandum phenomenon are at a higher level than descriptions of component parts and their activities. Advocates of the New Mechanism, particularly Craver, regard explanatory relevance in mechanistic explanation as *constitutive* relevance. Although the philosophical debate remains, we will tentatively assume here that the two concepts of relevance are identical.

[17] This kind of formal investigation of the explanatory relationship based on the 'because' complex sentence has been slightly shown in Wright and Bechtel's paper in 2006. However, they did not formulate their idea at all. I will adopt the counterfactual truth conditions of the 'because' sentence as an epistemic condition of counterfactual inference for explanatory relevance. They said, "Descriptions of mechanisms are not just coincident with, or derivative from, explanations—they *are* explanations. But explanations are not merely lists of descriptions of mechanisms or sets thereof: they include inferential and simulatory operations on them. (Considerations of the semantics of the explanatory connective 'because,' as well as what it is that arrow in

explanans, assuming that every phenomenon (P-clause) depends on its mechanisms (M-clauses). Suppose a description within the P-clause is false. In that case, all claims of explanatory relevance are false, as the first condition is not satisfied, regardless of the truth of the descriptions of mechanisms. It highlights that characterizing the explanandum phenomenon by determining its initial and terminal conditions is the primary cognitive activity of explanation. Similarly, if a description within the mechanism clauses is false, the 'because' sentence also becomes false, as the second condition is not satisfied, irrespective of the truth of the phenomenon description. These considerations presuppose that mechanistic explanation is a cognitive activity in the first conceptual dimension of explanation. It is important to note that some explanations can be false due to misrepresentations, not the non-existence of relevant components and their organizational features.

The third truth condition is crucial in judging the relevance between P and M descriptions. Let's assume that M is a description of a candidate for the relevant part of a mechanism. If the occurrence of the P-description is observed even when it is not the case that M (that is, $\sim$M), then it is certain that the referent in M is irrelevant. The counterfactual conditional "If it were not M, it would have been P" is false since the P-description is true even if the M-description is false. For instance, from the molecular biologists' perspective, let me first discuss a linguistic description of M-clause: "X is a genetic intermediate between DNA and protein" (M_1). Suppose a biologist makes M_1 by filling the description with ribosomal RNA within a microsome. In that case, ribosomal RNA is not relevant because it has been proven

---

box-and-arrow diagrams represent, help in grasping this point.)" (Wright and Bechtel, 2007, emphasis original). I will develop their basic idea by pursuing counterfactual conditions of the 'because' sentence.

that protein synthesis occurs even if M_1 is false. On the other hand, suppose another biologist makes M_1 by filling the description with messenger RNA. In this case, the counterfactual conditional "If it were not M_1, it would have been not P" is true since protein synthesis cannot be realized without mRNA.

It should be noted that we are not suggesting here that ribosomal RNA is an entirely incorrect component in protein synthesis. Rather, we found that ribosomal RNA was not an appropriate referent in the mechanistic description, "*X* is a genetic intermediate between DNA and protein." Ribosomal RNA is the correct referent for other mechanistic descriptions, such as "*X* synthesizes polypeptides based on messenger RNA sequence" (M_2). In other words, it is not realistic to determine whether a single referent is correct or incorrect for an entire mechanism. A referent may be correct for a description in one context and may be incorrect for other descriptions in another context.[18]

Additionally, from the biochemists' point of view, let me focus on enzymatic activities, particularly RNA polymerase. The counterfactual inference is well applicable to cases of step-by-step processes. An RNA polymerase is a key entity synthesizing mRNA with DNA. When explaining the production of mRNA, ontic theorists may provide a mechanistic explanation as follows. Suppose that an explanandum-phenomenon, *P*, is the production of mRNA sequence through transcription. The emergence of mRNA is the terminal out-

---

[18] Reversely, the occurrence of the P-description may not be observed even if a mechanism description is true. In this case, the counterfactual condition "If it were M, it would have been ∼P" is false so that the mechanism description includes irrelevant information. Here, all true mechanism descriptions can be proven as totally irrelevant to the phenomenon because all referents and their properties in M-clauses never contribute to generate the occurrence of the phenomenon. As a result, we easily ignore all such mechanism descriptions.

come of the mechanism of RNA synthesis. The initial condition of the production is nothing of any ribonucleic acids in a cell's nucleus. However, we must assume that there are a lot of building units of the mRNA, including adenine, uracil, guanine, and cytosine. A M-clause is "mRNA sequence emerges in a cell's nucleus." The truth condition of the M-clause could be determined by empirical manipulation like centrifugal identification. In the 1960s, the existence of mRNA was proven by Brenner et al. What are the M-clauses as explanans of the P-clause? Briefly speaking, there are five statements of the M-clauses as follows:

M_3. An RNA polymerase attaches a DNA region called the promoter sequence.

M_4. The RNA polymerase unwinds the two strands of DNA.

M_5. The RNA polymerase moves along the DNA.

M_6. At that time, the RNA polymerase synthesizes individual RNA nucleotides to the growing mRNA strand based on a single strand of DNA template.

M_7. The RNA polymerase detaches from the DNA template when stopping the synthesis.

If M_6 is not true, the P-description, "A transcription is generated," may also not be true since M_6 implies the P-description. M_6 temporally depends on M_5, which also depends on M_4, which depends on M_3. M_7 indicates the end of the series from M_3 to M_6. As a result, if a sequence of successive M-descriptions is false, then the P-description as a consequence of the terminal M-description is also false. That is, if M_3–M_7 is not true, then P is not true, too. In conclusion, M-descriptions, M_3–M_7, are explanatory relevance to the P-description. An inference with the counterfactual conditional is worthy of being an adequate epistemic condition to discern whether any mechanistic descriptions are relevant to the phenomenon description.

Finally, I do not argue the structural identity between mechanistic explanation and prediction. We do never adopt the demarcation problem because prediction is an explanatory virtue. Hempel's four conditions are insufficient for scientific explanations to satisfy but necessary to evaluate whether linguistic representations are adequate. No matter how logical necessity (or high probabilistic necessity) is satisfied from the premises and their conclusion, it does not guarantee that linguistic representations are scientific explanations. However, if all statements are employed to explain something, they must be logically consistent. If not so, independently of whether logical necessity is a sufficient condition to determine the eligibility of scientific explanation, all linguistic representations are not informative. Moreover, laws of nature are not sufficient conditions for scientific explanations. However, in the absence of laws of nature, it is difficult to identify the explanandum phenomenon and further understand how organizational features of mechanisms contribute to producing the occurrence of the phenomenon.

Hempel's account is epistemic because epistemic inferences, such as logical calculus or derivations, establish explanatory relevance. In contrast, Salmon's causal-mechanical account is ontic because explanatory relevance is established not by mental inferences based on logic but by actual target systems in nature, expected to bring about a phenomenon to be explained. Attempts to evade Hempel's shadow were unsuccessful in establishing the epistemic view in the relational dimension of explanatoriness in the original ontic-epistemic debate.

To advocate the epistemic conception of mechanistic explanation, I propose two epistemic ways to illuminate the explanatory relevance between the explanandum phenomenon and its explanans: the necessity thesis of linguistic representations and the counterfactual inference. Hempel's epistemic conditions are necessary for mechanis-

tic explanations, which are linguistic outcomes representing actual mechanisms under the assumption that linguistic descriptions express the contents of mechanistic explanation. Additionally, a counterfactual inference of the 'because' sentence helps check whether the relation between the explanandum and explanans is relevant. I illustrate this with molecular and biochemical cases. These alternative interpretations of Hempel's account of explanation in the relational dimension of explanatoriness are based on the previous consequence that the nature of explanation is a cognitive activity in the conceptual dimension.

4. On the norms of mechanistic explanation: Completeness vs. purpose-dependence

Most philosophers have discussed the normative dimension by emphasizing the superiority or interaction between ontic and epistemic norms (Kaplan and Craver, 2011; Illari, 2013; van Eck, 2015; Sheredos, 2016). In contrast, I will focus in this Section on these two types of norms as criteria for evaluating mechanistic explanations. Many ontic theorists defend ontic norms by mentioning that ontic norms are the most powerful standard for evaluating explanations (see Craver, 2014; Povich, 2018). Participating in the ontic-epistemic debate as a standard for evaluating mechanism explanations is noteworthy since it is an issue of the normative dimension of explanation that has yet to receive much attention among epistemic advocates. However, I would like to point out that, contrary to the wishes of ontic theorists, when they use ontic norms when evaluating mechanistic explanations, they encounter dilemmas that run counter to our common sense.

4.1 Pessimistic induction and demarcation of explanation

If we adopt the idea that mechanistic explanation is a cognitive achievement to represent a mechanism linguistically or diagrammatically in the conceptual dimension, what implications does it have for normative issues? Generally, mechanists refer to representations of mechanisms as mechanism schemes. Assuming this equivalence of abstract two terms, we can say that Craver emphasizes two ontic norms of mechanistic explanation: (i) completeness and (ii) correctness. He, with Darden, defines the two terms as critical criteria to evaluate mechanism schemes (Craver and Darden, 2013, p.9): "Completeness presumes that there is a complete target mechanism in the world, and one can assess the extent to which the schema includes all and only the entities, activities, and organizational features in the target. Correctness presumes that there is a fully instantiated target mechanism in the world, and one can assess the degree of fit or mapping between items in the scheme and items in the target mechanism." As noted previously, completeness is a chief ontic goal of mechanistic explanation. Correctness is closely related to Craver & Kaplan's 3M requirement between a scheme and its target mechanism.

Note that most ontic theorists, including Craver, Kaplan, and Povich, further employ those ontic norms to demarcate good from bad explanations. Ontic theorists seem to believe that a philosophical search for any criteria to evaluate the quality of explanation is a critical task of the philosophers of science. They think of mechanism sketches (or how-possibly models) as mere conjectures of the target. On the other hand, they appear confident that we can attain correct and complete representations of the entire target mechanism as how-actually models. They tend to regard mathematical descriptions of natural patterns as trifling or partial explanations. Hodgkin and Hux-

ley's equations of action potential are the typical case. No matter how the equations allude to regular patterns across a membrane of a neuron quantitively, ontic theorists underestimate the explanatory values of mathematical descriptions (see Craver, 2007). They stress the significance of only qualitative descriptions (or diagrams) with causal terms within Table 1. They seem to believe that referring to relevant components and their organizational patterns is the superior prerequisite of mechanistic explanation than depicting them mathematically.

Let's delve into ontic norms by addressing the following questions: Is the ontic norm of completeness attainable? How can we determine whether completeness is achieved without prior knowledge of the target mechanism? These inquiries pose metaphysical doubts about ontic norms. Moreover, most mechanism schemes are inevitably labeled as bad explanations when completeness is employed as a criterion for evaluation. The question then arises: Should we rely on completeness to distinguish between good and bad explanations?

Firstly, is the ontic norm of completeness applicable to the evaluation of hypothetical mechanism schemes? According to Craver & Darden's definition of completeness, an agent must possess complete information about a target mechanism before employing this ontic norm. This comprehensive information includes the types of entities, activities, and their organizational features. However, completeness is an unattainable criterion when applied to practical contexts. Information about the target must be known in advance to use the completeness norm in evaluation. Yet, as demonstrated earlier, direct access to a target mechanism is not feasible. We cannot grasp information about mechanisms all at once, and scientific instruments do not aid in perceiving the entire process of the explanandum phenomenon. Instead, we must partially grasp mosaic information about the mecha-

nism and then piece the information together.[19] Completeness is not
a tangible concept that can be achieved in one fell swoop. Instead, it
is a variable concept that gradually changes in magnitude and degree
as the obtained evidence increases.

Second, completeness is not an appropriate criterion for evaluat-
ing whether an explanatory representation is good or bad. This ontic
constraint is a variable concept depending on how far the research
has progressed. Is Crick's central dogma, DNA $\rightarrow$ RNA $\rightarrow$ Protein,
a mechanism scheme? Most ontic theorists seem to answer yes be-
cause the missing link was filled with RNA. Further, this scheme
opens the way to explain how proteins are synthesized from DNA.
However, this mechanism scheme seems to be a bad explanation under
the completeness criterion because the scheme does not include which
enzymes act when RNA is transcribed from DNA and when protein
is translated from RNA. Perhaps ontic theorists believe a complete
explanation can be achieved by enumerating the complex processes
by which RNA and protein are synthesized separately, using causal
terms. Or they may believe that a complete explanation can be more
easily achieved if the two syntheses are presented concisely, using
diagrams.

Perhaps ontic theorists would say this is on the way to complete
a mechanism description. Of course, pursuing completeness by adding
unknown sub-processes in more detail than the previous mechanism
descriptions is not problematic. The genuine problem arises from the
fact that proposing a mechanistic explanation depends on additional
research findings. More novel components will be continuously dis-
covered in pursuit of a complete mechanism description. And new

---

[19] Craver emphasizes the mosaic unity in neuroscience for a different reason (see
Craver, 2007). He argues that mechanistic explanations in neuroscience are not reduced
to the fundamental level but are unified by multilevel results of different fields.

interactions between the components will also emerge. The additional achievements of these discoveries are typical of biological research, which is entirely natural. But do biologists always classify their explanations as good or bad based on the yardstick of completeness? That rarely happens. All of their explanations are on a *continuum*, and it is highly unrealistic to divide the sequence of achievements into good and bad explanations solely based on completeness.

Let's reconsider the case of protein synthesis, particularly examining whether the mechanism schema in Fig. 6 serves as a complete and correct model to represent the full processes of protein synthesis. Regrettably, the schema is inherently incomplete. Firstly, the discovery of reverse transcriptase in 1970 unveiled a reverse flow of genetic information from an RNA virus to a DNA provirus. Secondly, eukaryotic mRNA is not immediately ready for translation. RNA at this stage is termed precursor-mRNA and must undergo additional processing before transitioning from the nucleus to the cytoplasm as mature mRNA. These processes involve RNA capping, polyadenylation, and splicing, modifying mRNA in various ways. These modifications enable a single gene to produce more than one protein.[20] Third, the full

---

[20] RNA capping modifies the 5' end of the RNA transcript, which is the end that is synthesized first. RNA is capped by the addition of an atypical nucleotide, which is a guanine nucleotide containing a methyl group attached to the 5' end of RNA in an unusual way. This capping occurs after RNA polymerase II has produced about 25 RNA nucleotides long before completing transcription of the entire gene. Polyadenylation provides newly transcribed mRNA with a special structure at the 3' end. Unlike bacteria, where the 3' end of an mRNA is simply the end of a chain synthesized by RNA polymerase, the 3' end of a eukaryotic mRNA is first trimmed by an enzyme that cleaves the RNA chain at a specific sequence of nucleotides. The transcript is then finished by a second enzyme that adds a series of repeating adenine nucleotides to the cleavage ends. These poly-A tails are typically hundreds of nucleotides long. Splicing removes introns from precursor-mRNAs. Introns are unnecessary regions for encoding proteins. The rest of the mRNA consists only of regions called exons that are needed to synthesize the protein. Editing is the process of changing some nucleotides in mRNA. For example, a human protein called APOB, which helps transport lipids in the blood, has two different forms due to editing. One form is smaller than the other

mRNA sequence after transcription is not the genetic code for synthe-
sizing proteins. Precursor RNAs synthesized by an RNA polymerase
in the nucleus must be edited. Particularly, the splicing mechanism
in eukaryotes is essential; non-genetic portions within the precursor
RNAs, the so-called introns, must be deleted, and genetic portions,
the so-called exons, must be ligated. Splicing implies that both syn-
thesizing processes, transcription, and translation, are not the only
processes within protein synthesis. In other words, Fig. 6 is never
a complete model of protein synthesis.
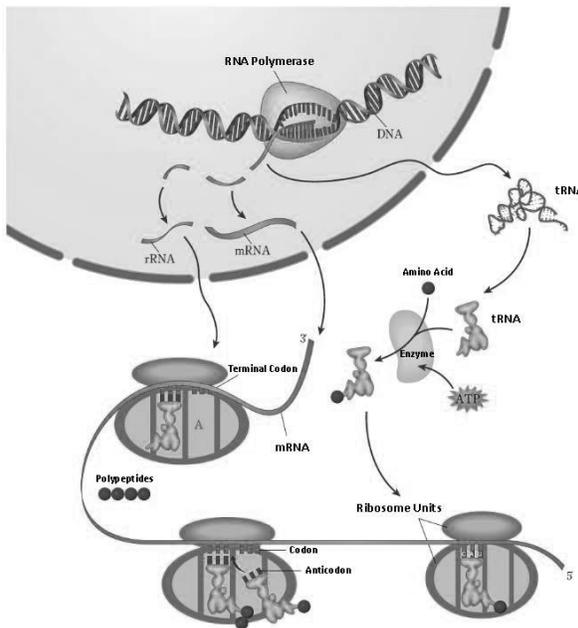
Figure 6: Mechanistic schema for protein synthesis

because editing adds an earlier stop signal to the mRNA. RNA editing processes such
as these occur all the time in real cells. In order for the mechanism by which RNA is
synthesized from DNA to meet the criteria for completeness, an explanation including
these processes is required.

Maybe ontic theorists will complain that those two cases, retrovirus and splicing reactions, do not refute their ontic norms to evaluate mechanism schemes because they are not counterexamples but evidence of their ontic criteria. The more processes are filled with the previous scheme, the more complete and correct explanations are achieved. Suppose that it is okay to accept the ontic theorists' claim such that the mechanism scheme, including (i) Watson's central dogma, (ii) biochemists' discoveries of enzymatic reactions from DNA to protein, (iii) a special case of genetic information flow in a retrovirus, and (iv) splicing mechanism in eukaryotes, is the complete and correct model of protein synthesis. Regrettably, again, Ochoa's discovery misrepresents RNA synthesis because he never found RNA polymerase but polynucleotide phosphorylase. Roger Kornberg successfully analyzes the structural features of RNA polymerase and its functions in eukaryotes in the early twentieth century. In other words, all mechanistic schemes in the absence of Roger Kornberg's achievements are certainly incomplete and incorrect.

Again, ontic theorists would complain that the system of mechanisms, up to and including Roger Kornberg's discovery, is a complete and accurate model of protein synthesis. The problem with answering this way is that ontic theorists can never seem to qualify themselves as tools for evaluating mechanism schemes other than holding their ground. Whenever a discovery emerges, the existing mechanism scheme becomes a false description, and a more complete and accurate explanation is achieved. As biologists' research accumulates, situations like this will occur more frequently in the future. It is clear, then, that any mechanism scheme currently believed to be complete and correct will one day become a bad explanation. The ontic norms applied to evaluating explanations inevitably allow for such *pessimistic* prospects in situations where theory propagation is

the norm. In this scenario, we must reluctantly conclude that achieving a complete and precise mechanism is an elusive goal, both at present and in the foreseeable future.

This conclusion is analogous to the conclusion of the pessimistic induction argument against scientific realism. The pessimistic induction argument shows that if past successful and truly accepted theories were false, we fail to believe the realist's claim that our currently successful theories are true because current theories will also be false as past successful theories were confronted. Similarly, if one adopts the realistic point of view of the existence of a mechanism and admits that mechanism schemes are continuously developed toward more complete and more correct models, then it is skeptical to attain the complete and correct model of the mechanism.

Advocates of the ontic conception of explanation believe (i) that we can one day achieve complete and accurate mechanism descriptions and (ii) we must assess mechanism schemes under the ontic norms. However, suppose the two beliefs juxtapose with each other. In that case, we are inevitably unable to obtain complete and accurate mechanism descriptions, or we are forced to brand almost all known mechanism descriptions as bad. Ontic theorists accept that mechanistic models must be developed from how-possibly to how-actually models. Mechanism sketches become just how-possibly models when acquiring more complete and correct models. By filling in intermediate components between the initial and terminal conditions, mechanism sketches must be developed into more complete and correct schemes as how-actually models. Suppose one realizes the acquisition of a complete and correct model of the explanandum phenomenon. In that case, the how-actually model is a good explanation, and all the early sketchy how-possibly models become bad explanations. Note

that both ontic norms of completeness and correctness are criteria for evaluating mechanistic schemes from a *realistic* point of view. Craver and Darden reveal their realistic stance by saying that:

> At the most abstract level, mechanism schemas are evaluated in terms of their depth, their completeness, and their correctness. We discuss these dimensions of evaluation and some tests for evaluating a schema along these dimensions. Our orientation toward these questions might be described as a *garden-variety realism* moderated by a sensible pragmatism (Craver and Darden, 2013, emphasis added).

Craver and Darden (2013, p.94) emphasize the close relationship between an ontic norm such as correctness and realism: "The difference between how-possibly, how-plausibly, and how-actually a mechanism works is a difference in correctness. Mechanists are garden-variety realists about such things: the goal is to describe correctly enough (to model or mirror more or less accurately) the relevant aspects of the mechanism under investigation." It is taken for granted that ontic norms are realistic criteria because they are based on the external existence of relevant entities and activities of the target mechanisms.

A problem of ontic norms as criteria for evaluating mechanism schemes is that all past achievements of mechanism schemes are just bad and provisional explanations. In other words, this decision is so hasty that it raises the problem of treating all existing scientific achievements as useless and bad things. But such a judgment is too extreme. Is the discovery that genes are passed from DNA to mRNA to proteins a bad achievement? Are all the diagrams missing the splicing process a bad diagram of how proteins are synthesized? Are all of Roger Kornberg's schemes where RNA polymerase is not named or omitted bad descriptions? No matter how imperfect or inaccurate these achievements may ultimately be, they must be our historically

significant assets. It is very questionable whether these achievements are appropriately treated as useless or bad simply because they are not complete and accurate mechanism schematics.

To summarize, ontic proponents may accept that the mechanism of protein synthesis, like the diagram of Fig. 6, explains the production of the primary sequences of proteins in molecular biology. Is this mechanism complete? If they think this mechanism is complete, this judgment is false because the splicing mechanism is ignored. If they think this mechanism is incomplete, then ontic norms seem sterile to distinguish how-actual from how-possible models. Ontic theorists believe that ontic norms such as correctness and completeness play a role in demarcating good from bad explanations. Unfortunately, as long as they continue to hold on to this belief, ontic theorists end up labeling a mechanistic scheme, considered a very important biological achievement, as a 'bad' explanation. This tragic situation has resulted from adhering to a particular philosophical position, the ontic criteria for evaluating mechanism schemes. Regardless of whether the splicing mechanism was discovered, Fig. 6 is still a good representation that explains a fairly important phenomenon in protein synthesis.

4.2 Ontic norms as methodological instructions

Suppose ontic norms such as completeness are difficult to achieve, and it is problematic to classify incomplete mechanism schemes into bad explanations according to these norms. What role do these norms play? Suppose we need not seriously consider mechanism schemes to be tagged as good or bad. Ontic norms can motivate researchers to provide more advanced explanations than previous explanations of mechanisms. As noted, mechanistic explanation relies on mech-

anistic inquiries. Completeness plays an instruction in developing existing schemes by filling unidentified processes or illuminating known entities' uncovered properties.

Again, for instance, the self-splicing mechanism in eukaryotes was discovered after most processes within the mechanism of protein synthesis were known in the 1980s. When discovering the structure of DNA in 1953, Watson and Crick never acknowledged the two distinctions, exons and introns, within nucleic acids. Exons are portions that include genetic information, whereas introns do not. So, introns must be removed before a ribosome synthesizes proteins based on the genetic sequences of mRNA. Thomas Cech discovered the splicing mechanisms in *Tetrahymena* (see Fig. 7) (Zaug, Grabowski and Cech, 1983). Due to the discovery of splicing processes, understanding of protein synthesis has increased. Moreover, Cech discovered a new property of RNA. Until the 1980s, most biochemists believed that all enzymes are proteins. However, while investigating the splicing mechanism, Cech was also aware of the enzymatic property of RNA molecules. That is, Cech shed light on a novel fact that biological reactions from pre-rRNA to spliced rRNA occur in the absence of enzymes (see Fig. 7). Cech (1986) proved it experimentally so that he won the Nobel Prize in 1989. Cech's discovery of ribozyme, a compound of RNA & enzyme, contributes to making more complete lists of referents of enzymes in the world. Cech's achievements are totally compatible with the previous outcomes. Biologists investigate their research to explain biological phenomena at the molecular level more completely. Ontic norms are methodological instructions in mechanism inquiries.[21]

---

[21] Although I only focus on completeness here, correctness is also an important ontic norm in mechanism inquiries. Without referring to referents as correct entities in a mechanism, we cannot explain the explanandum phenomenon successfully.
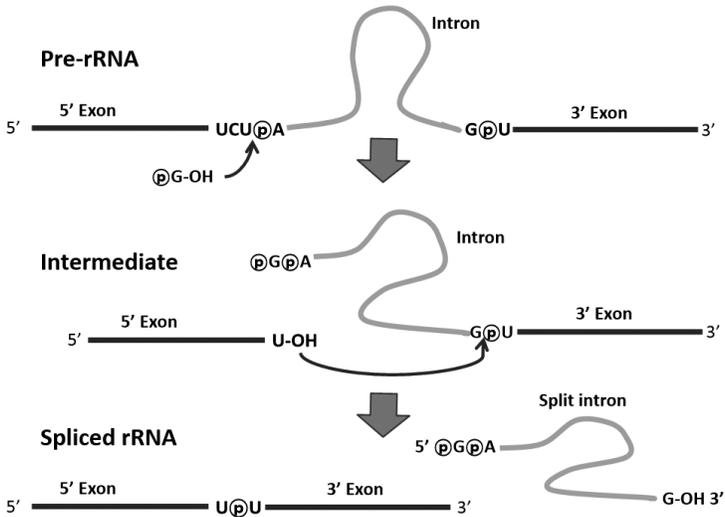
Figure 7: Cech's self-splicing mechanism (based on Zaug, Grabowski and Cech, 1983, p.582)

Finally, I would like to emphasize that whether an explanation is good or bad is not determined by any criteria but rather by its appropriateness or relevance to the topic determined in the given context. Ontic norms are instructions of mechanistic inquiries to discover patterns, not criteria to judge whether a description is good. Ontic norms are self-contradictory because scientific achievements are thrown away as bad explanations. Completeness is an infeasible Utopian standard to be realized in science. It rather is realizable following the explainer's interests. A philosophical distinction on explanatory normativity based on ontic norms seems a pseudo-project.

Interestingly, Craver and Darden also agree on the pragmatic aspect of mechanistic explanation by acknowledging that ontic norms are compatible with agents' interests or given contexts by saying:

> In speaking this way, we also intend to allow that a schema can be complete and correct enough for the *purposes* at hand without being fully complete or correct. One can acknowledge the ideals of completeness and correctness for descriptions of mechanisms while, at the same time, recognizing that science often traffics in *idealized* and *incomplete* schemas (Craver and Darden, 2013, emphasis added).

The pragmatic point of view was shown in the discussion of the conceptual dimension of the nature of explanation. Remind that the nature of mechanistic explanation is the tertiary relation among the representer, targets, and agent's goal in a given context. When judging whether a mechanistic explanation is successful, the explanation must be an adequate answer to a question about why the explanandum phenomenon happens. According to van Fraassen (1980), an explanation is an answer to a why-question. Here, I do not apply van Fraassen's theory of explanation to mechanistic explanation. Still, I emphasize that mechanistic explanations are evaluated by whether adequate representations explain the explanandum phenomenon under the agent's interest.

The same phenomenon can be explained differently depending on what aspects, to some degree, explain the phenomenon. For example, protein synthesis can be explained diversely concerning researchers' interests. Molecular biologists Watson and Crick were interested in the genetic flow from DNA to protein. The number of bases within nucleic acids is four, but the number of amino acids is twenty. So, the order of the bases in DNA determines what kinds of nucleic acids are in protein. By guessing the existence of a single linear sequence of

nucleic acids, or messenger RNA, they imagined how sequences of two or three bases correspond to 20 different types of amino acids. As a result, the simple mechanism scheme, DNA→RNA→Protein, explains a temporal order among essential entities within the mechanism. By contrast, biochemists Zamecnik and Hoagland focused on the energy flow in the form of ATP when amino acids are activated in synthesizing the peptide bond. Their chemical interest in protein synthesis led to the discovery of transfer RNA, which links the genetic code to amino acids. As a result, Zamecnik's diagram in Fig. 1 explains how biomolecules interact with each other to synthesize the peptide bond.

Further, another biochemist, Cech, was interested in what enzymes mediate the splicing reactions of precursor ribosomal RNA before mRNA moves out of a nucleus to synthesize protein. As a result, Cech's diagram in Fig. 7 explains how splicing reactions, including cleavage of pre-rRNA, ligation, and cyclization, occur without enzymatic proteins. Besides them, numerous molecular biologists and biochemists have proposed models representing various aspects of the protein synthesis process, depending on their respective fields of interest. For the same protein synthesis process, molecular biologists focus on genetic information, and biochemists focus on energy equivalence. Even among the same biochemists, someone discovers the energy components needed for chemical reactions, and someone else discovers new functions of RNA. Likewise, the topic, scope, and purpose of explaining the protein synthesis process vary depending on the researcher's interests. The important thing is that all of them are significant achievements in biological sciences, even if each incompletely explains the full processes of protein synthesis. Whether a mechanistic explanation is good or not depends upon the issue of *how much an answer to the question is attained.*

As noted, Craver, Bechtel, and Wright adopt Salmon's distinction between ontic and epistemic. However, they disagree about the nature of mechanistic explanations. Craver (2006; 2007; 2014) argues that the mechanistic explanation is also ontic, whereas Bechtel and Wright (Wright and Bechtel, 2007; Wright, 2012; 2015) claim it is epistemic. Both agree that the mechanistic explanation should be based on something other than Hempel's account of scientific explanation. More precisely, they claim that a mechanistic explanation does not include the laws of nature, and a linguistic form of argument does not constrain the explanation units. Based on a consensus of explanatory relevance in the first dimension, the three philosophers disagree on what mechanistic explanation is in the second dimension. Craver over-concentrates on the relational dimensions of mechanistic explanation in line with the first dimension of explanatoriness. On the contrary, Bechtel and Wright over-concentrate on the conceptual dimension of mechanistic explanation, such cognitive procedures independently from the first dimension. Remember that Craver rejects cognitive processes to identify causal factors of explanandum phenomena. As Illari mentions, Craver wants to proclaim an ontic constraint to examine mechanistic explanation in terms of mechanisms.

I restate the first and second dimensions to show that the normative dimension can depend on the two dimensions. Ontic norms of mechanistic explanation originate from the first relational dimension on which all mechanists agree. Epistemic norms of mechanistic explanation stem from the second conceptual dimension of the nature of explanation. I am emphasizing the non-existence of disagreement among ontic and epistemic theorists about a claim that mechanistic explanation is an achievement from cognitive performances in practice. What is important is that epistemic norms are also instructions in

mechanistic inquiries. This point has been similarly pointed out by many philosophers of science (Illari, 2013; van Eck, 2015; Sheredos, 2016; Kästner and Haueis, 2021).

Nonetheless, I cannot entirely agree with the defenders of the epistemic priority of the mechanistic explanation as they maintain that there are separate epistemic constraints, such as generality or systematicity, which are either independent of or additional to the ontic constraint (van Eck, 2015; Sheredos, 2016). The goal of understanding entities, activities, and their organizations is not independent of the ontic constraint, according to which we should discover the causal factors to explain a phenomenon. Other epistemic norms for the mechanistic explanation may be helpful only after discovering all causal factors. Without the ontic constraint, any epistemic norms independently do not support the explanatory relevance of the mechanistic explanation. Discoveries, encompassing not only the components but also their organizational features within a mechanism, serve as primary methodological guidance in biological sciences. Thus, ontic norms are not less worthy than any epistemic norms.

Are ontic and epistemic norms equally valuable? Illari (2013) and Kästner & Haueis (2021) appear to answer this question affirmatively, a stance with which I concur. However, my focus extends beyond the relative significance of these norms. I explored whether ontic norms warrant consideration in evaluating mechanistic explanations. I demonstrated that ontic norms alone are insufficient for distinguishing between good and bad explanations. Consequently, both ontic and epistemic norms play essential roles, as neither norm conflicts independently of the demarcation problem concerning the quality of mechanistic explanations.

Conclusion

The paper analyzes ontic and epistemic debates in scientific explanation through the three-dimensional interpretation. Each debate dispute concerns a particular issue: (1) explanatory formal framework, explanatory force, explanatory relevance between explanans and an explanandum; (2) nature of mechanistic explanation; and (3) normative constraints. This dimensional approach does not simply enumerate or give three names to different issues but is helpful to compare with participants' viewpoints in this debate. In addition, it provides a practical framework to comprehend several philosophical perspectives in scientific explanation. Furthermore, those three dimensions have their philosophical issues but are also interconnected, so it can be helpful to clarify how to reconcile different normative constraints.

Recent ontic-epistemic debates are displayed among proponents of the New Mechanism. The mechanistic explanation is a widespread explanatory practice in biological sciences. A consensus exists that the explanatory relevance between the explanans and the explanandum is a fundamental ontic constraint for a mechanistic explanation. However, some philosophers have either focused on the epistemic nature of the mechanistic explanation or insisted on the epistemic norms or constraints over the ontic constraints. I have shown that mechanistic explanations are better associated with the epistemic conception than with the ontic conception of explanation. I emphasized the epistemic nature of mechanistic explanations by highlighting their strong dependence on the activities that represent the mechanism based on van Fraassen's pragmatic viewpoint. Contrary to stereotypical views held by New Mechanism advocates, I proposed the linkage of mechanistic explanations to epistemic aspects such as logical validity, lawful grounds, and counterfactual inference. This is achieved by exploring

the necessary conditions in linguistic representations of these mechanistic explanations. Further, I addressed the philosophical issue of how mechanistic explanations should be evaluated.

Unlike many philosophers in the New Mechanism who stubbornly reject Hempel's model, I suggested to view Hempel's conditions as necessary conditions for linguistic representation rather than sufficient conditions for scientific explanation. By noting that the usage of the ontic norms as criteria for distinguishing between good and bad explanations makes the mistake of dismissing many existing scientific achievements as 'bad' explanations, I assert the compatibility of mechanistic explanations with pragmatic aspects of explanation. Through this, I carefully tried to reveal that the mechanistic explanation is significantly more related to the epistemic position than the ontic one.

Bibliography

Atkins, P.W., De Paula, J. and Friedman, R., 2014. *Physical Chemistry: Quanta, Matter, and Change*. Second edition. Oxford, United Kingdom: Oxford University Press.

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

Bechtel, W. and Abrahamsen, A., 2005. Explanation: a mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), pp.421–441. https://doi.org/10.1016/j.shpsc.2005.03.010.

Bechtel, W. and Richardson, R.C., 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press.

Bokulich, A., 2016. Fiction as a vehicle for truth: Moving beyond the ontic conception. *The Monist*, 99(3), pp.260–279. https://doi.org/10.1093/monist/onw004.

Carnap, R., 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press; Routlege & Kegan Paul.

Cech, S., 1986. A model for the RNA-catalyzed replication of RNA. *Proceedings of the National Academy of Sciences*, 83(12), pp.4360–4363. https://doi.org/10.1073/pnas.83.12.4360.

Coffa, J.A., 1973. *Foundations of Inductive Explanation*. PhD Thesis. University of Pittsburgh.

Craver, C.F., 2006. When mechanistic models explain. *Synthese*, 153(3), pp.355–376. https://doi.org/10.1007/s11229-006-9097-x.

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

Craver, C.F., 2014. The Ontic Account of Scientific Explanation. In: M.I. Kaiser, O.R. Scholz, D. Plenge and A. Hüttemann, eds. *Explanation in the Special Sciences: The Case of Biology and History*. Vol. 367, *Synthese Library*. Dordrecht: Springer Netherlands, pp.27–52. https://doi.org/10.1007/978-94-007-7563-3_2.

Craver, C.F. and Darden, L., 2013. *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago, IL: University of Chicago Press. Available at: <https://press.uchicago.edu/ucp/books/book/chicago/I/bo16123713.html> [visited on 29 September 2023].

Darden, L., 2006. *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*, *Cambridge studies in philosophy and biology*. Cambridge: Cambridge University Press.

van Eck, D., 2015. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Axiomathes*, 25, pp.5–22. https://doi.org/10.1007/s10516-014-9243-x.

Ellis, B., 1956. On the relation of explanation to description. *Mind*, 65(260), pp.498–506. https://doi.org/10.1093/mind/65.1.498.

van Fraassen, B.C., 1980. *The Scientific Image*, *Clarendon library of logic and philosophy*. Oxford; New York: Clarendon Press; Oxford University Press.

van Fraassen, B.C., 2008. *Scientific Representation: Paradoxes of Perspective*. 1st ed. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199278220.001.0001.

Giere, R.N., 2006. *Scientific Perspectivism*. Chicago, IL: University of Chicago Press.

Hempel, C.G. and Oppenheim, P., 1948. Studies in the logic of explanation. *Philosophy of Science*, 15(2), pp.135–175. https://doi.org/10.1086/286983.

Hempel, C.G., 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Hoagland, M.B. et al., 1958. A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of Biological Chemistry*, 231(1), pp.241–257.

Holley, R.W. et al., 1965. Structure of a ribonucleic acid. *Science*, 147(3664), pp.1462–1465. https://doi.org/10.1126/science.147.3664.1462.

Illari, P., 2013. Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, 78(S2), pp.237–255. https://doi.org/10.1007/s10670-013-9511-y.

Jacob, F. and Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), pp.318–356. https://doi.org/10.1016/S0022-2836(61)80072-7.

Judson, H., 2013. *The Eighth Day of Creation Makers of the Revolution in Biology*. New York: Cold Spring Harbor Laboratory Press. Available at: <https://www.academia.edu/32355053/Horace_Judson_THE_EIGHTH_DAY_OF_CREATION_Makers_of_the_Revolution_in_Biology> [visited on 8 December 2023].

Kaplan, D.M. and Craver, C.F., 2011. The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78(4), pp.601–627. https://doi.org/10.1086/661755.

Kästner, L. and Haueis, P., 2021. Discovering patterns: On the norms of mechanistic inquiry. *Erkenntnis*, 86(6), pp.1635–1660. https://doi.org/10.1007/s10670-019-00174-7.

Machamer, P., Darden, L. and Craver, C.F., 2000. Thinking about mechanisms. *Philosophy of Science*, 67(1), pp.1–25. https://doi.org/10.1086/392759.

Pardee, A.B., Jacob, F. and Monod, J., 1959. The genetic control and cytoplasmic expression of "Inducibility" in the synthesis of β-galactosidase by E. coli. *Journal of Molecular Biology*, 1(2), pp.165–178. https://doi.org/10.1016/S0022-2836(59)80045-0.

Povich, M., 2018. Minimal models and the generalized ontic conception of scientific explanation. *The British Journal for the Philosophy of Science*, 69(1), pp.117–137. https://doi.org/10.1093/bjps/axw019.

Rottman, F. and Nirenberg, M., 1966. RNA codons and protein synthesis: XI. Template activity of modified RNA codons. *Journal of Molecular Biology*, 21(3), pp.555–570. https://doi.org/10.1016/0022-2836(66)90027-1.

Salmon, W.C., 1971. *Statistical Explanation & Statistical Relevance*, *Pitt paperback 69*. Pittsburgh: University of Pittsburgh Press.

Salmon, W.C., 1984. Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, (2), pp.293–305. https://doi.org/10.1086/psaprocbienmeetp.1984.2.192510.

Salmon, W.C., 1989. Four decades of scientific explanation. In: P. Kitcher and W.C. Salmon, eds. *Scientific Explanation*, *Minnesota studies in the philosophy of science*, 13. Minneapolis: University of Minnesota Press, pp.3–219. Available at: <http://conservancy.umn.edu/handle/11299/185700> [visited on 4 October 2023].

Sheredos, B., 2016. Re-reconciling the epistemic and ontic views of explanation (or, why the ontic view cannot support norms of generality). *Erkenntnis*, 81(5), pp.919–949. https://doi.org/10.1007/s10670-015-9775-5.

Wright, C., 2012. Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science*, 2(3), pp.375–394. https://doi.org/10.1007/s13194-012-0048-8.

Wright, C., 2015. The ontic conception of scientific explanation. *Studies in History and Philosophy of Science Part A*, 54, pp.20–30. https://doi.org/10.1016/j.shpsa.2015.06.001.

Wright, C. and Bechtel, W., 2007. Mechanisms and psychological explanation. In: P. Thagard, ed. *Philosophy of Psychology and Cognitive Science*. Elsevier, pp.31–79. https://doi.org/10.1016/B978-044451540-7/50019-0.

Wright, C. and van Eck, D., 2018. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Ergo, an Open Access Journal of Philosophy*, 5(38). https://doi.org/10.3998/ergo.12405314.0005.038.

Zaug, A.J., Grabowski, P.J. and Cech, T.R., 1983. Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage–ligation reaction. *Nature*, 301(5901), pp.578–583. https://doi.org/10.1038/301578a0.

A conventionalist account of distinctively mathematical explanation

Mark Povich

University of Rochester

Abstract

Distinctively mathematical explanations (DMEs) explain natural phenomena primarily by appeal to mathematical facts. One important question is whether there can be an ontic account of DME. An ontic account of DME would treat the explananda and explanantia of DMEs as ontic items (ontic objects, properties, structures, etc.) and the explanatory relation between them as an ontic relation (e.g., Pincock, 2015; Povich, 2021). Here I present a conventionalist account of DME, defend it against objections, and argue that it should be considered ontic. Notably, if indeed it is ontic, the conventionalist account seems to avoid a convincing objection to other ontic accounts (Kuorikoski, 2021).

Keywords

scientific explanation, philosophy of science, philosophy of mathematics, mathematical explanation, ontic conception, conventionalism.

1. Introduction

Distinctively mathematical explanations[1] (DMEs) explain natural phenomena primarily by appeal to mathematical facts. DMEs have been receiving a lot of attention for a few good reasons (Steiner, 1978; Colyvan, 1998; Baker, 2005; 2009; Mancosu, 2008; Saatsi, 2011; 2012; 2016; Lyon, 2012; Lange, 2013; 2016; 2018; Pincock, 2015; Reutlinger, 2016; Craver and Povich, 2017; Povich, 2020; 2021). Some philosophers (e.g., Baker, 2005; 2009; contra Bangu, 2008; and Saatsi, 2011) take them to play a crucial role in enhanced indispensability arguments, providing good evidence for the existence of the mathematical objects to which they appeal. One important question is whether there can be an ontic account of DME. An ontic account of DME would treat the explananda and explanantia of DMEs as ontic items (ontic objects, properties, structures, etc.) and the explanatory relation between them as an ontic relation (e.g., Pincock, 2015; Povich, 2021) Here I present a conventionalist account of DME, defend it against objections, and argue that it should be considered ontic. Notably, the conventionalist account seems to avoid a convincing objection to other ontic accounts (Kuorikoski, 2021). I take my arguments to be far from conclusive, but to show that such a view is worth considering.

In Section 2, I elaborate on DME and present a paradigmatic example that I will use throughout the paper. My arguments ought to apply, *mutatis mutandis*, to other examples.[2] In Section 3, I briefly explain two recent ontic accounts of DME from Pincock (2015) and myself (Povich, 2021). I will also recount Kuorikoski's (2021) objection to ontic accounts of DME. In Section 4, I explain the kind of

---

[1] Also sometimes called "extra-mathematical explanations" (e.g., Baron, 2016; 2020).

[2] I discuss many more examples in my forthcoming book (Povich, 2024).

conventionalism to which I will appeal. In Section 5, I explain how to give the previously presented ontic accounts a conventionalist twist, which deflates their platonism. This deflating allows ontic accounts to escape Kuorikoski's objection. However, one might legitimately wonder whether deflated accounts are still ontic.[3] In Section 6, I argue that they are. In Section 7, I respond to objections.

2. Distinctively Mathematical Explanations

DMEs work primarily by showing a natural explanandum to follow in part from a mathematical fact—a fact modally stronger than any fact about causes, mechanisms, and even natural laws. A DME shows that the explanandum had to happen, in a sense stronger than any ordinary causal law can supply (Lange, 2013). One example of DME, which I will use throughout, is Trefoil Knot (Lange, 2013). The explanandum is the fact that Terry failed to untie his knot. The explanantia are the empirical fact that the knot is a trefoil knot and the mathematical fact that the trefoil knot is distinct from the unknot (i.e., mathematically cannot be untied). The explanantia mathematically ensure that Terry fails to untie the knot, for his success is mathematically impossible.

There are several distinctive features of DME, accounting for which serve as desiderata for any account of DME:

The Modal Desideratum: an account of DME should accommodate and explicate the modal import of some DMEs. (Baron, 2016)

---

[3] One might also legitimately wonder whether there *are* DMEs. For the sake of this paper, I assume that there are.

As I just mentioned, there is a modal robustness to Terry's failure—
he had to fail. An account of DME should capture and, preferably,
explicate that modal force.

> *The Distinctiveness Desideratum*: it should distinguish uses of
> mathematics in explanation that are distinctively mathematical
> from those that are not. (Baron, 2016)

This is emphasized by defenders of the enhanced indispensability
argument (EIA, e.g., Baker, 2009) According to the EIA, we ought to
believe in the existence of (certain) mathematical objects because they
play an indispensable explanatory role in science. The examples to
which defenders of the EIA appeal are DMEs. For them, what is dis-
tinctive about DMEs is the explanatory—not merely representational—
role that mathematics plays in them. Bromberger's (1966) flagpole is
a well-known example of an explanation that uses mathematics but is
not a DME. The explanandum is the fact that the length of a flagpole's
shadow is L. The explanantia are the empirical facts that the angle of
elevation of the sun is θ and that the height of the flagpole is H and
the mathematical fact that *tan $\theta = H/L$*. Most party to the debate on
DME agree that this is not a DME. Precise explanations of why may
depend on one's account of DME, but the central idea is that in this
example the mathematics is playing a merely representational role,
where this means that the mathematics is merely representing what
is in fact doing the real explanatory work (e.g., the physical causes).
Any account of DME should count Trefoil Knot as a DME and not
Bromberger's flagpole.

> *The Directionality Desideratum*: it should accommodate the di-
> rectionality of DMEs. (Craver and Povich, 2017)

Craver and I argue that Trefoil Knot can be "reversed"[4] to form an argument that fits Lange's (2013) account of DME but is not explanatory. Simply take the explanandum and the empirical premise, swap and negate them, akin to turning a *modus ponens* into a *modus tollens*. Thus, the "reversed" explanandum is the fact that Terry's knot is not trefoil. The empirical explanans is the fact that Terry untied his knot. The mathematical explanans is the same: the trefoil knot is distinct from the unknot. Reversed Trefoil Knot and other such reversals should not count as DMEs.

3. Ontic Accounts of DME

In this section, I briefly present two ontic accounts of DME: Pincock's (2015) abstract dependence account and my Narrow Ontic Counterfactual Account (NOCA; Povich, 2021). I focus on ontic accounts since these are especially well-suited for EIAs, and I focus on Pincock's and mine in particular since these are two of the few that are explicitly billed as ontic.[5] According to these accounts, purely mathematical claims refer to Platonistic facts and applied mathematical claims refer to instantiations of mathematical objects. Other accounts of DME that are not explicitly billed as platonistic could be given an platonistic spin, and my arguments that follow plausibly undermine their ability

---

[4] These are not strict reversals—simple swaps of explanandum and explanans—like the well-known reversal of Bromberger's flagpole. Henceforth, I will drop the scare quotes.

[5] Reutlinger (2016) also gives a counterfactual account of DME, but it is not explicitly ontic, nor does it rely on countermathematicals. If it is ontic, my arguments to follow apply.

to serve in EIAs as well.[6] Ontic accounts like mine explicitly rely on countermathematicals, which I address below, but I do not take such reliance to be essential to ontic accounts.

I think that, ultimately, my account is basically a version of Pincock's, though I elaborate the kinds of counterfactual that the ontic relation of instantiation supports, or must support to figure in a DME, and I explicitly argue that the resulting account satisfies the above desiderata. Also, it is important to note that Pincock did not intend to give an account of DME. He wanted to argue 1) that there is a kind of explanation involving abstract entities, which he called "abstract explanation," 2) that abstract explanation is not causal, and 3) that causal explanation and abstract explanation both count as explanation in virtue of providing information about objective dependence relations. It is clear though that the examples usually given of DME, including Trefoil Knot, are abstract explanations in Pincock's sense. Being ontic accounts of DME that rely on ontic relations between abstract mathematical and concrete phenomena, Pincock's and mine would be especially suited to enhanced indispensability arguments (Baker, 2009)—if either account is the right account of DME, then, it would seem, platonism straightforwardly follows. However, my cen-

---

[6] For example, Lange (2013) specifically refers to his account as a modal, rather than an ontic, one. According to his, in a DME, the purely mathematical premises refer to facts "modally stronger than ordinary causal laws" and the empirical premises refer to facts that are "understood to be constitutive of the physical task or arrangement at issue," a condition that is never clearly explicated. Lange (2021) explicitly addresses the metaphysics of DME and defends "Aristotelian realism," according to which "mathematics concerns mathematical properties possessed by physical systems," which is explicitly anti-Platonist. Lange's Aristotelian realist construal of DME will obviously be of no use in an EIA, but one could give Lange's (2013) basic account an ontic spin, e.g., by claiming that the facts "modally stronger than ordinary causal laws" are Platonistic facts and the empirical facts that are "understood to be constitutive of the physical task or arrangement at issue" are instantiations of mathematical objects. The arguments of this paper would then straightforwardly apply to that ontic account.

tral argument is that conventionalism about mathematical necessity undermines this inference by deflating ontic accounts of DME, yet, arguably, does not undermine their status as ontic accounts.

I start with Pincock's (2015) account. Pincock motivates his abstract dependence account using the explanation of Plateau's three laws for soap-film surfaces and bubbles:

> First, a compound soap bubble or a soap film spanning a wire frame consists of flat or smoothly curved surfaces smoothly joined together. Second, the surfaces meet in only two ways: Either exactly three surfaces meet along a smooth curve or six surfaces (together with four curves) meet at a vertex. Third, when surfaces meet along curves or when curves and surfaces meet at points, they do so at equal angles. In particular, when three surfaces meet along a curve, they do so at angles of 120 with respect to one another, and when four curves meet at a point, they do so at angles of close to 109. (Almgren and Taylor, 1976, p.82; quoted in Pincock, 2015, p.858)

The explanation for these laws relies on the mathematical proof that certain mathematical objects called "almost minimal sets" satisfy Plateau's three laws and that soap films instantiate almost minimal sets. As Pincock writes, "Many mathematical structures have concrete systems as instances. The almost minimal sets have soap films as some of their instances, and this is what makes facts about sets relevant to facts about soap films" (Pincock, 2015, pp.865–866).

Pincock suggests that the kind of explanation involved here—so-called abstract explanation—is akin to causal explanation on Woodward's (2003) interventionist account, though shorn of its interventionism. Woodward emphasizes that the ability to answer what-if-things-had-been-different questions (i.e., w-questions) regarding the explanandum—thus, knowledge of information about counterfactual

dependence relations—is constitutive of explanation. Woodward even suggests that there could be non-causal explanations in cases where information about counterfactual dependence relations is provided, but those relations cannot sensibly be interpreted as involving interventions (Woodward, 2003, p.221). Similarly, for Pincock, what makes both causal and abstract explanations *explanations* is that they reveal "objective dependence relations". The relation between almost minimal sets and soap films is not a causal one, but it is, Pincock argues, a kind of objective dependence relation—what he calls abstract dependence[7]. In the case at hand, Pincock suggests that the abstract dependence relation in question is that of being an instance or instantiation (Pincock, 2015, p.865). Thus, an abstract explanation (or at least one important kind of abstract explanation) seems to be an explanation in which a concrete object is shown to be a certain way because it is an instantiation of an abstract object that is that way.

I also generalize Woodward's interventionism within an ontic account of explanation and suggest that the ontic relation involved in DMEs is instantiation (Povich, 2018; 2021). However, I was specifically concerned with giving an account of DME that satisfies the above desiderata. According to my Narrow Ontic Counterfactual Account (NOCA), an explanation is a DME just in case either a) it shows a natural fact (weakly) necessarily to depend counterfactually only on a mathematical fact, or b) it shows a natural event to be necessitated by a component natural fact that (weakly) necessarily counterfactu-

---

[7] Pincock is not clear about the relation between abstract dependence and counterfactual/countermathematical dependence, but the very idea of a dependence relation, as well as the comparison with Woodward's account, seems to imply counterfactual dependence, regardless of whether the dependence relation in question can be *reduced* to or *analyzed* in terms of counterfactuals. Regardless, my argument that conventionalism renders invalid any EIA starting from Pincock's account and ending in Platonism does not depend on this.

ally depends only on a mathematical fact. For example, the fact that Terry's trefoil knot is distinct from the unknot fact (weakly) necessarily counterfactually depends only on the mathematical fact that the trefoil knot is distinct from the unknot. This means that in every world where Terry's trefoil knot is distinct from the unknot, that fact counterfactually depends only on the mathematical fact that the trefoil knot is distinct from the unknot. The fact that Terry did not untie his trefoil knot is necessitated by the fact that Terry's trefoil knot is distinct from the unknot; and the former fact "contains" the latter fact (i.e., all the objects and properties that compose the latter fact are present in the former). Thus, both the fact that Terry's trefoil knot is distinct from the unknot and the fact that he did not untie his trefoil knot are (according to clauses a) and b) of NOCA, respectively) distinctively mathematically explained by the mathematical fact that the trefoil knot is distinct from the unknot.

I argued that ontic structures, objects, etc. are required to be the truthmakers for the counterfactuals to which NOCA appeals. According to NOCA, in DMEs there is a kind of counterfactual dependence of natural facts on mathematical facts. What *is* this dependence relation? I suggested two possible ontic relations—grounding and instantiation—but I prefer instantiation. For example, the concrete natural fact that Terry's trefoil knot is distinct from the unknot is an instantiation of the abstract mathematical fact that the trefoil knot is distinct from the unknot.[8] Thus, the following is equivalent (at least extensionally) to NOCA: an explanation is a DME just in case either a) it shows a natural fact to be an instantiation of a mathematical fact, or b) it shows a natural event to be necessitated by a component natural fact that instantiates a mathematical fact.

---

[8] I.e., the object and property that compose the natural fact are instantiations (or realizations) of the object and property that compose the mathematical fact.

I will not go through the rigmarole of explaining how NOCA is supposed to satisfy the three desiderata above, but I want to point out that Platonism plays no role in NOCA's ability to satisfy them. I appealed to Platonic objects to provide truthmakers for the relevant counterfactuals. That Platonism plays no role in NOCA's ability to satisfy the three desiderata can be seen by the fact that the two clauses of NOCA satisfy all three desiderata by themselves, without relying on any specific metaphysics of mathematics. In the case of NOCA, unlike other appeals to the ontic (e.g., to causation in the case of Bromberger's flagpole; Salmon, 1984; 1989) the ontic is unnecessary to secure explanatory directionality and thus satisfy the directionality desideratum. In the present paper, I will argue that Platonism is not required to provide truthmakers for the relevant counterfactuals. Conventionalism gives the nominalist a way to understand the counterfactuals involved in DMEs without positing abstract mathematical facts or abstract ontic dependence relations.

Note that ontic dependence accounts of DME must rely on countermathematicals: counterfactuals with mathematically impossible antecedents. For example, I argued that of the following countermathematicals, the first but not the second is weakly necessarily true, that is, true in every world where Terry's trefoil knot is distinct from the unknot:

> Were the trefoil knot isotopic to the unknot, Terry's trefoil knot would have been isotopic to the unknot.

> Were the trefoil knot isotopic to the unknot, Terry's untieable[9] knot would have been trefoil.

Of course, on a standard Lewis' (1973) semantics of counterfactuals, all counterpossibles (i.e., counterfactuals with impossible antecedents)

---

[9] By "untieable," I mean "able to be untied," not "unable to be tied".

are vacuously true. One straightforward amendment to the Lewisian account is to introduce impossible worlds (Brogaard and Salerno, 2013; see Kocurek forthcoming for a survey of approaches to counterpossibles). One feature of conventionalism is that it can provide an account of countermathematicals that avoids ontological commitment to impossible worlds. More on this in the next section.

Before that, I want to describe Kuorikoski's (2021) objection to ontic accounts of DME. Kuorikoski argues that any ontic account of DME, such as Pincock's or mine, cannot accommodate the Woodwardian 'same-object condition,' which requires that in counterfactual reasoning we really are reasoning about the *same* object under different conditions. According to Kuorikoski, when reasoning *countermathematically* we cannot distinguish whether we are conceiving of a change in a given mathematical structure or simply a different mathematical structure. As Kuorikoski puts the objection, "if there is no difference between changing a specific property of a mathematical object into something else and simply contemplating the properties of a different mathematical object, we lose the very distinction between explanatory and classificatory information" (Kuorikoski, 2021, p.197). The idea is that, in countermathematicals like "Were the trefoil knot isotopic to the unknot, Terry's trefoil knot would have been isotopic to the unknot," I did not give a stipulation-independent reason to think that a 'trefoil knot' isotopic to the unknot would still *be* a trefoil knot. (Nor did Pincock provide anything similar, *if* his account needs to rely on countermathematicals.) This is required for the counterfactual to express an explanatory relationship between antecedent and consequent. Such stipulation-independent reasons would basically amount to a theory of the essential and accidental properties of all mathematical objects involved in DMEs. Not only is this task daunting, but there is no guarantee that upon its completion, all the countermathemati-

cals involved in DMEs will come out as same-object-satisfying, i.e., that they will involve countermathematicals whose antecedents state changes in the object's accidental properties. And even if by sheer luck all countermathematicals involved in current DMEs come out as same-object-satisfying, there seems nothing to prevent a DME that appeals to the essential properties of a mathematical object, failing to make the associated countermathematical same-object-satisfying. E.g., suppose that being prime is an essential property of 3-3 wouldn't be 3 if it weren't prime. There's no guarantee that there are no DMEs that appeal to the fact that 3 is prime. The countermathematical in that case would be "if 3 weren't prime, ..." which by assumption isn't same-object-satisfying.

4. Conventionalism

There are two conventionalistic philosophies that for the purposes of this paper I will treat as equivalent:Amie Thomasson's (2020; 2021) modal normativism and Jared Warren's (2020) conventionalism (see also Kocurek, Jerzak and Rudolph, 2020; Sidelle, 1989). I will help myself to the language of both in the following sections, and I think that either can provide adequate deflations of the previously described ontic accounts. I will briefly explain both views, and why I will treat them as equivalent.

Thomasson's (2020; 2021) modal normativism is somewhat similar to expressivism about metaphysically necessity and possibility. Although Thomasson's normativism concerns specifically *metaphysical* modality, it is easily generalizable to mathematics. According to mathematical normativism, mathematical claims do not describe, in any *substantive* sense, anything, but instead are object language *ex-*

pressions of conceptual/semantic[10] rules[11] or consequences thereof[12]. A mathematical claim such as "3 is prime" is an expression of a semantic rule according to which it is correct to apply "is prime" when it is correct to apply "3".[13]

I say that Thomasson's normativism is "somewhat" similar to expressivism, because she accepts the existence of modal truths, facts, etc., as long as all of these terms are understood in suitably deflationary senses that are clearly distinguished from the senses these terms have in talk of non-modal, empirical truths, facts, etc. (Thomasson, 2020; see also Baker and Hacker, 2009).[14]

According to Warren's (2020) conventionalism, all mathematical truths in a language are fully explained by (the validity of) the basic inference rules of that language. Warren isn't exactly clear about what

---

[10] I ignore this distinction here. This should not affect my argument.

[11] Rules which may include empirical variables to account for *a posteriori* necessities (Sidelle, 1989; Thomasson, 2020; Warren, 2022b), but these, as well as *de re* necessities, are irrelevant to the present discussion.

[12] This is reminiscent of Wittgenstein (1978; 2013).

[13] Throughout, when I say "mathematics" or "mathematical," I am talking about pure mathematics. I leave aside how conventionalists can account for applications of mathematics. See Warren (2020) for a discussion of conventionalism and applicability.

[14] Obviously, the problem of so-called "creeping minimalism" in metaethics (Dreier, 2004)—i.e., how to distinguish moral expressivism from moral realism once the expressivist adopts semantic minimalism or deflationism—applies here. Here I merely point the reader to what I think is a promising way of solving this problem (Simpson, 2020). Adjusting Simpson's solution to the topic of modality, he would hold that normativism differs from its rivals by not having to appeal to modal facts to explain (the content of) modal language and thought. See also Brandom's (2008) explanation of modal language. The conventionalist might also appeal to analyticity: what distinguishes conventionalism from Platonism is that the conventionalist takes mathematical claims (including existence claims) to be analytic, whereas the Platonist doesn't. This might seem to misclassify neo-Fregeans who are Platonists yet think that mathematical claims (including existence claims) are analytic (e.g., Hale and Wright, 2001). However, Thomasson (2014) and Warren (2020) are both generally sympathetic to neo-Fregeanism, with Warren (2020, pp.198, 203) calling it "conventionalist-adjacent," and it is often grouped with metaontological deflationisms or minimalisms.

"fully explained" means. Being derivable from the basic inference rules is clearly sufficient for "full explanation," as the example in the next paragraph shows.

Notably, according to Warren's (2020) conventionalism, it is not the case that mathematical truths *describe* conventions. You could say that arithmetical truths describe numbers because their terms refer to numbers, but such reference—and, therefore, existence—is a trivial byproduct of our arithmetical language. For example, let us assume our arithmetical language is formally modeled by first-order Peano arithmetic, one of whose basic inference rules allows the derivation of "N0" (i.e., "zero is a number") from no premises. From this, we can easily derive "there is a number" via the introduction rule for the existential quantifier. Thus, the existence of numbers is a trivial byproduct of our arithmetical language. Thomasson and Warren are both deflationary "trivial realists" in mathematical ontology.

There are some obvious differences between Warren's and Thomasson's views, but I will treat them as equivalent for my purposes because the differences will not affect the use to which I put them. One difference that is unimportant is Warren's emphasis on inference rules and Thomasson's emphasis on application conditions. This should not affect my arguments since Thomasson is aware that application conditions might not be the only kind of semantic rule that is expressed by necessary statements, and Warren accepts that application conditions can be meaning-determining (Warren, 2022a). A potentially more significant difference is the apparent fact that Thomasson is an expressivist and Warren is not—he's an inferentialist. However, Thomasson is not an expressivist in the traditional sense, which is one reason she prefers the term 'normativism'. Normativism is not a semantic or metasemantic thesis, like traditional expressivism, which has an 'ideational' (meta)semantics; normativism

is for Thomasson a functional thesis, a thesis about the function of a piece of language. In fact, like Warren, she is an inferentialist about meaning (Thomasson, 2020, p.79), and the functional thesis is entirely open to Warren. As my argument progresses, I will make clear how normativists and conventionalists can say the same thing.From now on, when I say "conventionalist," I will tend to mean Thomasson's version, but everything I argue is open to Warren as well.

As we saw, ontic accounts of DME rely on non-vacuous countermathematicals. Fortunately, modal normativism has already been used to provide an account of non-vacuous counterpossibles with metaphysically impossible antecedents (Locke, 2021) and of non-vacuous counterpossibles with (meta)logically impossible antecedents (Kocurek and Jerzak, 2021). Consider the counterpossible: Were Goliath (the statue) to survive being flattened, it would be an abstract object. According to Locke, such counterpossibles have non-vacuous readings that express[15] the consequences of changing our semantic rules only as much as the antecedent demands.[16] This counterpossible

---

[15] It is important to note that expressing what would be the case if actual semantic rules had been different is not the same as expressing actual semantic rules. According to normativism, only necessities express actual semantic rules, so only if a counterpossible is necessary (and some may be; see below) does it express actual semantic rules. One could simply avoid talk of "expressing" here by saying that non-vacuous readings of counterpossibles involve changing semantic rules. Thus, when I say, "conventionalist account of counterpossibles," I do not mean conventionalism *about* counterpossibles, viz., the view that counterpossibles express actual semantic rules or consequences thereof; most do not. I simply mean what the conventionalist says is going on in non-vacuous counterpossibles, viz., that we consider actually adopting different semantic rules.

[16] This is the conventionalist analogue of "in the nearest possible world where the antecedent is true". In general, the conventionalist can give a semantic interpretation of Baron, Colyvan, and Ripley's (2017) account of the evaluation of countermathematicals—instead of conceiving ourselves as "twiddling" mathematical facts and thinking through the ramifications, we "twiddle" concepts, their application conditions, etc., and think through the ramifications.

expresses the claim that if the application conditions of statue-names like "Goliath" were changed so as to continue to apply after flattening, Goliath would be an abstract object. Locke argues that this is false, because when we imagine changing the application conditions of "Goliath" only so much that it continues to apply after being flattened, we have not changed that part of the application conditions that ensures it only applies to concrete objects. Kocurek and Jerzak (2021) argue for the same idea regarding counterfactuals with (meta)logically impossible antecedents. According to them, a counterlogical such as "If intuitionistic logic were correct, the continuum hypothesis would be either true or not true" has a non-vacuous reading that expresses the consequences of accepting the intuitionist's semantic rules for 'or' and 'not'. On this reading, it is false. Note that these ideas are entirely open to Warren,[17] who could treat counterpossibles as expressing what would be true according to different conventions.

For the previous claims to make sense, it is important to introduce a distinction. Einheuser (2011) called readings of counterfactuals on which we consider actually adopting different semantic rules "counterconceptual" readings and readings on which we do not change our semantic rules "countersubstratum" readings.[18,19] Note that these do not refer to kinds of counterfactual but to ways of reading coun-

---

[17] He in fact cites Einheuser approvingly several times. Her work is discussed in the following paragraph of the main text.

[18] A counterconceptual reading of a counterfactual bears similarities to some two-dimensionalists' notion of considering a possible world as actual (Stalnaker, 2001). I think there are many significant commonalities between conventionalism and some versions of two-dimensionalism, especially Stalnaker's, but that is beyond the scope of this paper. (See also the mention of Chalmers and Stalnaker in Section 7 below.)

[19] The conventionalist can agree with the Lewisian that all countersubstratum readings of counterpossibles are vacuously true. This is Kocurek and Jerzak's view.

terfactuals. Using this distinction, we can say that according to the conventionalist, counterpossibles are non-vacuous on counterconceptual readings.

In many instances of counterfactual reasoning, we automatically give countersubstratum readings of counterfactuals, that is, we continue to use our actual semantic rules (Kripke, 1980; Wright, 1985; but see Kocurek, Jerzak and Rudolph, 2020 for cases where it is natural to give counterfactuals counterconceptual readings). It is plausible that this is how we naturally read so-called "independence conditionals" such as "even if our semantic rules had been different, the necessities would not have been different" (Thomasson, 2007a; see also Sidelle, 2009; Thomasson, 2020). The conventionalist can accept this: countersubstratum readings of that counterfactual are indeed true.

5. Conventionalism about DME

I propose to extend the conventionalist treatment of counterpossibles with metaphysically and (meta)logically impossible antecedents to counterpossibles with mathematically impossible antecedents. Thus, I take non-vacuous countermathematicals to express consequences of changes in the rules governing mathematical concepts.[20] For example, return to the countermathematicals:

> Were the trefoil knot isotopic to the unknot, Terry's trefoil knot would have been isotopic to the unknot.

---

[20] Here I will not rely on any particular account of the distinction between mathematical and non-mathematical concepts, which should not matter for my argument. The distinction may turn out to be disjunctive—a mathematical concept is either an arithmetical concept or a geometrical concept or. . . , where an arithmetical concept is a concept of quantity, a geometrical concept is a concept of space, etc. I do not think it is necessary for my argument that there should even be a clear distinction.

> Were the trefoil knot isotopic to the unknot, Terry's untieable knot would have been trefoil.

The conventionalist should interpret these as expressing something like:

> Were the semantic rules governing the application of the term 'trefoil knot' such that, wherever it applied, 'isotopic to the unknot' also applied, Terry's trefoil knot would have been isotopic to the unknot.

> Were the semantic rules governing the application of the term 'trefoil knot' such that, wherever it applied, 'isotopic to the unknot' also applied, Terry's untieable knot would have been trefoil.

where we consider actually adopting the semantic rule specified in the antecedent. These conventionalist interpretations should preserve the original countermathematicals' truth values. Thankfully, it seems that they do—the first counterfactual above is (weakly) necessarily true and the second is not. That is, in every world where Terry has a trefoil knot, the first counterfactual, but not the second, is true. Remember that we are to consider actually adopting the semantic rule specified in the antecedent: were we to imagine actually adopting the semantic rule that wherever 'trefoil knot' applies, 'isotopic to the unknot' applies, then 'isotopic to the unknot' would have applied to the knot of Terry's to which 'trefoil' applies; so, via semantic descent, Terry's trefoil knot would have been isotopic to the unknot.

I just characterized these conventionalist interpretations *modally*: the first is (weakly) *necessarily* true and the second is not. The necessity involved here does not seem to be mathematical necessity, so the *mathematical* conventionalist *need* not say that it expresses a semantic

rule, but I will suggest a way to say just that: the first counterfactual, but not the second, is a consequence of actual semantic rules governing the terms therein. Given that the first countermathematical is similar to a case of universal instantiation (i.e., "if for all x, x is F, then a is F"), I suggest that it follows from semantic rules governing the logical terms involved, such as "wherever" and/or the counterfactual conditional "if… were…, then… would be…".[21]

What about clause b) of NOCA? This clause deals with necessitation, and though it is unclear whether this is *mathematical* necessitation, I will suggest a way to say that such necessitation claims express semantic rules. According to NOCA, the fact that Terry's trefoil knot is distinct from the unknot necessitates that he will fail to untie his trefoil knot. Necessitation claims are usually cashed out as necessary conditionals: necessarily, if Terry's trefoil knot is distinct from the unknot, then he will fail to untie his trefoil knot. That conditional is arguably an expression of actual semantic rules governing the terms therein—it expresses what Thomasson class the 'analytic entailment' of the consequent by the antecedent.

Pincock's account of abstract explanation appeals to instantiation. Instantiation (or perhaps some kind of realization) is the relation that many Platonists take to hold between a mathematical object (or property) and the concrete objects (or properties) that are its instances. Instantiation is appealed to by both Pincock and I, and we both take it to be an ontic relation between abstract mathematical and concrete objects. For the conventionalist, instantiation should be seen not as an ontic relation but as a semantic one. To say that some concrete

---

[21] Does this commit me to a kind of logicism? I do not think so. I would only be worried if I were committed to the claim that all purely mathematical truths are expressions of rules governing logical concepts. I certainly am not committed to that; and note that the countermathematical in question is not purely mathematical.

object instantiates a mathematical object is just to say that the relevant mathematical concept applies to it. Instantiation is concept application.[22] Notice that the relation of concept application has features that are important for the explanatory aims of Pincock and I. They both rely on the asymmetry of instantiation to buttress their theories' explanatory credentials and exclude certain reversals and other potential counterexamples. Concept application too is an asymmetric relation[23]: 'trefoil knot' (or 'almost minimal set') applies to some concrete object, but that concrete object does not apply to 'trefoil knot' (or 'almost minimal set'). On NOCA, the mathematical fact that the trefoil knot is distinct from the unknot does not explain why the knot Terry untied is not trefoil because that mathematical fact is not instantiated in that natural fact.[24] According to the conventionalist interpretation of NOCA, the mathematical fact that the trefoil knot is distinct from the unknot does not explain why the knot Terry untied is not trefoil because the mathematical concept "trefoil knot" does not apply to anything in that natural fact (for the same reason that the trefoil knot is not instantiated in that natural fact—there is no trefoil knot there!). On Pincock's account, abstract explanations show that a concrete object possesses a certain property because it is an instanti-

---

[22] I cannot here say anything about what I take the relation of concept application to be (though I am sympathetic to Thomasson, 2007b). I do not think my arguments require any particular account of that relation, though obviously an across-the-board nominalist (not merely a nominalist about mathematical objects) will want an account that does not appeal to abstract objects.

[23] At least in the relevant cases, such as in DMEs, where concept application is intended to take the place of the instantiation of an abstract object by a concrete object. As Earl Conee (personal communication) pointed out to me, perhaps the application of the concept *concept* to itself is not asymmetric. However, I do not think this is a case where we would say a concrete object instantiates an abstract object.

[24] As I have already noted, NOCA does not need the instantiation relation to exclude this reversal—the counterfactual clauses of NOCA already do that. I suggested that the instantiation relation is what makes true the counterfactual clauses of NOCA.

ation of an abstract object that possesses a that property. According to the conventionalist interpretation of Pincock's account, in abstract explanations some predicate is shown to apply to a concrete object in virtue of the fact that its predication is analytically entailed by the predication of a mathematical predicate to it. Soap films are shown to satisfy Plateau's laws because they instantiate almost minimal sets, which instantiate Plateau's laws. According to the conventionalist interpretation, soap films satisfy Plateau's laws because 'almost minimal set' applies to them, and the application of 'almost minimal set' to a concrete object analytically entails the application of 'satisfies Plateau's laws' to it.

Let us take stock so far. I presented Pincock's (2015) and my (Povich, 2021) ontic accounts of DME. I then explained conventionalism and showed how it can accommodate non-vacuous countermathematicals.[25] Finally, I gave a semantic account of the instantiation relation as concept application.

Before addressing the question of whether conventionalist interpretations of Pincock's abstract dependence account and NOCA are still ontic, let me explain how the conventionalist can escape Kuorikoski's objection. Recall that the challenge is to adhere to the same-object condition. For conventionalists, the same object is the term/concept, individuated syntactically, merely with a different meaning/content. Of course, this means that Kuorikoski is right that countermathematicals are importantly different from standard ontic counterfactuals, and that there is something more "representational" about countermathematicals—this shouldn't be surprising since, after all, conventionalists take mathematical truths to express rules for the use of language—but also note how for the conventionalist counter-

---

[25] Baron (2020) presents an account of DME that relies on countermathematicals, but it is unclear if he takes his account to be an ontic one.

mathematicals are importantly *different* from the clearly *epistemic* counterfactuals with which Kuorikoski contrasts ontic counterfactuals, such as his Sisley example (Kuorikoski, 2021, p.196). We are to imagine that a museum has a policy that all and only Sisleys are hung in room 18. The counterfactual "If this painting were in room 18, then it would be a Sisley" is false when read ontically but true when read epistemically as a claim about what it would be rational to believe if the antecedent were true. But counterconceptual interpretations of countermathematicals are *not* epistemic claims like this, for what is true according to a convention is not an epistemic matter. The distinction between ontic and epistemic readings of counterfactuals cuts across the distinction between countersubstratum and counterconceptual readings of counterfactuals. The counterconceptual reading of, e.g., "Were the trefoil knot isotopic to the unknot, Terry's trefoil knot would have been isotopic to the unknot" does not concern what it would be rational to believe were the trefoil knot were isotopic to the unknot, nor what it would be rational to believe were the semantic/conceptual rules governing the application of the term/concept 'trefoil knot' such that, wherever it applied, 'isotopic to the unknot' also applied. Neither of those epistemic questions involves a shift in conceptual scheme or convention; in that sense epistemic readings are like countersubstratum readings. The counterconceptual reading concerns what is true according to the convention specified in the antecedent. This leads us directly to our central question—whether conventionalism strips accounts of DME of their ontic status.

6. The Ontic Status of Conventionalist Accounts of DME

We have seen that counterconceptual readings of countermathematicals are not epistemic in Kuorikoski's sense; they do not concern what it would be rational to believe. But the question still remains whether conventionalism deprives ontic accounts of DME of their ontic status. Here I present an argument that it does not. First, I need to be clear about the ontic. By "ontic," I mean mind-independent, in the sense that whether the explanans, explanandum, and explanatory relation between them exist is not up to *us explainers*. But I do not think this should rule out conventions as ontic. I don't think the definition of "ontic" should rule out the possibility of ontic explanation in the cognitive and social sciences, including sociology and linguistics; brains, beliefs, social (including linguistic) conventions, and so on are perfectly objective in the sense that matters for ontic explanation, and only on exceedingly controversial and exceedingly rare philosophical views can such things not enter into causal or other natural relations. Brains and linguistic conventions are in principle scientifically manipulable and apt to figure in causal explanations, as explanantia and as explananda.

To illustrate the ontic status of conventionalist explanations, let us simply think about Woodwardian (2003) interventionism from the conventionalist perspective, using the distinction between counterconceptual and countersubstratum readings of counterfactuals. Take the countermathematical: "Were the trefoil knot isotopic to the unknot, Terry's trefoil knot would have been isotopic to the unknot". The conventionalist says we should interpret this as expressing something like: "Were the semantic/conceptual rules governing the application of the term/concept 'trefoil knot' such that, wherever it applied, 'iso-

topic to the unknot' also applied, Terry's trefoil knot would have been isotopic to the unknot". Read in the usual, countersubstratum way, this is simply false. Read counterconceptually, it is true. But note that this can be interpreted in terms of Woodwardian interventions, from which many ontic proponents, including Pincock and I, take our inspiration. Thus, imagine an intervention on the concept/term "trefoil knot" that changes the semantic/conceptual rules governing it. I take it that such an intervention would amount to an intervention on people's brains or on their social conventions (an intervention which would again presumably have its intended effect via changes in people's brains). How exactly this could work depends on the metaphysics of concepts, which I cannot get into here. I will just note two important things about this suggestion: 1) Woodward does not require that interventions be physically possible, so difficulty in imagining what this would look like in practice is no objection to it. 2) We need to individuate terms/concepts syntactically, or some other way such that changes in the rules governing the term/concept do not change the term/concept itself. As I mentioned above when addressing Kuorikoski's objection, the 'same object' here is a syntactically individuated term/concept. (Terms/concepts are individuated this way for Chalmers' (2004, pp.169–170) orthographic contextual intensions and Stalnaker's (1978; 2001) diagonal propositions.) Otherwise, we will not have the same term/concept pre- and post-intervention.

Our intervention would change which claims the people upon whom we intervened make and which beliefs they have—they would now assert that Terry's trefoil knot is isotopic to the unknot. Of course, we, the interveners, using our actual semantic rules, would not say that, post-intervention, Terry's trefoil knot is isotopic to the unknot. We would say that Terry's trefoil knot is still distinct from the unknot, and we would take our intervention merely to have demonstrated

a causal or mechanistic relation between their brain states or social conventions and what they think and say. Of course, that is true, but the conventionalist about mathematics can say more. If *we* were actually to adopt their post-intervention semantic rules, *we* would say that Terry's trefoil knot is isotopic to the unknot. Counterconceptual readings of interventionist counterfactuals show that there is a kind of "counterconceptual causal" dependence here—a dependence that one can see only by switching semantic rules.[26] The idea here is that x counterconceptually depends on y just in case the counterfactual "were ∼y the case, then ∼x would be the case" is true on a counterconceptual reading. So, since the counterfactual, "Were the semantic/conceptual rules governing the application of the term/concept 'trefoil knot' such that, wherever it applied, 'isotopic to the unknot' also applied, Terry's trefoil knot would have been isotopic to the unknot" is true on a counterconceptual reading, the fact that Terry's trefoil knot is distinct from the unknot counterconceptually depends on the semantic/conceptual rules governing the application of the term/concept 'trefoil knot'. If we think of the antecedent as brought about by an intervention, à la Woodward, then there is "counterconceptual causal" dependence. When we give the previous counterfactual a counterconceptual reading and conceive the antecedent as brought about by an intervention, it is true. Thus, the fact that Terry's trefoil knot is distinct from the unknot "counterconceptually causally" depends on the semantic/conceptual rules governing the application of the term/concept 'trefoil knot'.

---

[26] Craver (2007; Craver, Glennan and Povich, 2021) adjusts Woodward's interventionism to give an account of mechanistic/constitutive, rather than causal, relevance. Perhaps it would be better to say that there is a "counterconceptual mechanistic" dependence here, depending on what is intervened upon (e.g., the brain or social conventions).

It is important to note that the view is not that Terry failed to untie his trefoil knot because the way the mathematical concepts "trefoil knot" and "unknot" are used. The conventionalist can recognize that falsity of that claim just as she can recognize the falsity of the standard reading of the counterfactual, "If the concept 'trefoil knot' were used differently, Terry would've untied his trefoil knot". Nevertheless, the view is that "Terry failed to untie his trefoil knot because the way the mathematical concepts 'trefoil knot' and 'unknot' are used" is getting at *something* important in a *roundabout* way, a way which was the purpose of this section, and the concept of counterconceptual dependence, to explicate.[27] On an account like Lange's or NOCA, the explanandum of a DME is "narrow" or as Lange puts it, the empirical explanans is presupposed in the context of the why-question. This makes the explanandum-statement basically analytic for a conventionalist. Suppose we want to explain why Terry failed to make his triangle four-sided or failed to make his sister a bachelor, say by widowing her. In these cases, I think it is uncontroversial that it would be *adequate* for an explanation of Terry's failure to cite *only semantic conventions*. (See Donaldson, 2021 for a defense of this kind of idea.) One *needn't* cite semantic conventions; one could also explain his failure by appeal to the fact that bachelors are (necessarily) men. But for the conventionalist that is simply an expression of a conceptual rule and the explanation that cites the rule directly is adequate on its own. I submit that on an account like Lange's or NOCA any impression that DMEs are different than this is an illusion. On these views, one is simply metaphysically confused if one has in mind some

---

[27] I think the "something important" is also brought out by similar work on conventionalism and analyticity (e.g., Topey, 2019; Warren, 2020; Donaldson, 2021). These authors, each in their own way, argue that there are some non-linguistic facts (e.g., those expressed by analytic truths) that can be explained by convention, contra opponents of truth by convention (e.g., Boghossian, 1996).

more metaphysically robust explanandum for DMEs. (Compare the objection: "But you can't explain the *FACT* that Terry failed to make his sister a bachelor by appeal to only semantic conventions!" This betrays a confusion about the metaphysical lightness of the fact itself. See again Topey (2019), Donaldson (2021), and Warren (2020).)

On other accounts of DME, like Pincock's, the explanandum is not narrow, so its description is not analytic. However, the conventionalist will say similar things about the mathematical premises in Pincock's account. Soap films satisfy Plateau's laws because they instantiate almost minimal sets and it is a mathematical fact that almost minimal sets satisfy Plateau's laws. The conventionalist can agree that the fact that soap films instantiate almost minimal sets (or that "almost minimal set" applies to them) is an empirical, non-conventional fact. However, for the conventionalist, the mathematical fact that almost minimal sets satisfy Plateau's laws is an expression of conceptual rules. The explanatory status of this fact has the same two-faced character as the one discussed in the previous paragraph. The conventionalist 1) can accept that this mathematical fact partly explains the explanandum, and 2) can hold that "soap films satisfy Plateau's laws in part because of how terms are used" is false on the standard reading of that claim, yet 3) can hold that it is true on a counterconceptual reading. For the conventionalist, DMEs on Pincock's account are no different than the following: why is Bob an unmarried man? Because Bob instantiates the property of being a bachelor and bachelors are unmarried men. The fact that Bob is (or instantiates the property of being) an unmarried man is an empirical, non-conventional fact. "Bachelors are unmarried men" is an expression of conceptual rules. The explanation would be just as adequate if it appealed to a *semantic* fact here: because Bob

is a bachelor and "bachelor" means *unmarried man*. According to the conventionalist, DMEs on a Pincockian view are no different than this.

Note that accepting that this kind of counterconceptual dependence is explanatory seems not to require any significant revision in our ordinary concept of explanation. Kocurek, Jerzak, and Rudolph (2020, p.7) point out that there are many times when we accept that counterconceptual dependence is explanatory. They give the following nice example. In 2006, the International Astronomical Union's (IAU) revised the scientific definition of 'planet'. According to this new definition, Pluto is no longer classified as a planet. Kocurek et al. maintain that the following claims are literally true:

> Whether or not Pluto is a planet *depends* on what definition the members of the IAU agree on.
>
> Part of what *explains* why Pluto is not a planet is the IAU's decision in 2006 to redefine 'planet'.
>
> *Because* of the IAU's decision in 2006, Pluto is not a planet. (Kocurek, Jerzak and Rudolph, 2020, p.7, my emphasis)

Kocurek et al. consider a Gricean attempt to explain this away. According to the Gricean, these claims express literal falsehoods, and we should instead understand them as communicating something explicitly metalinguistic, e.g., "Part of what explains why Pluto is not classified as a planet is the IAU's decision in 2006 to redefine «planet»" (2020, p.7). Kocurek et al. argue in response that "[t]he defender of this line owes us a theory of how these utterances are transformed into explicitly metalinguistic ones. We think that the prospects for such a theory are not good because the exact nature of the transformation into an explicitly metalinguistic sentence is highly unsystematic"

(2020, p.7). They go on to defend this last claim, but we needn't continue it here. My point is just that accepting counterconceptual dependence as explanatory doesn't seem to do significant damage to our ordinary concept of explanation. One might try to argue that a proper philosophical explication of the ordinary concept should exclude counterconceptual dependence, but I have argued here we have good reasons for including it.

Now, one might claim that this isn't *really* an ontic account of DME. On ontic accounts, the explanandum, explanans, and the dependence relation between them are distinct, *ontic* things. Yet, what I described in the previous paragraph is merely a case where the *same* fact is *described* or *conceptualized* in two different ways. It is not a case where the explanandum ontically depends on some other fact: nothing about Terry's knot or the soap films really changed, only what people think and say about them. As I conceded in the last section, I think there is something right about this, namely, that the conventionalist views pure mathematics as expressing rules for the use of language. This is especially on accounts that narrow the explanandum-statements until they are analytic, like Lange's or NOCA. (On a conventionalist interpretation of Pincock's account, there is still the empirical, non-conventional fact that soap films instantiate almost minimal sets [or that "almost minimal set" applies to them] that plays an explanatory [causal] role.) Still, I do not think the ontic proponent need fear. First, the hypothetical intervention into peoples' brains or social conventions or whatever clearly *is* one to which no ontic proponent would object—it plainly illustrates an ordinary ontic (causal or mechanistic) explanation of what people think and say. Second, according to the conventionalist, there *just is not anything else here to explain*. Mathematics is just a reflection of how people talk—shadows of our syntax (Warren, 2020). So, everything there *is* to explain can be ex-

plained ontically. No worry for the ontic proponent, then. I think those who would object that conventionalist accounts are not really ontic are really objecting to conventionalism, they are objecting that there must be something more to explain.

I know that I have not conclusively established that conventionalist accounts of DME are ontic, but I hope to have convinced you that the claim that they are isn't as obviously false as it might have seemed.[28]

7. Objections

One objection to my conventionalist treatment of countermathematicals is that it only works for certain explananda, namely those that only depend on a mathematical fact, and not for cases like the well-known magicicadas or hexagonal honeycombs (Lyon and Colyvan, 2007), whose explananda also depend on natural facts. However, the countermathematicals in these cases work the same way as the others.

We need a conventionalist evaluation procedure for countermathematicals. We could, like Baron et al. (2017), hold fixed the "morphism" between mathematical structures and empirical structures,

---

[28] Here is another brief argument for the ontic status of conventionalist DMEs. There are three extant conceptions of explanation: the ontic, the modal, and the epistemic (Salmon, 1989). Conventionalist DMEs certainly don't seem to fall into a modal or epistemic conception, for they don't show that their explananda had to occur, nor that they were expected to occur. By "epistemic conception," some just mean that explanation is a representational act. Nothing I've said here disagrees with that. For, by "ontic conception" I don't mean that explanations are themselves ontic; I just mean that they explain by appeal to the ontic (see Craver, 2014). Since conventionalist DMEs aren't modal or epistemic, they are ontic. Of course, this argument relies on there being only three conceptions of explanation. I challenge those who don't think conventionalist DMEs are ontic to explain what they are and why.

so that changes in mathematical structures have ramifications into empirical structures. For a conventionalist, this would mean holding fixed that the concept in question applies, ramifying into the world accordingly, i.e., we imagine a world where our changed concept still applies. I don't think this idea damages the conventionalist picture in any way, but I think we can do something much simpler, without talk of worlds altogether, which seems to track better how we actually reason through countermathematical scenarios. Consider the following evaluation procedure: a countermathematical is true when and only when[29] the new rule of inference expressed in the antecedent licenses the derivation of the consequent from given empirical (and other unchanged mathematical) background premises. Obviously, other inference rules that have not been changed are also allowed in the derivation. This is similar to what Lewis (1973) called the metalinguistic theory of counterfactuals.[30]

This evaluation procedure works for the countermathematicals at issue in DMEs regardless of whether the explananda are narrowed. Take the countermathematical 'if 13 hadn't been a prime number, then

---

[29] "When and only when" means just that—this is just your standard biconditional. I am not here giving truth-conditions for countermathematicals. That would imply that I take countermathematicals to *mean* something about rules of inference and derivability, but I don't. Nor am I giving truthmakers for countermathematicals.

[30] Unlike most other defenders of metalinguistic theories, I do not prefer it because I have some metaphysical problem with possible worlds. Conventionalists let a thousand languages bloom. And I don't intend to commit myself to the linguistic ersatzist view that possible worlds just *are sets* of sentences or something of that sort, though what I say is consistent with such a view. (Even the modal realist will admit that to each world there corresponds a unique set of propositions describing it—the linguistic ersatzist simply claims that the correspondence is identity (Bennett, 2003, p.303).) I prefer this metalinguistic theory of countermathematicals not for metaphysical reasons but because it doesn't require the complication of holding fixed that the changed concepts apply, it seems to describe better what we actually do when evaluate countermathematicals, and I feel it just comports better with the conventionalist idea that mathematical truths express rules of inference.

North American cicadas wouldn't have had 13-year life-cycles'. In the present case, we imagine actually adopting the rule that 'prime' does not apply to anything '13' applies to. This can be seen as a rule of inference for descriptions containing 'prime' and '13'. Since the consequent follows from given background premises (e.g., that having a life-cycle period that minimizes intersection with other periods is evolutionarily advantageous and that prime periods minimize intersection) using the new rule of inference expressed in the antecedent, it is true on this evaluation procedure that the cicadas wouldn't have had 13-year life-cycles. Let us examine the reasoning of Baron et al. (2017, p.11) regarding this countermathematical and show that this is exactly what they are doing—using the mathematical claim in the antecedent as an inference rule to reach the consequent via given background premises—their platonistic excesses are just that. They are concerned to show specifically the truth of the countermathematical, "If, in addition to 13 and 1, 13 had the factors 2 and 6, North American periodical cicadas would not have 13-year life cycles". Here is their reasoning:

> To evaluate this counterfactual, we start in the mathematics. [1] We hold fixed as much as we can by changing multiplication to behave like multiplication*. This leaves 13's factors as desired. This gives us a structure, S*, that is just like the natural numbers, except that 13 is not prime, and factorises via 2 and 6. [2] Because we are holding fixed the relationship between the mathematical and physical structures, the physical structure that is now being mapped onto S* must twist to keep up with the counterfactual change. [3] The result is that an interval of 13 years is now divisible into six two-year segments, or into two six-year segments. [4] It follows from this that a cicada with a 13-year life cycle will overlap with predators that have two-year and six-year life cycles and [5] thus that 13

> is not an optimal way to avoid predation. [6] So cicadas won't
> evolve 13-year life cycles. [7] So [the countermathematical] is
> true. (11)

The first three claims, which I've collectively labeled [1], are, accord-
ing to the conventionalist, simply telling us to imagine adopting a new
inference rule according to which 13 has the factors 1, 2, 6, and 13,
while leaving all other inference rules unchanged. Claim [2] is the
morphism claim we discussed above, the conventionalist analog of
which would be holding fixed that the concept in question applies.
Since we are here illustrating a different evaluation procedure that
relies simply on descriptions and not on worlds, we can ignore claim
[2]. Claim [3] is simply an application of our new inference rule: since
13 has the factors 1, 2, 6, and 13, an interval of 13 years is divisible
into six two-year segments, or into two six-year segments. Claim [4]
is inferred from [3] using normal inference rules that have not been
changed, which is fine since [1] tells us that only one inference rule
has changed. Similarly, claim [5] is inferred from [4], and claim [6] is
inferred from [5] and empirical background premises (e.g., that noth-
ing suboptimal will evolve), using normal inference rules that have
not been changed. They conclude [7], that the countermathematical
is true. Thus, they have concluded that the countermathematical is
true because its consequent can be inferred from given background
premises using the inference rule specified in the antecedent (and any
other unchanged inference rules).

The same points apply, *mutatis mutandis*, to the hexagonal hon-
eycombs and other cases. For example, if the structure that divides
a planar region into regions of equal area using the least total perime-
ter were not a hexagonal grid, then the honeybees' combs would
not have been a hexagonal grid. Here we imagine adopting the rule
that 'hexagonal grid' does not apply to anything 'structure that di-

vides a planar region into regions of equal area using the least total perimeter' applies to. Again, since the consequent follows from given background premises (e.g., that producing the largest honeycomb cells using the least wax is evolutionarily advantageous) using the new rule of inference expressed in the antecedent, it is true on the metalinguistic evaluation procedure that the combs would not have formed a hexagonal grid. In general, whenever mathematical necessities appear ineliminably in a scientific explanation, they play the normative role of making explicit the conceptual norms linking the mathematical concepts applied in its empirical explanans-statement(s) to mathematical concepts applied in its empirical explanandum-statement. That is their function as expressions of rules of inference, rules for transforming empirical descriptions.

The metalinguistic theory of counterfactuals faces a notorious problem: the problem of cotenability (Goodman, 1947). Consider the counterfactual, "If this match had been struck, it would have lit". According to the metalinguistic theory, this is true if and only if "it lights" can be derived from "this match is struck". Obviously, this derivation doesn't work without further premises. But what further premises is it legitimate to include? Certainly allowed are laws of nature and premises that are implicit in the context of the conversation we are having. And equally certainly, we cannot allow the truth that the match was not struck. That would generate a contradiction and, assuming classical logic, every consequent would follow. Goodman (1947) argued that cotenability with the initial premise (i.e., "this match is struck") was a condition for inclusion into the further premises, where a sentence S is cotenable with the initial premise P if and only if it is not true that if P were true, then S would be false. Of course, he knew

that this was circular, since the definition of cotenability was given in counterfactual terms. The problem of cotenability is to provide a definition that isn't in counterfactual terms.

My metalinguistic approach to countermathematicals may avoid this problem because we are only changing a rule of inference. We are not changing any premises; we are using actual, rather than counterfactual, premises. Obviously, we need to know what the given premises are, and there are decisions that need to be made about what premises and rules of inference to hold fixed in counterconceptual scenarios, but these decisions exactly parallel the decisions about what to hold fixed in any account of countermathematical reasoning (e.g., the decisions Baron et al. must make about what ontic facts and morphic relationships to hold fixed). I want to emphasize that I am merely showing what the conventionalist must hold fixed to make the countermathematicals come out true. As Baron et al. (2017, p.12) note, "To ask whether it is reasonable to hold these facts fixed when evaluating counterfactuals is to call into doubt the truth of the counterfactuals at issue". However, even if this does not avoid the problem of cotenability, I am not trying to give a reductive account of counterfactuals generally, so I have no problem relying on counterfactuals to explicate countermathematicals. Lewis (1973, p.69) offers a possible worlds solution to the problem. According to Lewis, "χ is cotenable with an entertainable antecedent φ at a world i if and only if χ holds throughout some φ-permitting sphere around i". Defining cotenability this way makes the metalinguistic approach logically equivalent to the possible worlds approach (Lewis, 1973, p.69). Perhaps the conventionalist wouldn't have a problem adopting this so long as the talk of possible worlds can be understood deflationarily. Again, though, I don't think there is any problem with the evaluation procedure where we hold fixed that the changed concept applies.

Another objection is that counterconceptual readings of counter-mathematicals incorrectly make the dependence of natural fact on mathematical fact into a dependence of the meanings of descriptions of natural fact on the meaning of descriptions of mathematical fact. I think there is something right about this objection, but it is obviously question-begging: it assumes anti-conventionalism—it assumes that pure mathematics offers substantive descriptions of mathematical facts. What is right about it is something I do not take to be objectionable: that mathematical truths express rules of description—semantic rules—and there is nothing more to them. That is objectionable to many, but it is just conventionalism.

Another objection is that conventionalism makes a mystery of why everyone in the world adopts the same semantic conventions and why mathematics has any explanatory power (in the standard, "non-distinctive" sense not at issue in the DME debate). First, the objection would prove far too much if it were correct. I do not think either the agreement on, or the explanatory power of, mathematics is any less mysterious according to Platonism or any other (non-empiricist[31]) anti-Platonism. Take fictionalism, for example. The objection applies equally to it—fictionalism makes just as much a mystery of why everyone in the world adopts the same fictions and why some of these mere fictions have explanatory power. Second, and more substantively, there are a few direct answers the conventionalist can give as to why mathematics has explanatory power and why everyone in the world adopts the same semantic rules—though, importantly, note that there does exist disagreement in mathematics, just as in logic (e.g., Balaguer, 2017; Beall and Restall, 2006; Davies, 2005; Priest, 2013;

---

[31] The empiricist could explain (some of) the agreement of mathematicians by appeal to the empirical regularities to which all mathematicians have access and that, according to them, (at least basic arithmetic and geometrical) mathematical truths describe.

2021) Some answers to the latter question can and have been given, mutatis mutandis, by the fictionalist. For example, as Colyvan (2011) notes, the fictionalist can appeal to constraints on writing the fiction of mathematics, such as that "new installments" (i.e., theories) in the fiction be self-consistent, consistent with past installments, and not introduce unnecessary "characters" (i.e., entities). The conventionalist can appeal to these as constraints on the creation of semantic rules as well. The conventionalist (and fictionalist) can also appeal to a shared (culture- or species-specific) aesthetic sense (see Steiner, 1998 for a provocative discussion of the role of aesthetics in mathematical theorizing). Mathematicians whose proposed semantic rules fail to meet these constraints are sanctioned by the mathematical community, inducing further agreement. Finally, the conventionalist can also explain agreement, at least in basic arithmetic and geometry, in the way that the empiricist does—by appeal to empirical regularity.[32] This is an idea prominent in Wittgenstein, who argued that the propositions of basic arithmetic and Euclidean geometry were empirical generalizations "hardened into rules" (i.e., rules of inference) and "put in the archives" (i.e., made immune from empirical refutation) (Bangu, 2018; Steiner, 1996; 2009; Wittgenstein, 1976; 1978).[33]

Regarding the standard, 'non-distinctive' explanatory power of mathematics, it is perfectly consistent for the conventionalist to say that (many) mathematical concepts have empirical content and that applied mathematical propositions are straightforwardly descriptive. Conventionalism is a theory of mathematical modality, not of the content of mathematical concepts. Conventionalism is thus compatible

---

[32] If Maddy (1990) is right that we can perceive some sets, perhaps some of basic set theory can also be accounted for this way.

[33] Perhaps this accounts for Kant's judgment that "7 + 5 = 12" is synthetic *a priori* (Kant, 1781/1787, B15).

with the claim that mathematical concepts have (or can have, after suitable empirical interpretation) empirical, descriptive content[34]" and that this content contributes to mathematics' (non-distinctive) explanatory power by mapping (Pincock, 2011; Bueno and French, 2018), indexing (Melia, 2000), or representing (Saatsi, 2011) explanatorily relevant quantities, magnitudes, etc. If this seems strange, consider a comparison. Conventionalism about metaphysical modality is compatible with the claim that empirical concepts with descriptive content can figure in necessary truths. 'Bachelor' is a concept with empirical, descriptive content and "bachelors are unmarried men" expresses a semantic rule governing it. Similarly, 'triangle' is a concept with empirical, descriptive content and "triangles have three sides" expresses a semantic rule governing it. Conventionalism about mathematical modality does not rob mathematical concepts of their explanatory power. Furthermore, mathematical semantic rules need not be 'arbitrary,' as evidenced by the Wittgensteinian idea mentioned above that basic arithmetic and geometric truths are empirical generalizations "hardened into rules".[35] Perhaps there remains for the conventionalist some version of Wigner's (1960) problem of the "unreasonable effectiveness of mathematics"—though I think the Wittgensteinian idea goes a long way to dispelling this—but this is a problem for everyone, and here I can only refer the reader elsewhere (see e.g., Steiner, 1998; Bangu, 2006; Clark, 2017; Bueno and French, 2018)

---

[34] Compare Waismann's (1986, p.66) description of Russell's position: "For Russell the propositions of mathematics are, to be sure, *a priori*—they are tautologies—but the concepts are empirical.

[35] Paul Audi (personal communication) helpfully suggested another sense in which semantic rules generally are non-arbitrary: presumably, the *reason* that, e.g., "unmarried" applies if "bachelor" applies, is that the features of the *world* in virtue of which the former applies are a subset of those in virtue of which the latter applies. A similar point is made by Thomasson (2007b, p.70).

Next, there is the worry that deflating instantiation by treating it as expressing facts about concept application results in too many things being counted as DMEs.[36] For example, we want to explain why Claire has 5 apples. Because she has 2+3 apples and 2+3=5. The narrow explanandum would be the fact that Claire, who has 2+3 apples, has 5 apples. This weakly necessarily counterfactually depends only on the mathematical fact that 2+3=5; if 2+3 were not equal to 5, then Claire, who has 2+3 apples, would not have had 5 apples. Thus, we have a DME of why Claire has 5 apples. I can't appeal to instantiation qua concept application to exclude this case, since the concepts '5' and '2+3' both apply in this scenario. Thus, there are as many DMEs as there are equations.

This is a great example, but I think the objection misses the mark. Let me note four things. First, it is not obvious to me that explanations like this are always bad. It seems like this would be a good explanation for someone who didn't know that 2+3=5, although I admit that perhaps my intuitions are conflating explanation and evidence. Second, the truth of the relevant countermathematical doesn't depend on its being read counterconceptually nor on my semantic account of instantiation. For example, I don't see why it wouldn't be true according to Baron, Colyvan, and Ripley's (2017) platonistic evaluation procedure. Third, why might some say 2+3 can't be (inflationarily) instantiated? Presumably they would say it's because 2+3 is not a mathematical object. But why? Don't '2+3' and '5' both refer to the same object? After all, that's when an identity statement is true—when the expressions flanking the identity symbol refer to the same object. 2+3 is not a strange conjunctive object composed of 2 and 3; it *is* 5. 2+3 is instantiable, because 5 is instantiable and 2+3 is identical to 5 (via identity elimination or Leibniz's law—the indiscernibility of identicals, not

---

the identity of indiscernibles). So, non-conventionalists will also have to accept the instantiability of 2+3. Could non-conventionalists argue that 5, and, so, 2+3, isn't instantiable? But then what is the relation between the number 5 and Claire's 5 apples? If it isn't instantiation, call it "shminstantiation". Clearly shminstantiation is a relation that can figure in DMEs, since many DMEs appeal to numbers to represent various quantities and magnitudes. Surely, we don't want to say there are no DMEs that appeal to numbers. Fourth, because of the last two points, many other accounts seem to render this a DME too. If Mary's having 23 strawberries is constitutive of the physical task or arrangement at issue in Lange's (2013) famous example, presumably having 2+3 apples is in this case, so Lange's account counts it as a DME. Perhaps Lange could say that having 5 apples can be constitutive of a physical task or arrangement at issue, but having 2+3 apples can't. It's hard to see how that could be, given that having 5 apples and having 2+3 apples are identical facts.[37] Pincock's (2015) abstract dependence account similarly seems to have to accept this as a DME. Claire has 5 apples because her apples instantiate the property of being 2+3 (in quantity), and 2+3=5.

Finally, I may be able to rule this case out by arguing that it is a case of denying one of the why-question's presuppositions, which is something distinct from explanation (Roski 2021). The narrowed why-question presupposes that Claire's having 2+3 apples and her having 5 apples are distinct facts, and the putative explanation undermines

---

[37] Lange (2016, pp.xviii–xix) mentions identity explanations favorably: "that Samuel Clemens and Mark Twain are identical explains non-causally why they have the same height, weight, and birth dates". Similarly, Lange might accept that there is some context where the fact that Claire's 2+3 apples are identical to her 5 apples explains non-causally why they have the same mass, price, etc. However, Kim's (2011, pp.104–105) arguments against identity explanations in the philosophy of mind may be relevant here.

this. The why-questioner thus gains understanding, certainly, but this understanding is not explanatory. Furthermore, it is merely in virtue of learning that 2+3=5 that the why-questioner learns that Claire's having 2+3 apples and her having 5 apples are not distinct facts, so the putative explanation succeeds in undermining the presupposition regardless of whether the content of that knowledge (that 2+3=5) is interpreted conventionalistically. In other words, this move doesn't require any particular metaphysics of what the fact that 2+3=5 consists in. However, this move is not available to Pincock because for him the explanandum is not narrowed, i.e., the empirical explanans is not presupposed, so the why-question doesn't presuppose that Claire's having 2+3 apples and her having 5 apples are distinct.[38] I conclude that, if counting this case as a DME is a problem for conventionalism, it's a problem many non-conventionalists seem to have as well.

Finally, Benacerraf (1973) provides one of the most powerful objections to conventionalism.[39] His is a challenge to provide a homo-

---

[38] The move is available to Lange, since he narrows the explananda of DMEs, but it would seem to be inconsistent with his approval of identity explanations. He could still accept the legitimacy of identity explanations if he could show that there are contexts wherein non-identity is not presupposed. This seems implausible though. In his (2022) debate with Roski (2021) over "really statistical" explanations, he says that an indication that *p* is a presupposition of the question "Why is *p* the case?" is that it is pragmatically infelicitous to say "I do not want to assume that *p* is the case. But why is *p* the case?". However, it seems to me similarly pragmatically infelicitous to say "I don't want to assume that Samuel Clemens and Mark Twain are distinct. But why are they so similar?"

[39] Many consider Gödel's incompleteness theorems to provide the most powerful objection to conventionalism. According to some (including Gödel himself), the theorems decisively refute conventionalism and support Platonism. Much ink has been spilled on this, and I merely point the reader to some ideas that might be helpful to the conventionalist (Moore, 1999; Floyd and Putnam, 2000; Stalnaker, 2001; Awodey and Carus, 2004; Berto, 2009; Lampert, 2018; Warren, 2020).

geneous semantics for mathematical and non-mathematical discourse. A homogeneous semantics would treat the following two sentences as both having the logical form of the third:

1. There are at least three large cities older than New York.
2. There are at least three perfect numbers greater than 17.
3. There are at least three FG's that bear R to a (Benacerraf, 1973, p.663).

I believe Creath (1980) has mostly adequately addressed this problem and that he is right that Benaceraff begs a central question in demanding a substantive referential conception of truth for a homogeneous semantics. I return to this below. Here I argue that the conventionalist can agree that sentences 1 and 2 have the logical form of sentence 3. Conventionalists tend to be deflationists. Deflationists of the kind I have mind (Carnap, 1950; Schiffer, 2003; Thomasson, 2014; see also Price, 2011) have argued that, e.g., a proposition like 'It is possible that p' analytically entails 'There is a possible world where p,' which analytically entails 'There are possible worlds'. Deflationists take such analytic entailments to have no substantive ontological implications; possible worlds are hypostatizations of our possibility-talk. Deflationists have made similar arguments for other kinds of entity. For example, 'The ball is red' analytically entails 'The ball has the property of being red,' which analytically entails 'There are properties'. Again, such analytic entailments have no substantive ontological implications; properties are hypostatizations of predicates. The conventionalist can similarly say that 'There are three mice' analytically entails 'The number of mice is 3,' which analytically entails 'There are numbers' (see Hale and Wright (2001) for similar arguments, though I do not intend to commit myself to their entire neo-Fregean program). Thus, nothing prevents the conventionalist from saying

that sentence 2 has the logical form of sentence 3. This is analytic; sentence 2 analytically entails sentence 3. In fact, 'There are at least three perfect numbers greater than 17' analytically entails 'Certain mathematical objects stand in a certain relation to each other', just as 'There are at least three large cities older than New York' analytically entails 'Certain cities stand in a certain relation to each other'. This is just what Benaceraff demands of a homogeneous semantics.

Here is where I think Creath (1980) hits the nail on the head: to demand more than this, to demand that the semantics invoke a *substantive* notion of reference, so that the ontology mirrors the semantics, is to beg the question. Conventionalism need not be a primitive expressivism. Just as the sophisticated metaethical expressivist (e.g., Blackburn, 1993; 2005) can speak of moral truths, reference, beliefs, knowledge, assertions, propositions, facts, and descriptions, the conventionalist can speak of mathematical truths, facts, etc. She could even say that a true mathematical proposition describes a mathematical fact, as long as all of these terms are understood in suitably deflationary senses that are clearly distinguished from the senses these terms have in talk of empirical, facts, etc., as I mentioned earlier (Thomasson, 2020; see also Baker and Hacker, 2009).

8. Conclusion

I presented Pincock's (2015) and my (Povich, 2021) ontic accounts of DME. I explained conventionalism and extended it to DME. I proposed counterconceptual readings of countermathematicals, and I gave a semantic account of the instantiation relation as concept application. These resources allow the nominalist to accept the existence of DMEs while denying Platonism, thus blocking the indispensability-inference

from DMEs to Platonism. The conventionalist can also disagree with the *critics* of indispensability arguments who simply deny the existence of DMEs by arguing that the mathematics is merely playing a representational (Saatsi, 2011) or indexing role (Melia, 2000), not an explanatory one. The critics often also mistakenly think that DME entails Platonism. The conventionalist can accept that the mathematics *is* doing something explanatory, and she can even accept ontic[40] accounts of its explanatoriness, such as Pincock's or NOCA, suitably deflated.

There is much work to be done. I do not pretend to have provided a complete defense of conventionalism, nor of the metaontological deflationism that might be required for the conventionalist to answer Benaceraff's objection. All this is attempted in more detail in my forthcoming book (Povich, 2024). I am under no illusion that it is impossible for there to be a distinctively mathematical explanation that conventionalism will not be able to accommodate. What I hope to have convinced the reader of here is that conventionalism at least provides a promising avenue for the nominalist to accept DMEs while denying Platonism, thus undermining the enhanced indispensability argument, and that conventionalist DMEs as still arguably ontic.

Bibliography

Almgren, F.J. and Taylor, J.E., 1976. The geometry of soap films and soap bubbles. *Scientific American*, 235(1), pp.82–93. https://doi.org/10.1038/scientificamerican0776-82.

---

[40] Baron (2020) presents an account of DME that relies on countermathematicals, but it is unclear if he takes his account to be ontic. Conventionalism can be applied to Baron's account as well.

Awodey, S. and Carus, A.W., 2004. How Carnap could have replied to Gödel. In: S. Awodey and C. Klein, eds. *Carnap Brought Home: The View from Jena*. LaSalle, IL: Open Court, pp.203–223.

Baker, A., 2005. Are there genuine mathematical explanations of physical phenomena? *Mind*, 114(454), pp.223–238. Available at: <https://www.jstor.org/stable/3489104> [visited on 12 January 2024].

Baker, A., 2009. Mathematical explanation in science. *The British Journal for the Philosophy of Science*, 60(3), pp.611–633. https://doi.org/10.1093/bjps/axp025.

Baker, G.P. and Hacker, P.M.S., 2009. *Wittgenstein: Rules, Grammar and Necessity: Essays and Exegesis of §§185–242*. 1st ed. John Wiley & Sons. https://doi.org/10.1002/9781444315691.

Balaguer, M., 2017. Mathematical pluralism and platonism. *Journal of Indian Council of Philosophical Research*, 34(2), pp.379–398. https://doi.org/10.1007/s40961-016-0084-4.

Bangu, S., 2006. Steiner on the applicability of mathematics and naturalism. *Philosophia Mathematica*, 14(1), pp.26–43. https://doi.org/10.1093/philmat/nkj002.

Bangu, S., 2008. Inference to the best explanation and mathematical realism. *Synthese*, 160(1), pp.13–20. https://doi.org/10.1007/s11229-006-9070-8.

Bangu, S., 2018. Later Wittgenstein and the Genealogy of Mathematical Necessity. In: K.M. Cahill and T. Raleigh, eds. *Wittgenstein and Naturalism*. 1st ed., *Wittgenstein's thought and legacy*. New York, NY: Routledge, Taylor & Francis Group, pp.151–173. [visited on 13 January 2024].

Baron, S., 2016. Explaining mathematical explanation. *The Philosophical Quarterly*, 66(264), pp.458–480. https://doi.org/10.1093/pq/pqv123.

Baron, S., 2020. Counterfactual scheming. *Mind*, 129(514), pp.535–562. https://doi.org/10.1093/mind/fzz008.

Baron, S., Colyvan, M. and Ripley, D., 2017. How mathematics can make a difference. *Philosopher's Imprint*, 17(3), pp.1–29. Available at: <http://hdl.handle.net/2027/spo.3521354.0017.003> [visited on 13 January 2024].

Beall, J.C. and Restall, G., 2006. *Logical Pluralism*. Oxford; New York: Clarendon Press; Oxford University Press.

Benacerraf, P., 1973. Mathematical truth. *The Journal of Philosophy*, 70(19), pp.661–679. https://doi.org/10.2307/2025075.

Bennett, J., 2003. *A Philosophical Guide to Conditionals*. 1st ed. Oxford: Oxford University Press.

Berto, F., 2009. The Gödel paradox and Wittgenstein's reasons. *Philosophia Mathematica*, 17(2), pp.208–219. https://doi.org/10.1093/philmat/nkp001.

Blackburn, S., 1993. *Essays in Quasi-Realism*. New York: Oxford University Press.

Blackburn, S., 2005. Quasi-Realism no Fictionalism. In: M.E. Kalderon, ed. *Fictionalism In Metaphysics*. Oxford University PressOxford, pp.322–338. https://doi.org/10.1093/oso/9780199282180.003.0012.

Boghossian, P.A., 1996. Analyticity. In: B. Hale and C. Wright, eds. *A Companion to the Philosophy of Language*. Oxford: Blackwell, pp.331–368.

Brandom, R.B., 2008. *Between Saying and Doing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199542871.001.0001.

Brogaard, B. and Salerno, J., 2013. Remarks on counterpossibles. *Synthese*, 190(4), pp.639–660. https://doi.org/10.1007/s11229-012-0196-6.

Bromberger, S., 1966. Why-Questions. In: R.G. Colodny, ed. *Mind and Cosmos: Essays in Contemporary Science and Philosophy*. Pittsburgh: University of Pittsburgh Press, pp.86–111.

Bueno, O. and French, S., 2018. *Applying Mathematics: Immersion, Inference, Interpretation*. Oxford: Oxford University Press.

Carnap, R., 1950. Empiricism, semantics, and ontology. *Revue Internationale de Philosophie*, 4(11), pp.20–40. Available at: <https://www.jstor.org/stable/23932367> [visited on 15 January 2024].

Chalmers, D.J., 2004. Epistemic two-dimensional semantics. *Philosophical Studies*, 118(1), pp.153–226. https://doi.org/10.1023/B:PHIL.0000019546.17135.e0.

Clark, B., 2017. *Wittgenstein, Mathematics and World*. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-63991-8.

Colyvan, M., 1998. Can the eleatic principle be justified? *Canadian Journal of Philosophy*, 28(3), pp.313–335. https://doi.org/10.1080/00455091.1998.10715975.

Colyvan, M., 2011. Fictionalism in the Philosophy of Mathematics. In: E.J. Craig, ed. *Routledge Encyclopedia of Philosophy*. New York.

Craver, C.F., 2007. *Explaining the Brain*. Oxford University Press, USA, p.272.

Craver, C.F., 2014. The Ontic Account of Scientific Explanation. In: M.I. Kaiser, O.R. Scholz, D. Plenge and A. Hüttemann, eds. *Explanation in the Special Sciences: The Case of Biology and History*. Vol. 367, *Synthese Library*. Dordrecht: Springer Netherlands, pp.27–52. https://doi.org/10.1007/978-94-007-7563-3_2.

Craver, C.F., Glennan, S. and Povich, M., 2021. Constitutive relevance & mutual manipulability revisited. *Synthese*, 199(3), pp.8807–8828. https://doi.org/10.1007/s11229-021-03183-8.

Craver, C.F. and Povich, M., 2017. The directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A*, 63, pp.31–38. https://doi.org/10.1016/j.shpsa.2017.04.005.

Creath, R., 1980. Benacerraf and mathematical truth. *Philosophical Studies*, 37(4), pp.335–340. https://doi.org/10.1007/BF00354902.

Davies, E.B., 2005. A defence of mathematical pluralism. *Philosophia Mathematica*, 13(3), pp.252–276. https://doi.org/10.1093/philmat/nki017.

Donaldson, T., 2021. Analyticity. In: M.J. Raven, ed. *The Routledge Handbook of Metaphysical Grounding*, *Routledge handbooks in philosophy*. New York; London: Routledge, pp.288–299.

Dreier, J., 2004. Meta-ethics and the problem of creeping minimalism. *Philosophical Perspectives*, 18(1), pp.23–44. https://doi.org/10.1111/j.1520-8583.2004.00019.x.

Einheuser, I., 2011. Toward a conceptualist solution of the grounding problem. *Noûs*, 45(2), pp.300–314. https://doi.org/10.1111/j.1468-0068.2010.00765.x.

Floyd, J. and Putnam, H., 2000. A note on Wittgenstein's "Notorious Paragraph" about the Godel Theorem. *The Journal of Philosophy*, 97(11), p.624. https://doi.org/10.2307/2678455.

Goodman, N., 1947. The problem of counterfactual conditionals. *The Journal of Philosophy*, 44(5), p.113. https://doi.org/10.2307/2019988.

Hale, B. and Wright, C., 2001. *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*. Oxford : Oxford ; New York: Clarendon Press ; Oxford University Press. https://doi.org/10.1093/0198236395.001.0001.

Kant, I., 1781/1787. *Critique of pure reason*. Ed. and trans. by P. Guyer and A.W. Wood. 2018 edition. Cambridge; New York: Cambridge University Press.

Kim, J., 2011. *Philosophy of Mind*. 3rd ed. Boulder, CO: Westview Press.

Kocurek, A.W., 2021. Counterpossibles. *Philosophy Compass*, 16(11), e12787. https://doi.org/10.1111/phc3.12787.

Kocurek, A.W., Jerzak, E. and Rudolph, R.E., 2020. Against conventional wisdom. *Philosopher's Imprint*, 20(22). Available at: <http://hdl.handle.net/2027/spo.3521354.0020.022> [visited on 15 January 2024].

Kocurek, A.W. and Jerzak, E.J., 2021. Counterlogicals as counterconventionals. *Journal of Philosophical Logic*, 50(4), pp.673–704. https://doi.org/10.1007/s10992-020-09581-6.

Kripke, S.A., 1980. *Naming and Necessity*. Cambridge, Mass: Harvard Univ. Press.

Kuorikoski, J., 2021. There are no mathematical explanations. *Philosophy of Science*, 88(2), pp.189–212. https://doi.org/10.1086/711479.

Lampert, T., 2018. Wittgenstein and Gödel: An attempt to make 'Wittgenstein's objection' reasonable. *Philosophia Mathematica*, 26(3), pp.324–345. https://doi.org/10.1093/philmat/nkx017.

Lange, M., 2013. What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science*, 64(3), pp.485–511. https://doi.org/10.1093/bjps/axs012.

Lange, M., 2016. *Because Without Cause: Non-Causal Explanations in Science and Mathematics*, *Oxford Studies in Philosophy of Science*. Oxford; New York: Oxford University Press.

Lange, M., 2018. A reply to Craver and Povich on the directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A*, 67, pp.85–88. https://doi.org/10.1016/j.shpsa.2018.01.002.

Lange, M., 2021. What could mathematics be for it to function in distinctively mathematical scientific explanations? *Studies in History and Philosophy of Science Part A*, 87, pp.44–53. https://doi.org/10.1016/j.shpsa.2021.02. 002.

Lange, M., 2022. In defense of really statistical explanations. *Synthese*, 200(5), p.388. https://doi.org/10.1007/s11229-022-03868-8.

Lewis, D., 1973. *Counterfactuals*. 1st ed. Blakwell.

Locke, T.D., 2021. Counterpossibles for modal normativists. *Synthese*, 198(2), pp.1235–1257. https://doi.org/10.1007/s11229-019-02103-1.

Lyon, A., 2012. Mathematical explanations of empirical facts, and mathematical realism. *Australasian Journal of Philosophy*, 90(3), pp.559–578. https://doi.org/10.1080/00048402.2011.596216.

Lyon, A. and Colyvan, M., 2007. The explanatory power of phase spaces. *Philosophia Mathematica*, 16(2), pp.227–243. https://doi.org/10.1093/philmat/nkm025.

Maddy, P., 1990. *Realism in Mathematics*. Oxford; New York: Clarendon Press; Oxford University Prress.

Mancosu, P., 2008. Mathematical explanation: Why it matters. In: P. Mancosu, ed. *The Philosophy of Mathematical Practice*. Oxford University Press, pp.134–150. https://doi.org/10.1093/acprof:oso/9780199296453.003. 0006.

Melia, J., 2000. Weaseling away the indispensability argument. *Mind*, 109(435), pp.455–480. https://doi.org/10.1093/mind/109.435.455.

Moore, A., 1999. More on 'The Philosophical Significance of Gödel's Theorem'. *Grazer Philosophische Studien*, 55(1), pp.103–126. https://doi.org/10.1163/18756735-05501006.

Pincock, C., 2011. *Mathematics and Scientific Representation*. Oxford; New York: Oxford University Press.

Pincock, C., 2015. Abstract explanations in science. *The British Journal for the Philosophy of Science*, 66(4), pp.857–882. https://doi.org/10.1093/bjps/axu016.

Povich, M., 2018. Minimal models and the generalized ontic conception of scientific explanation. *The British Journal for the Philosophy of Science*, 69(1), pp.117–137. https://doi.org/10.1093/bjps/axw019.

Povich, M., 2020. Modality and constitution in distinctively mathematical explanations. *European Journal for Philosophy of Science*, 10(3), pp.1–10. https://doi.org/10.1007/s13194-020-00292-y.

Povich, M., 2021. The narrow ontic counterfactual account of distinctively mathematical explanation. *The British Journal for the Philosophy of Science*, 72(2), pp.511–543. https://doi.org/10.1093/bjps/axz008.

Povich, M., 2024. *Rules to Infinity: The Normative Role of Mathematics in Scientific Explanation*. Oxford University Press.

Price, H., 2011. *Naturalism Without Mirrors*. Oxford, New York: Oxford University Press.

Priest, G., 2013. Mathematical pluralism. *Logic Journal of the IGPL*, 21(1), pp.4–13. https://doi.org/10.1093/jigpal/jzs018.

Priest, G., 2021. A note on mathematical pluralism and logical pluralism. *Synthese*, 198(20), pp.4937–4946. https://doi.org/10.1007/s11229-019-02292-9.

Reutlinger, A., 2016. Is there a monist theory of causal and noncausal explanations? The counterfactual theory of scientific explanation. *Philosophy of Science*, 83(5), pp.733–745. https://doi.org/10.1086/687859.

Roski, S., 2021. In defence of explanatory realism. *Synthese*, 199(5), pp.14121–14141. https://doi.org/10.1007/s11229-021-03413-z.

Saatsi, J., 2011. The enhanced indispensability argument: representational versus explanatory role of mathematics in science. *The British Journal for the Philosophy of Science*, 62(1), pp.143–154. https://doi.org/10.1093/bjps/axq029.

Saatsi, J., 2012. Mathematics and program explanations. *Australasian Journal of Philosophy*, 90(3), pp.579–584. https://doi.org/10.1080/00048402.2012.665374.

Saatsi, J., 2016. On the 'indispensable explanatory role' of mathematics. *Mind*, 125(500), pp.1045–1070. https://doi.org/10.1093/mind/fzv175.

Salmon, W.C., 1984. Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, (2), pp.293–305. https://doi.org/10.1086/psaprocbienmeetp.1984.2.192510.

Salmon, W.C., 1989. Four decades of scientific explanation. In: P. Kitcher and W.C. Salmon, eds. *Scientific Explanation*, *Minnesota studies in the philosophy of science*, 13. Minneapolis: University of Minnesota Press, pp.3–219. Available at: <http://conservancy.umn.edu/handle/11299/185700> [visited on 4 October 2023].

Schiffer, S., 2003. *The Things We Mean*. Oxford and New York: Clarendon Press.

Sidelle, A., 1989. *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Ithaca, NY: Cornell University Press. https://doi.org/10.7591/9781501746260.

Sidelle, A., 2009. Conventionalism and the contingency of conventions. *Noûs*, 43(2), pp.224–241. https://doi.org/10.1111/j.1468-0068.2009.00704.x.

Simpson, M., 2020. Creeping minimalism and subject matter. *Canadian Journal of Philosophy*, 50(6), pp.750–766. https://doi.org/10.1017/can.2020.20.

Stalnaker, R., 1978. Assertion. In: P. Cole, ed. *Syntax and Semantics: Pragmatics*. Vol. 9. New York: Academic Press, pp.315–332.

Stalnaker, R., 2001. On considering a possible world as actual: Robert Stalnaker. *Aristotelian Society Supplementary Volume*, 75(1), pp.141–156. https://doi.org/10.1111/1467-8349.00083.

Steiner, M., 1978. Mathematics, explanation, and scientific knowledge. *Noûs*, 12(1), pp.17–28. https://doi.org/10.2307/2214652.

Steiner, M., 1996. Wittgenstein: 'Mathematics, regularities, rules. In: A. Morton and S.P. Stich, eds. *Benacerraf and His Critics*, *Philosophers and their critics*, 8. Cambridge, MA: Blackwell Publishers, pp.190–212.

Steiner, M., 1998. *The Applicability of Mathematics as a Philosophical Problem*. Cambridge, MA: Harvard Univ. Press.

Steiner, M., 2009. Empirical regularities in Wittgenstein's philosophy of mathematics. *Philosophia Mathematica*, 17(1), pp.1–34. https://doi.org/10.1093/philmat/nkn016.

Thomasson, A.L., 2007a. Modal normativism and the methods of metaphysics. *Philosophical Topics*, 35(1/2), pp.135–160. https://doi.org/10.5840/philtopics2007351/27.

Thomasson, A.L., 2007b. *Ordinary Objects*. Oxford; New York: Oxford University Press.

Thomasson, A.L., 2014. *Ontology Made Easy*. Oxford: Oxford University Press.

Thomasson, A.L., 2020. *Norms and Necessity*. Oxford: Oxford University Press.

Thomasson, A.L., 2021. How can we come to know metaphysical modal truths? *Synthese*, 198(8), pp.2077–2106. https://doi.org/10.1007/s11229-018-1841-5.

Topey, B., 2019. Linguistic convention and worldly fact: Prospects for a naturalist theory of the a priori. *Philosophical Studies*, 176(7), pp.1725–1752. https://doi.org/10.1007/s11098-018-1088-5.

Waismann, F., 1986. The Nature of Mathematics: Wittgenstein's Standpoint. In: S. Shanker, ed. *Ludwig Wittgenstein: Critical Assessments, Vol. III*. London: Croom Helm, pp.60–67.

Warren, J., 2020. *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism*. Oxford; New York: Oxford University Press.

Warren, J., 2022a. Inferentialism, conventionalism, and a posteriori necessity. *The Journal of Philosophy*, 119(10), pp.517–541. https://doi.org/10.5840/jphil20221191034.

Warren, J., 2022b. *The A Priori Without Magic*. 1st ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009030472.

Wigner, E.P., 1960. The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Communications on Pure and Applied Mathematics*, 13(1), pp.1–14. https://doi.org/10.1002/cpa.3160130102.

Wittgenstein, L., 1976. *Wittgenstein's Lectures on the Foundations of Mathematics Cambridge, 1939*. Ed. by C. Diamond. Ithaca, NY: Cornell University Press.

Wittgenstein, L., 1978. *Remarks on the Foundations of Mathematics*. Ed. by G. von Wright, R. Rhees and G. Anscombe (G. Anscombe, Trans.). Oxford: Blackwell.

Wittgenstein, L., 2013. *Tractatus Logico-Philosophicus* (D. Pears and B. McGuinness, Trans.), *Routledge great minds*. London: Routledge.

Woodward, J., 2003. *Making Things Happen: A Theory of Causal Explanation*, *Oxford Studies in Philosophy of Science*. Oxford, New York: Oxford University Press.

Wright, C., 1985. In Defence of the Conventional Wisdom. In: I. Hacking, ed. *Exercises in Analysis*. Cambridge: Cambridge University Press, pp.171–197.

Can fiction and veritism go hand in hand?

Antoine Brandelet

Université de Mons, Université de Namur

Abstract

The epistemology of models has to face a conundrum: models are often described as highly idealised, and yet they are considered to be vehicles for scientific explanations. Truth-oriented—veritist—conceptions of explanation seem thereby undermined by this contradiction. In this article, I will show how this apparent paradox can be avoided by appealing to the notion of fiction. If fictionalism is often thought to lead to various flavours of instrumentalism, thereby weakening the veritist hopes, the fiction view of models offers a framework much richer than it seems at first sight. To do so, I will call upon the concepts of modality, counterfactual structure and credible worlds. In the end, veritism of explanation and fiction can indeed go hand in hand, but the scope of explanations we can hope to draw from models must be more precisely delineated.

Keywords

fiction, explanations, idealisations, models, veritism.

1. Setting the Problem

Models are ubiquitous in science. It is widely acknowledged that they are of central epistemological importance, and un-

derstanding their nature and function has become one of the most discussed topics in philosophy of science. There is no agreement on the general framework in which we need to address the modelling problem, and finding one is probably not even desirable. The reason for the variety of available approaches is obvious: diversity of uses and diversity of objects. Models are supposed to serve many functions: represent physical systems, provide explanations or help in theory construction, to cite only a few. Also, many kinds of objects, abstract and concrete, are considered as models: sets of mathematical equations, algorithms, graphs, drawings, or scale-models, for example. If we want to understand the explanatory power of models, which is our main topic here, we must tackle this problem of diversity.

With the acknowledgement of the wide use of models also comes a much more difficult conundrum. One of the main features of modelling practices is the idealisation[1]: models are always simpler than the systems or situations they depict or explain. "Models are generally caricatures of the natural world" (Chakravartty, 2001). The examples abound: the billiard-ball model of gases considers molecule collisions as perfectly elastic, predictions of planetary motion are obtained via the hypothesis that planets are perfectly spherical and with constant mass density, the Lotka-Volterra predator-prey model describes two populations with no outside influences, models in economics are concerned with perfectly rational agents making their decisions from all the available information.

This seems at first problematic, not to say paradoxical: models are vehicles for explanations, and yet they contain idealisations, distortions, purely fictional objects or even impossibilities. Following

---

[1] Here, I call "idealisation" the general process of simplification, the exact distinction between abstraction, idealisation and/or approximation is not relevant for my point.

Elgin (2017), facing the use of these "felicitous falsehoods", one could argue for the relaxation of truth as the main epistemic goal of science, hence her critique of *veritism*.

In the context of scientific explanation, veritism is the view that considers truth as a necessary condition for explanation (Pincock, 2021). Broadly conceived, I think a position needs to meet two conditions to be considered as veritist: firstly, there must be a form of correspondence or similarity between the model and the physical system it represents and, secondly, the framework must provide a way to distinguish between "good" and "bad" explanations. I will elaborate on that in section 5.

We are faced with a dilemma: either the idealised models refute the veritist position, or the idealised models are not epistemically legitimate and the explanations provided by such models must be rejected. The second horn is clearly not in line with scientific practice: highly idealised models are often used to explain the phenomena observed, and the model-based explanations are considered a fully legitimate part of scientific knowledge. Veritism then seems refuted.

A similar debate is taking place in the context of representation. Besides explanation, models are also used to represent target physical systems. Explanation and representation are two distinct but closely related problems. If explanation faces the conundrum of veritism, representation is concerned with the more general problem associated to the realism/antirealism debate. Representations cannot be strictly qualified as true or false, but there seems to be something like a "resemblance" or "correspondence" between the model and the target that is at work.

In both cases, the worries are concerned with conceiving a model-world relation as a basis of knowledge, even if the "veritist need not endorse any specific account of how models represent" (Pincock,

2021). I think this is true, but here, I want to show how one could use resources provided by responses to the representation problem to clarify the veritism conundrum.

Recently, Roman Frigg (see for example (Frigg, 2009), or, with James Nguyen (Frigg and Nguyen, 2016; 2020)) has developed the so-called "fiction-view of models", in which models are broadly conceived as *Epistemic Representations*. This solves the aforementioned problem of diversity, as representative models are any object that is used as a vehicle to support surrogative reasoning. According to Frigg, models are fictions in the sense that they are an invitation to engage in a prop-oriented game of make-believe, understood in its Waltonian sense (Walton, 1990). Leaving technical details aside (see e.g. (Toon, 2012) for a discussion of Walton's view), the main idea is that when using a model, we accept assumptions knowing that they are false but which acceptance function as principles of generation. We enter the fictional world they describe by pretending they are true. In other words, every proposition of a model could be preceded by an "as-if" clause referring to the rules of the game of make-believe.

The fiction view is intended to solve the problem of representation (i.e. how can models provide inferences from the model to the represented target system?). Here, I will focus on the explanatory power of models, but I think these fictional resources can be of great help to the explanation problem. This is what I will defend in section 2.

Expressed in this way, the fiction view seems to support purely instrumentalist and non-veritist conceptions of explanation. Models can incorporate any kind of false assumptions as soon as they empower explanatory or representational power, undermining all the truth-oriented hopes of the veritist or the realist. I think that even if make-believe puts fiction at the center, there is still a way to defend a slightly modified version of veritism. In subsequent sections,

I will examine in more detail the consequences of this fiction-oriented view to find a way for veritism in the make-believe framework. More specifically, engaging in a game of make-believe is better understood as the process of building a counterfactual story, which highlights the importance of counterfactual reasonings in model-based explanation (Section 3). Models depict possible worlds from which we hope to draw inferences, and justifying these inferences is at the core of our present problem. I claim it is the counterfactual structure exemplified by the model that supports these inferential steps, and as models contain something inherently modal, the kind of explanations models provide also have something to do with modality. This is what I will explore in Section 4.

Section 5 sets the counterfactual structure as the focal point of the model-world relation the veritist is looking for. I show how the fiction view can help demarcate between valid and invalid explanations once the modal aspect of the model is appropriately understood. I then examine how it influences the kind of explanations one could expect to derive from models.

In Section 6, I discuss how the fictional approach can help to clarify the opposition between the ontic and epistemic conceptions of explanation. The focus on idealisations and representations in model building is often taken as an argument in favour of epistemic approaches, that is conceptions that consider explanations to be products of some scientific activity involving various techniques, such as the fictional processes described in this paper. I argue on the contrary that critiques of the ontic conception focusing on idealisations and representations often miss the point: the ontic conception suitably understood does not deny the importance of fictional processes, but puts the focus on the referents of explanatory texts and representations. The

fiction view is able to provide a framework in which questions about explanation and questions about the ontic or epistemic expectations of explanations are clearly distinguished.

2. Fiction and Models

In this section, I will (very) briefly summarise the main aspects of the fiction view of models.

Epistemology was of course not the first to use the concept of fiction. This term refers primarily to "works of fiction", understood in the purely artistic sense of the term, and it has been the subject of much discussion in aesthetics, for example. Recent interest for fiction in epistemology and philosophy of science often refers to Walton's seminal work, *Mimesis as Make-Believe*, which deals mainly with artistic representation, but offers a framework suitable to address epistemological issues.

The idea that science deals with theories or models that simplify reality is not new either, and one could find the premises of such an analysis for instance in Vaihinger's extensive use of the "as-if" statement (Vaihinger, 1911). That is an important shift in understanding the explanatory function of models. Models don't just simplify matters by isolating variables, idealising processes or abstracting properties. They are an invitation to think "as-if" things were so and so, just as a work of fiction is a prescription to imagine situations, people and places. The simple pendulum is a fiction in the sense that reading its theoretical description, we imagine a point mass oscillating at the end of a non-extendable, massless string, even if we know point masses and massless strings don't exist in reality.

Specifying a model by stipulating principles of generation is of course not the whole story. One could imagine any set of (at least coherent) assumptions and claim that it is a model by a mere *modelling fiat*. But it is not. For a set of hypotheses to constitute a model, it needs to be applicable to the target system: the terms involved in the assumptions need to be interpretable in terms of the target. Most of the time, scientific models come already partially interpreted by a theory. The simple pendulum is not a model of real world pendula because scientists just say so, but because we can assign to each term (e.g. m, l, g) a target-interpretation (respectively mass of the bob, length of the cord, gravitational field magnitude), and the relevance of these variables is inherited from Newtonian mechanics, the theory in which the model takes place.

The model is then studied, investigated and manipulated. New fictional truths (i.e. propositions true in the model) are derived from the principles of generation. In a famous quote, Hacking writes that "a model in physics is something you hold in your head rather than your hands" (Hacking, 1983). I think we can understand this quote literally. I take the derivation of new fictional truths as analogous to the manipulation, for instance, of a scale-model or a map: generally conceived as epistemic representations, models are manipulable because we can discover new propositions from postulated ones. It is clear for the example of the pendulum, but finding your way by making inferences from a map works in the exact same way: interpreting lines and color patches as roads and forests enable the map user to determine his position.

According to the fiction view of scientific representation, that is where the representative power of models comes from. It also explains why scientists often talk about abstract models as-if they were real objects, as we talk about Sherlock Holmes as-if he was a real person.

Starting from Walton's make-believe theory, Frigg and Nguyen (2020) have developed the DEKI account of scientific representation, but for our present purpose, only the basic concepts of the make-believe view are necessary.

As a matter of fact, fiction is often thought to clash with truth. In libraries, there is a strict distinction between fiction and non-fiction sections. In everyday language, fiction is associated with imagination, fantasy or lies. If models incorporate such imaginary or false assumptions, it seems impossible to reconcile this view with veritism.

In her book *True Enough* (2017), Catherine Elgin establishes this incompatibility on the very first page:

> Modern science is one of humanity's greatest cognitive achievements. To think that this achievement is a fluke would be mad. So epistemology has the task of accounting for science's success. A truth-centered, or veritistic, epistemology must treat models, idealizations, and thought experiments as mere heuristics, or forecast their disappearance with the advancement of scientific understanding. Neither approach is plausible. We should not cavalierly assume that the inaccuracy of models and idealizations constitutes an inadequacy; quite the opposite. I suggest that their divergence from truth or representational accuracy fosters their epistemic functioning. When effective, models and idealizations are, I contend, felicitous falsehoods. They are more than heuristics. They are ineliminable and epistemically valuable components of the understanding science supplies. (Elgin, 2017, p.1)

As the make-believe view bets on the central importance of fictional aspects of modelling, truth-based explanations cannot be derived from models at all, it seems, and the conclusion is the same as Elgin's: *Felicitous falsehoods* cannot be removed from the success of science, then *veritism* must be abandoned.

I think there is nonetheless a way to defend a modified form of veritism in the context of the fictional view. Such a defense must proceed in two steps. The first concerns epistemic virtues. Elgin acknowledges the virtue of idealisations, but dismisses that of truth. On the other side, the veritist claims that the main epistemic virtue is truth, but still, that doesn't mean idealisations cannot also have some kind of value and that the two cannot be articulated in a common framework. The second step, which will be our main concern in the remainder of this article, is to establish what limitations, if any, the fiction view imposes on the scope of model-based explanations.

There is a large literature about epistemic virtues, and various authors have defended that veritism and idealisation may not be as incompatible as we may think. For instance, Nawar argues that "in grasping and idealizing claim *as an idealizing claim*, if seems that one in facts grasps a truth" (Nawar, 2019, his emphasis). Sullivan and Khalifa (2019) admit that idealisation has virtues, but that they are non-epistemic. Idealisations are used for convenience, simplicity or tractability, and in this sense they are felicitous falsehoods, but only their true components can provide understanding. In the same vein, Lawler argues that "falsehoods can play an epistemic enabling role in the process of obtaining understanding but are not elements of the explanations or analyses that constitute the content of under-standing" (Lawler, 2019). Making the process/content of explanation distinction, called the *extraction view*, is interesting for our purpose. Concerning the former, the fiction view enables all the usual virtues granted to idealisations by taking these fictionalising procedures as the central feature of models. Concerning the latter, to be fully adequate in the fiction context, we must clarify in terms of make-believe what exactly the content of the explanations provided by the models is.

Admitting there is a place for the virtue of idealisations even if truth still constitutes the main goal of inquiry in providing explanations, we must now turn to the question of explanations themselves. How can fictional models generate virtuous explanations? How can the false explain the real? This is what we will see in subsequent sections.

3. Counterfactuals at work

In this section, I will examine the lessons we can draw from the fiction view in the way models generate explanations. I will also see how my account may provide insights to understand the relation between models and laws via the use of counterfactual inferences.

Remember the fiction view proceeds in two stages: firstly, rules of the game that generate the fictional world are postulated; then, secondly, the model is explored and fictional truths are derived from these principles of generation.

Modulo the applicability of the model, any kind of assumption is *a priori* acceptable, whether it be idealisation, abstraction or distorsion. These are not the whole story. Some of the fictional processes the scientist may use to build the model are not reducible to these simplifying assumptions. As a matter of fact, models sometimes feature impossibilities, assumptions that are incompatible with the theory in which the model takes place.[2] Take the case of the simple pendulum as an example: point masses not only do not exist in reality, but are also impossible according to the Newtonian picture of the

---

[2] I here use the term theory in a very loose sense: a system of concepts and general principles. The question of the theory-model relationship is a debate in its own right and is beyond the scope of this article, but I will briefly sketch a possible answer that naturally arises in the fiction framework in the remainder of the article.

world. I take this observation to be a good reason to turn ourselves to a modal conception of models, where notions of possibility are directly implemented in the framework.

Models, like fiction, seem to describe possible worlds, that is worlds that resemble ours in many aspects, but where Sherlock Holmes is a detective, rabbits talk or point masses oscillate without friction. If we acknowledge models may contain impossibilities, the fiction view itself faces a conundrum, as coherence and consistency are often presented as necessary conditions for a work of fiction to be acceptable.[3] In the context of epistemology, this seems to suggest that some models contain clashing propositions, thus creating a contradictory story and undermining the potential veridic ground of explanations. What is an impossible world, according to the fiction view? The distinction is important, as any account oriented towards truth must be able to offer acceptability criteria. At this point, we also face the modal version of the initial conundrum: how can models be impossible descriptions of real (therefore possible) phenomena?

To resolve the apparent paradox, we must clarify the use of (im)possibilities in models to understand how they fit together. I think the resources of the fiction view are of great interest, here, as it draws our attention to the important distinction between the inside of the model (fictional—intradiegetic[4]—propositions) and the outside.

When qualifying assumptions as impossible, it is always with an implicit frame of reference in mind. Something is possible or impossible only according to a set of hypotheses or axioms. In this regard, all the fictional propositions are, by definition, diegetically possible. When we say that a model features impossibilities, it is

---

[3] This is also related to debates around the *willing suspension of disbelief*. Interestingly, Kendall Walton is one of the critics of this approach (see e.g., Walton, 1978, p. 7).

[4] Literally: inside (intra) the narrative (diegetic).

with regard to the exterior, to the laws, theories or principles we believe to be true in reality. Models are often thought to be interpreted structures that link the theories to the empirical world, the fiction view generalises this idea to any kind of inference vehicles (theoretical or material) and principles of generation (theories, laws or imaginary entities).

Newtonian mechanics is testable only if we build a model that generates, when applicable to a target system, empirical propositions. The theory itself acquires representational or explanatory power via the models that depict concrete situations. Remember the Ian Hacking quote. What does it mean to manipulate a theoretical model? Taking the simple pendulum example, that means plugging values for the free parameters and seeing what comes out. The model does not contain number values, but a network of relations between variables with potential inputs. Manipulating a model is then playing with counterfactual propositions: when applicable to the target, they generate propositions of the form "had the parameters had such and such values, the system would behave as such and such". That is how models are empirically testable. As the "as-if" clause is characteristic of the fictional process, these models generate "what-if" propositions about their targets. Models are sets of propositions arranged in a counterfactual structure and potentially applicable to a target physical system and some, but not all, their principles of generation are derived from a more general theory: the simple pendulum is a Newtonian model because it embodies what are considered to be the Newtonian laws.

So far so good, but does it still make sense to talk about being the model of a theory in this context? As we have seen, models often contain impossibilities as premisses, which makes them incompatible

with the theory. If models are believed to serve as intermediaries, they must respect the properties they pretend to exemplify. Models may be caricatures of the world, but not of the theories they are models of.

The analogy with fiction sheds light on the complex relation of models not only with physical systems, but also with theoretical principles. A model is neither a strict exemplification of theoretical principles, nor it is a faithful representation of target systems. Models are strange, sometimes abstract, objects, made of heterogeneous parts like an epistemic Frankenstein's creature.

Yet, the relation of fiction to truth is also a more complicated story, and we sometimes use them to learn about the real world. It is widely acknowledged that fiction is not reducible to falsity. One of the most used examples is Tolstoy's *War and Peace*, which contains lots of accurate details about the Napoleonian wars. This example shows there is a use of outside truths for diegetic purposes, and that a reader could learn about the real world by reading the novel. However, this could hardly be described as genuine knowledge as, in the absence of prior background knowledge, the reader may be as justified to believe in the love story of Pierre and Natasha. In the case of scientific models, we have seen that, because of postulated impossibilities, the same problem arises.

There is, I think, another example more suitable to our epistemic concerns that will illustrate the way the fiction view may solve the problem. Let's consider the fables of La Fontaine. At first glance, they describe anthropomorphic animals in imaginary situations. The characters themselves are less human-like in their attitudes than they are archetypes of certain behaviours. Yet, the fictional world depicted by the fable has the function of providing information on actual human behaviour. This is the role of the final moral. Here, I think, the analogy makes sense: idealised entities are postulated, some of their properties

are intended to be interpreted literally while others are not, and the final epistemic goal is to state something true of the exterior of the fiction.

The analogy also seems to suggest an important role assigned to the theories. What makes the moral of a fable a good indication of human behaviour? I claim it is a kind of legal compliance with respect to laws that are supposed to govern human behaviour. In this sense, the fable functions as a model: the fictional entities are embedded in a web of relations and these relations are described by certain laws. It is no threat to the compatibility with the background theory because it is made to describe humans and not animals[5]. The model is a model of that theory because the objects it depicts, even if purely fictional, enter the web of relations described by the theory. In this sense, models are intermediary objects between a theory and target systems.

This is, I think, the main contribution of the fiction view: it leads to a sort of structural *divide et impera* strategy. Models are considered as modal structures, mechanisms used to exemplify a web of counterfactual relations. The modal structure is the skeleton of the model, fictional hypotheses are the flesh that makes the model more tractable or easy to interpret. More importantly: the modal structure exemplified is partially independent from the fictional assumptions.

Manipulating the simple pendulum, we find a relation for the period that is independent of time, regardless of knowingly false hypotheses. That is this counterfactual structure that makes the work. When a model is empirically adequate, we have to ask ourselves: what makes the predictive job? When explaining with models, the question is: what is doing the explanatory job? Explaining in terms of the validity of the principles of generation (rules of the game) is a no-go (at least for the veritist): they are knowingly false. Then if something

---

[5] I leave aside the symbolic aspect of the use of certain animals in that particular case.

is doing the explanation, it is the modal structure itself, inherited from the laws of the theory the model is a model of.[6] Asking for an explanation of the independence on mass of the period of the pendulum, one may present the simple pendulum model, show how the equation $T = 2\pi\sqrt{l/g}$ is derived and explain why it is applicable to the target pendulum. In this context, idealisations, abstractions and all the fictional processes are epistemically valuable not because they constitute the explanation, but because they help make the model tractable by expliciting its counterfactual structure. As Bokulich (2016, p.1) puts it: "Fictional models can succeed in offering genuine explanations by correctly capturing relevant patterns of counterfactual dependence and licensing correct inferences."

Exploring the real via falsehoods still seems a dangerous strategy if we are not able to distinguish between valid and invalid explanations, as there is still some kind of *pessimistic meta-induction* (PMI) threat here. Bokulich takes the example of a non-explanatory fiction: the epicycles model of the Solar system. From a purely fictional-counterfactual point of view, this and the Newtonian-heliocentric models are on a par, but the former fails in capturing the right counterfactual pattern. This is of course circular reasoning if we can't provide a justification for what "right" here means. This is what Bokulich calls the "justificatory step", but she claims that a general account is impossible, thereby undermining the veritist hopes:

> However, what does it mean to say that a fictional representation is adequate? It has to be more than mere empirical adequacy. Unfortunately, here is where I think abstract philosophical generalizations purporting to hold across all model

---

[6] My account does not rely on any particular conception, or metaphysics, of laws, I use the term in a minimal sense: I take counterfactuals as the focal point of explanation and laws are known for supporting counterfactuals. Models exemplifying counterfactual dependencies are in this sense inherited from laws.

explanations give out, and one needs to turn to the nitty-gritty
details of the science in question. What is to count as an ad-
equate fictional representation is something that has to be
negotiated by the relevant scientific community and will de-
pend on the details of the particular science, the nature of the
target system, and the purposes for which the scientists are
deploying the model (Bokulich, 2012, p.734).

I think that objection is no fundamental threat to my account
for two reasons. Firstly, finding the common ground veridically in-
terpretable for explanations and providing a general acceptability
criterion are two different things. In this sense, my account is minimal,
as it proposes a necessary condition. I agree that the sufficiency ar-
gument may be context dependent. Secondly, the resources provided
by the fiction view seem compatible with the general structural ar-
guments put forward by many scientific realists. The response to the
PMI-like worries raised by Bokulich may follow the same path.

In the same vein as Bokulich, Potochnik claims that

depicting causal patterns regularly motivates departures from
accuracy of any given phenomenon; this is why idealizations
are used to represent as-if. Put in these terms, the present
idea is that idealizations positively contribute to generating
understanding by revealing causal patterns and thereby en-
abling insights about these patterns that would otherwise be
inaccessible to us (Potochnik, 2017, p.95).

Again, the highlighting role of idealisations is considered as one
of the main aspects of models, but the exact role and nature of those
causal patterns remain unclear. Potochnik writes her view does not
rely on any particular metaphysics of causation, but she acknowledges
causal patterns must be real: "How, then, can we tell if understand-
ing is actual and not merely apparent? For this, the causal pattern

apparently grasped must be real. Briefly, for a causal pattern to be real, it must be embodied (to some degree or other) in some range of phenomena" (Potochnik, 2017, p.115). But still, she dismisses truth as the main epistemic goal of science:

> The clearest illustrations of this are idealizations themselves, which are quite far from the truth but, in the right circumstances, are epistemically acceptable nonetheless. So, in my view, science simply is not after the truth. There are some important ways in which truth still may be involved in the scientific enterprise, but in each case, it is only a means to other ends (Potochnik, 2017, p.117).

So, when truth is indeed put forward by scientists, it is always with other, more important, goals in mind: understanding of phenomena, which is not truth-oriented. But what would it mean to highlight real causal patterns in a non-truth-centered way? To me, it seems clear that qualifying something as real in a model, even highly idealised, involves some kind of correspondence and, in the end, (at least partial) truth. There is a tension at play, here: accuracy is a requirement of epistemic acceptability, but we must refuse to align it with truth, as idealisation helps generate explanations and understanding. The initial puzzle is still unsolved.

The fiction view seems particularly adequate to treat this problem, as it makes clear the distinction between the fictional process by which the world of the model (the description of possibilities potentially not realised) and the counterfactual structure it brings out. The role of the fictional part is to bring counterfactual structure to the front, but the structure itself is immune to fiction, as we have seen. I propose to take this observation as the focal point of our veritist considerations.

One counterargument would be to argue that once fiction enters the picture, it propagates to the entire model. Setting aside approxi-

mations and idealisations, when a purely fictional entity like silogen atoms (Winsberg, 2006) are ineliminable parts of a model, even the counterfactual structure makes a truth-oriented interpretation and explanation impossible. Here again, I think fiction solves the conundrum.

From a make-believe point of view, postulating new knowingly-unreal particles like silogen atoms is generating a model in which these impossible particles exist, have properties and interact. The model itself is predictive and has good empirical adequacy, but to fully understand its use in generating explanation, we also must examine how it is applied to the target system. As fictional assumptions, silogen atoms have no interpretations in terms of the target system, they are not strictly applicable. Engaging in the silogen game of make-believe means we accept the assumptions inside the model-world, but not outside. The model is still capable of being veridically evaluated because what is confirmed is not a matter of entities or even physical processes. Silogen atoms are fictional entities that ground a set of properties, properties that feature in laws (quantum mechanics and solid-state physics, say), laws that are exemplified in the model, thus exhibiting a counterfactual structure. Interpreting silogen atoms as existing would be a misuse of the model, just like asserting that the moral of the fable is only valid for anthropomorphic animals.

The veritist base of the model explanation is then to be found in the way the laws are generating the skeleton of the model, and the model is evaluated by manipulating this structure to make it generate empirically testable propositions. Giving an interpretation of the counterfactual dependencies is the ground of any explanation. We are not forced to identify these patterns as causes and consequences. My proposition remains agnostic and offers room for different interpretations and levels of ontological commitments. This process is fiction-blind but empowered by the fictional freedom.

4. Modality and Explanations

Let us now turn to the kind of explanation such models can generate.

Verreault-Julien (2019) argued that models may provide "how-possibly explanations", which are propositions of the form $\diamond(p \text{ because } q)$. Facing the issue of how highly idealised models featuring impossibilities may provide such an explanation, Verreault-Julien insists on the importance of making a clear distinction between model-propositions (translated in my account, we could say fictional propositions, i.e. propositions true in the fiction) and world-propositions. "What model propositions (e.g., unrealistic assumptions) do is to give reasons to believe in the truth of the possibility claim", he writes. In his view, models may depict impossibilities and still support possibility claims, which are non-fictional world propositions.

With this I agree, only if we consider not propositions of the form $\diamond(p \text{ because } q)$ for given p and q, but a counterfactual function that assigns a value of q to a value of p. The model is used not to support a set of definite statements about the target, but a counterfactual structure supposed to be embedded in the phenomena we are interested in explaining. These functions also often have a higher arity, as they link several physical values.

More generally, and to avoid any metaphysical commitment to the term "because", I think suitably applicable models support claims of the form $\diamond(p \sim q)$, where $\sim$ is a relation between quantities exemplified by the model. The simple pendulum depicts an impossible object but nonetheless supports the modal claim that connects a certain number of quantities applicable to real-world pendula when interpreted as mass, length, etc.

Another support for the need for a clear fictional/world propositions distinction comes from the fictional process itself. As Verreault-

Julien makes clear, it is possible for a model to non-trivially support possibility claims only if we already have $\diamond p$ and $\diamond q$. But of course, these fictional assumptions may be impossible, understood as world propositions, hence the need to distinguish "diegetic" from "extradiegetic" claims, as I suggested in the previous section.

Sugden suggests that the posited similarities between the model and reality may license inductive inferences (see e.g. Sugden, 2000; 2013, p.240). He gives an argument that takes the following form: $p \rightarrow q$ in the model and p and q in the world gives good reason to infer that $p \rightarrow q$ in the world. As in any inductive argument, a similarity between specific observations is posited, and in the case of model-based inference, it is the model-world similarity that supports induction. The second step of Sugden's argument is analogous to what I call applicability, and the focus on the counterfactual structure may support the inferential step by providing ground for the justification of the similarity.

Sugden also takes modality to be an important feature of models: "So what might increase our confidence in such inferences? I want to suggest that we can have more confidence in them, the greater the extent to which we can understand the relevant model as a description of how the world could be" (Sugden, 2000, p.24). There are many ways the world could be, and Sugden proposes *credibility* as a way to sort them, but in his view, it is not clear how credible worlds (i.e. worlds that could be real) would deal with postulated impossibilities. Are impossible worlds credible?

Elsewhere (Brandelet, 2021), I have defended the view that the inductive framework proposed by Mill (Mill, 1843 / 2011) is transformed into a deductive system when causal laws have highlighted the relevant structure in experiments. Regularities are inductively explained and then serve as the ground for deductively making new

predictions. Avoiding the discussion about the nature of causality in Mill's work, I think we can export his view on the notion of laws to consolidate our fictional and counterfactual conception.

I disagree with Sugden when he writes: "To put this another way, the real world is equivalent to an immensely complicated model: it is the limiting case of the process of replacing the simplifying assumptions of the original model with increasingly realistic specifications" (Sugden, 2000, p. 23) because fiction allows for impossibilities that are not only simplifications or idealisations. Natural laws, understood as Mill's inductive generalisations, generate models. Models are then credible only if compatible with these laws, but credibility does not prevent impossibilities.

Laws express sets of relations obtained via inductive reasoning over observed regularities. These counterfactual relations are exemplified in models, and exemplification may involve all sorts of fictional processes such as, but not limited to, idealisations. That is where the felicitous falsehoods draw their epistemic virtues. Postulating impossibilities (i.e. incompatibilities with laws) is the main modelling freedom offered by the fiction view, but as the laws generating the models express a web of counterfactual dependences, only these relations need to be compatible for the model to be acceptable. In Sugden's terms, impossible worlds can be credible, as long as the model is robust through manipulation and the counterfactual structure remains applicable to the modelised phenomena.

There seems to be some kind of "truth-eligibility" in fictional statements. Manipulating a fictional model does not only mean deriving new fictional statements from old ones, it also means being able to give an interpretation of those propositions in terms of the target. Using the model adequately is also a matter of knowing what is not to be supposed to be true, or even possible at all. If a simple pendu-

lum user claims that the empirical adequacy of the model supports the existence of point masses, he is obviously not using the model properly, the non-existence of point masses is no argument against the model itself. We may be wrong about some aspects or properties of the depicted entities, even about their existence, but the conservation of the counterfactual structure through manipulation of the model is the focal point of our understanding and explanations. That is how we shed light on the fixed point of counterfactual dependencies the model exemplifies.

Model propositions must not be taken at face value. A model is more like a dynamical entity, a counterfactual engine that generates sets of propositions about a physical system and provides a justification for inferences to the world. They are "descriptions of how the world could be" (Sugden, 2013, p.241) equipped with inference rules (inherited from the laws) that guide the fictional reasonings.

5. Veritism reloaded?

As in many subfields of epistemology and philosophy of science, realist and truth-oriented positions have faced strong arguments from all sides. If, as I claim, veritism of explanation can be retained in the fiction view of models, what limitations does it impose on veritism?

I think a position needs to meet two conditions to be considered as veritist:

1. There needs to be some kind of correspondence or similarity between the world and the model at play, and this relation must be, at least partially, the ground of the explanation. In

particular, pure empirical adequacy, acceptance by the scientific community or compliance to scientific norms is insufficient if not based on the correspondence relation.

2. The correspondence must be equipped with a demarcation mechanism: we must be able to link the validity of the model-based explanation with the other epistemic virtues, such as empirical adequacy, and the difference between acceptable and non-acceptable explanation must be constructible in the framework.

Firstly, as I have explained in the previous sections, I take the counterfactual structure exemplified by the model to be the ground of the explanation. Clearly, the model must reproduce and predict empirical data in order to be veridically evaluated. Robustness for a range of input values, i.e. manipulability of the model, offers the ground for inductive inferences. This is where the correspondence relation comes into play: as in Sugden's example of inductive (and abductive) argument, a similarity gives the argument its skeleton. I claim this similarity to be a similarity of counterfactual structure inherited in the model from the laws, understood in the minimal sense of inductive generalisations supporting counterfactual reasonings.

Secondly, the lesson from the fiction view is that the demarcation between valid and invalid explanations is possible, but that we must refrain to interpret veridically anything that is not part of the counterfactual structure in our explanations. For example, using a model of silogen atoms to explain phenomena, we may talk about these atoms as-if they were real, but the veritist ground of the explanation are the quantum-mechanical processes at play in solid-state physics. Explaining with silogen would be a misuse of the silogen fiction.

6. Fiction and the ontic/epistemic accounts of explanation

In this section, I will show how the fiction-view of models defended in this paper can help clarify issues related to the conceptions of explanation and the OC/EC opposition.

Conceptions of explanation are typically classified into two categories: ontic and epistemic. Commonly conceived, ontic conceptions (OC) consider explanations to be concrete things that exist "out-there", independently of any theorising about them. On the other hand, epistemic conceptions (EC) take explanations to be the product of a scientific activity involving various techniques, such as representation of the phenomena to be explained. According to the EC, there is then no explanation if no scientist is doing the explaining.

In (Bokulich, 2016), the focus on the central role of idealisations and fictional processes in modelling is taken to be the sign of a need to move "beyond the ontic conception". Her line of argument is quite straightforward and easily phrased in fictional terms. Models feature deliberate falsities and idealisations which are part of their explanatory power. Models considered as fictions have representational power, and these representations are epistemic products built and used by scientists to explain phenomena. In this context, it seems clear that EC is a much more natural way of conceiving explanations.

Differently put, if, as it is claimed in OC, explanations are "full-bodied things" (Craver, 2013, p.40), "objective features of the world" (Craver, 2007, p.27) that exist independently of the epistemic goals, arguments and activities performed by scientists, then how could deliberate falsities be explanatory? The recognition of the explanatory power of fictions rules out this possibility, and the fiction-view provides an argument in favour of EC. That, of course, should be no

surprise: it is at the very heart of *make-believe* oriented approaches to take models as epistemic products designed and used by certain agents in order to achieve certain goals according to certain norms of evaluation.

Also, as Wright (2015) notes, the ontic conception, for example in Salmon's phrasing, involves something as an "exhibition". If it is deeply unclear to point out what could be an exhibition "in re", it is quite natural in the fictional perspective, where exhibition can be understood as a form of model-representation: it is exactly the role of idealisations and fictional hypotheses to exhibit features in models to support the explanation. This is the act of representing which is at the core of model-explanations and in which, as I have emphasised earlier, fiction plays a positive role.

Bokulich (2018) takes her analysis a step further where she proposes what she calls the "eikonic" conception. In this perspective, explanations also are "the product of an epistemic activity involving representations of the phenomena to be explained". Leaving the details aside, the eikonic conception is close to the fiction view discussed in this paper. What I think is more important is the distinction she introduces, as it may help clarify the roots of the OC/EC debate.

One of the key aspects of the eikonic proposition is that it constitutes a conception of explanation and not an account of explanation. An account of explanation is a claim about how explanations work, while a conception is a claim about what explanations are. The problem is that there is an ambiguity in the general treatment of the OC/EC debate: are these conceptions in the general sense, or in the restrained use of Bokulich?

This is an important point because examples traditionally associated with the ontic conception may fall under the epistemic umbrella if seen not as explanations, but as ways of explaining. For example,

causal or mechanistic accounts are on the ontic side: it is the electron that hit the screen that explains the presence of a white dot. But one could also argue that causes, mechanisms and unobservable entities are features of idealised models that are not explanations but descriptions of how explanations work, thus constituting an account of explanation. A mere reference to a cause or mechanism is not sufficient to decide whether it is part of an ontic or epistemic conception of explanation. A causal account is compatible with both the ontic and epistemic conceptions depending on what is considered to be the explanation itself.

If, as Bokulich asserts, early literature on explanation fails to draw the account/conception distinction, it becomes necessary to examine if such a distinction helps clarify the ontic/epistemic opposition. Unlike her, I don't think it does, and I claim the concept of fiction I am focusing on in this paper may explain why. That is what I will illustrate in the remainder of this section.

The lack of the aforementioned distinction and the only recent critique of the ontic conception may seem surprising at first glance. The central theses of the ontic conception, namely that explanations are not arguments but things and that explaining doesn't involve representation, have been left uncriticised for a long time, even if they are obviously false. Some proponents of EC, including Bokulich, share this surprise, but on the contrary, we might as well be surprised that such obviously false positions could be attributed to anyone. This type of astonishment could just as easily have been the result of a misunderstanding as of an ill-considered philosophical position.

To shed light on the tension at play in this defence of EC, let us extend the analysis proposed in the eikonic conception. As a conception of explanation, the eikonic position is a claim about what explanations are and it falls under the epistemic umbrella for reasons

discussed just before. Also, the eikonic conception is independent of the particular account of explanation one chooses to defend. Bokulich also claims that her view is compatible with scientific realism, just as I claim with the fiction view. Explanations are then considered as epistemic products involving a representational activity via idealised models. The compatibility with different accounts comes from the fact that the model may produce the explanation via different means, e.g. by using covering laws, causes, causal patterns, mechanisms, etc.

If a realist justification is to be found in this context, it must be via some kind of correspondence. I elaborated on the realist requirements and the correspondence with the causal patterns the model exemplifies (or exhibits!) in the previous sections. The question that arises is then the following: how could a purely epistemic explanation be realistically justified? Objective features of the world cannot be deliberately false as models are, but if we look for a justification in the realist sense, it must be grounded on a correspondence with objective features, may it be entities, mechanisms, causes or structures, depending on the flavour of realism one prefers.

In his response to Bokulich (2016), Craver develops an argument along these lines. Contrary to many critics of OC, Craver starts by asserting that proponents of the ontic conception simply do not dismiss the explanatory use of texts and arguments:

> When defenders of the ontic view write about explanations as if they are, "out there", as they are, independently of what anyone knows or thinks about them, they are expressing realism about the appropriate referents of explanatory texts, not abandoning the idea that scientists use texts to express explanations (Craver, 2019).

If proponents of EC reject OC because of the obvious use of epistemic products to express explanations, it is a misunderstanding

of what makes OC ontic. Now, if we come back to the fictional or the eikonic view, we may as well claim they fall under the ontic conception. Idealisations and fictional models are ways to extract and convey information, but they do not constitute the explanation. What explains is the correct correspondence with real causal patterns. On that topic, Craver is explicit:

> But conveying explanatory information about X and truly representing the explanation for X are not the same thing. Friends of the ontic conception should say that idealized models are useful for conveying true information about the explanation, but that they are not true representations of the explanation (Craver, 2013, p.50).

And, in the next paragraph, he adds:

> Once these are separated, the problem of idealization is clearly not a problem for theories of scientific explanation; rather it is a problem for philosophical theories of reference. The question at the heart of the problem of idealization is this: What is required for a given explanatory text to convey information about the ontic structure of the world? This is an important question, but it is a question about reference, not a question about explanation. We only invite confusion if we fail to keep these questions distinct (Craver, 2013, p.50).

This is exactly what the fictional view is able to provide: accounting for the representational and idealised component of explanation while justifying realistically these explanations. Fiction makes a clear distinction between the world of the model and what it may refer to. Both OC and EC can make room for various epistemic strategies in conveying the relevant information: idealisations, representations, modelling in terms of causes, mechanisms, laws, etc. In the fictional

view I defend, the model is not the explanation, the model is sup-porting the explanation by making a set of hypotheses manipulable. When asking, in the vein of Bokulich, about the fictional view as a conception of explanation, there are two equally valid answers: it is ontic if the model provides a justified representation of real causal patterns, it is epistemic if the fictional models are considered as tools for predicting and manipulating physical systems with no reference to real entities or causes or mechanisms. In this understanding of "conception", the OC/EC debate aligns with the realism/antirealism opposition: what makes the "electron hit the screen" explanation ontic is the reference to the real photon and its properties. Following Craver, as conceptions of explanation conflate with reference problems, if Bokulich claims her eikonic view is compatible with realism, it then falls under the ontic umbrella, contrary to her claim.

On the other hand, fiction naturally handles the division of labour between questions about what models refer to if they are to count as explanatory and questions about how explanations work, i.e. the question of *accounts* of explanation. Make-believe and exhibition of causal pattern is a description of how model-based explanations work, while the justification in terms of realist-oriented correspondence provides an ontological understanding. In the same vein, Bokulich's eikonic view is an *account* of explanation, neither ontic nor epistemic, but compatible with realist as well as antirealist views on reference, just as the fiction view is to be considered as an account of explanation, irrespective of questions of reference and OC/EC classification. As I showed earlier, this account is compatible with veritism and, more broadly, with a realist attitude towards science.

7. Conclusion

Is there a place for veritism of explanation in the fiction view of models? I think there is, but exploring the aspects of models as games of make-believe about credible worlds imposes limits on the kind of veritism one can hope to achieve.

First of all, the fictional view can accomodate the epistemic virtue of idealisations and approximations: they epistemically contribute to explanations and understanding by simplifying matters, and the fictional freedom in the construction of models offers all the strategies to scientists in doing so.

The central conundrum of the fiction/veritist approaches may be clarified by turning our attention to the modal aspects of models. Models are descriptions of credible worlds which we can manipulate to generate, when suitably interpreted, propositions about physical target systems. Manipulating a model means we explore the robustness across a range of input values of the embedded counterfactual structure. Fictional processes, like postulating non-existent entities, can help in this exploration, but the counterfactual structure itself is immune to fiction and remains veridically interpretable.

The relation between laws and models also appears clearly in the fictional context. Laws generate models in the sense that they are the expression of the counterfactual dependencies the model contains. The structure is the skeleton, fictional hypotheses are the flesh that facilitates reasoning and interpretation. When the model is found to be robust, we can say that the counterfactual structure reproduces the one of the physical system. This is the ground of explanations, and it remains veritist in the sense that it is the agreement between the two structures that strengthens the inductive inferences we draw from the model to build explanations. That also explains why laws are

so important for explanations. The observation of regularities needs to be explained, the law is the explanation, and the models make the link between laws and the world by making them manipulable. Models are not the explanation, but we need models in order to express explanations.

Regarding the OC/EC debate, the fiction view puts forward the positive role of representations and idealisations while making room for realist and truth-oriented justification of explanations. The ontic slogan according to which explanations are objective features of the world must be interpreted as a claim about the reference of explanations rather than a claim about the nature of explanations qua arguments or representations. This shows that the opposition at play is less about explanations that it is about the more general problem of reference and realism. The fiction view marks clearly the distinction by making a claim about how explanations work (i.e. it provides an account of explanation) and by providing the conceptual resources to tackle the problem of the reference of idealised claims and models.

Bibliography

Bokulich, A., 2012. Distinguishing explanatory from nonexplanatory fictions. *Philosophy of Science*, 79(5), pp.725–737. https://doi.org/10.1086/667991.

Bokulich, A., 2016. Fiction as a vehicle for truth: Moving beyond the ontic conception. *The Monist*, 99(3), pp.260–279. https://doi.org/10.1093/monist/onw004.

Bokulich, A., 2018. Representing and explaining: The eikonic conception of scientific explanation. *Philosophy of Science*, 85(5), pp.793–805. https://doi.org/10.1086/699693.

Brandelet, A., 2021. Induction et loi naturelle chez Mill. *Le Philosophoire*, 55(1), pp.205–224. https://doi.org/10.3917/phoir.055.0205.

Chakravartty, A., 2001. The semantic or model-theoretic view of theories and scientific realism. *Synthese*, 127(3), pp.325–345. https://doi.org/10.1023/a:1010359521312.

Craver, C.F., 2019. Idealization and the ontic conception: a reply to bokulich. *The Monist*, 102(4), pp.525–530. https://doi.org/10.1093/monist/onz023.

Craver, C.F., 2007. *Explaining the Brain*. Oxford University Press, USA, p.272.

Craver, C.F., 2013. The ontic account of scientific explanation. *Explanation in the special sciences*. Springer Netherlands, pp.27–52. https://doi.org/10.1007/978-94-007-7563-3_2.

Elgin, C.Z., 2017. *True Enough*. MIT Press Ltd. 352pp.

Frigg, R., 2009. Models and fiction. *Synthese*, 172(2), pp.251–268. https://doi.org/10.1007/s11229-009-9505-0.

Frigg, R. and Nguyen, J., 2016. The fiction view of models reloaded. *The Monist*, 99(3), pp.225–242. https://doi.org/10.1093/monist/onw002.

Frigg, R. and Nguyen, J., 2020. *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Springer International Publishing. 260pp.

Hacking, I., 1983. *Representing and Intervening*. Cambridge University Press. https://doi.org/10.1017/cbo9780511814563.

Lawler, I., 2019. Scientific understanding and felicitous legitimate falsehoods. *Synthese*, 198(7), pp.6859–6887. https://doi.org/10.1007/s11229-019-02495-0.

Mill, J.S., 1843 / 2011. *A System of Logic, Ratiocinative and Inductive*. Cambridge University Press. https://doi.org/10.1017/cbo9781139149839.

Nawar, T., 2019. Veritism refuted? Understanding, idealization, and the facts. *Synthese*, 198(5), pp.4295–4313. https://doi.org/10.1007/s11229-019-02342-2.

Pincock, C., 2021. A defense of truth as a necessary condition on scientific explanation. *Erkenntnis*. https://doi.org/10.1007/s10670-020-00371-9.

Potochnik, A., 2017. *Idealization and the Aims of Science*. University of Chicago Press. 288pp.

Sugden, R., 2000. Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, 7(1), pp.1–31. https://doi.org/10.1080/135017800362220.

Sugden, R., 2013. How fictional accounts can explain. *Journal of Economic Methodology*, 20(3), pp.237–243. https://doi.org/10.1080/1350178x.2013.828872.

Sullivan, E. and Khalifa, K., 2019. Idealizations and understanding: Much ado about nothing? *Australasian Journal of Philosophy*, 97(4), pp.673–689. https://doi.org/10.1080/00048402.2018.1564337.

Toon, A., 2012. *Models as Make-Believe*. Palgrave Macmillan. 148pp.

Vaihinger, H., 1911. *The Philosophy of 'as If '*. 2009th ed. Martino Fine Books. 420pp.

Verreault-Julien, P., 2019. How could models possibly provide how-possibly explanations? *Studies in History and Philosophy of Science Part A*, 73, pp.22–33. https://doi.org/10.1016/j.shpsa.2018.06.008.

Walton, K., 1978. Fearing fictions. *The Journal of Philosophy*, 75(1), pp.5–27. https://doi.org/10.2307/2025831.

Walton, K., 1990. *Mimesis as make-believe : on the foundations of the representational arts*. Cambridge, Mass: Harvard University Press.

Winsberg, E., 2006. Models of success versus the success of models: Reliability without truth. *Synthese*, 152(1), pp.1–19. https://doi.org/10.1007/s11229-004-5404-6.

Wright, C., 2015. The ontic conception of scientific explanation. *Studies in History and Philosophy of Science Part A*, 54, pp.20–30. https://doi.org/10.1016/j.shpsa.2015.06.001.

Ontic or epistemic conception of explanation: A misleading distinction?

Michał Oleksowicz
Nicolaus Copernicus University in Toruń

Abstract

In this paper, I discuss the differences between ontic and epistemic conceptions of scientific explanation, mainly in relation to the so-called new mechanical philosophy. I emphasize that the debate on conceptions of scientific explanation owes much to Salmon's ontic/epistemic distinction, although much has changed since his formulations. I focus on the interplay between ontic and epistemic norms and constraints in providing mechanistic explanations. My conceptual analysis serves two aims. Firstly, I formulate some suggestions for recognising that both sets of norms and constraints, ontic and epistemic, are necessary for scientific theorising. Secondly, I emphasize that there are multiple dimensions involved in scientific explanation, rather than clear-cut alternatives between ontic and epistemic aspects. I conclude with a general observation that although contextual aspects of explanations are unavoidable, the epistemic-relativity of our categories, explanations and models can in fact be compatible with their objectivity. Instead of making hastily drawn ontological implications from our theories or models, we should carefully scrutinize them from the ontic-epistemic perspective.

1. Introduction

The word "explain" is used in very different contexts. Explaining some phenomenon involves performing operations on its representations to understand the "how" or "why" of this phenomenon. The explanation is then a matter of representing what depends upon what. In this paper, I want to explicate the difference between the ontic and epistemic conceptions of scientific explanation (OC and EC, respectively), mainly linked to the so-called new mechanical philosophy (NMP). For those who are not familiar with the latter approach, I will briefly mention that the NMP is a novel revision of Old Mechanism that takes on theoretical problems from the last fifty years of post-logical empiricist philosophy of science. It is particularly focused on the issue of causal explanations of natural phenomena and offers an overview of various methodologies employed in different sciences (Andersen, 2014a,b). According to the new mechanists, the idea of mechanisms as complex causal systems in the world, and mechanistic explanations (MExs) as tools for discovering such complex systems, are both crucial for understanding natural phenomena (Machamer, Darden and Craver, 2000). In the beginning, NMP aimed at examining the causal "talk" within different sciences and discussed the normative properties that a good explanation ought to have (Craver, 2014). New Mechanists undertook both these aspects, focusing on the mechanisms

in their ontic and epistemic aspects, i.e., as real causal systems in the world and representations of worldly things (2005; 2006; 2008; 2013; 2013; 2014).

In what follows, I will briefly introduce the origins of the NMP's distinction between the ontic and epistemic conception of explanation. Those who are proponents of NMP literature can skip this section and go directly to the second section, in which I analyze the debate over the ontic and epistemic norms and constraints of explanation. In the third section, I discuss the solutions offered by Illari (2013), Kästner (2018), and Kästner and Haueis (2021) to the long-lasting opposition between OC and EC. In the fourth section I point out further problems linked with this debate, i.e., the ambiguity of the term "mechanism", and I articulate a dual ontic-epistemic approach, showing in what sense it may benefit for the philosophical approach to discuss the conceptions and accounts of scientific explanation. I conclude with general observation that although contextual aspects of explanations are unavoidable, the epistemic-relativity of our categories, explanations and models can be compatible with their objectivity. Instead of hastily drawing out ontological implications from our theories or models, we should carefully scrutinize them from an ontic-epistemic perspective.

2. New wave of ontic and epistemic conceptions

W. Salmon, in his analysis regarding scientific explanation, points out that OC originated with José Alberto Coffa, who was:

> a staunch defender of the ontic conception of scientific explanation, and his theory of explanation reflects this attitude. For Coffa, what explains an event is whatever produced it or

> brought it about. [. . .] The linguistic entities that are often
> called 'explanations' are statements reporting on the actual
> explanation. Explanations, in his view, are fully objective and,
> where explanations of nonhuman facts are concerned, they
> exist whether or not anyone ever discovers or describes them.
> Explanations are not epistemically relativized, nor (outside of
> the realm of human psychology) do they have psychological
> components, nor do they have pragmatic dimensions (Salmon,
> 1989, p.133).

This conception was further developed by W. Salmon, who at the
same time wavered between two ways of thinking about it (Bokulich,
2016, p.262): whether explanations exist in the world or whether they
are something that reports such facts (Salmon, 1989, p.86). Without
entering into the historical complexities of the development of the OC
and EC to the present, it suffices to say that Salmon further contrasted
the OC with EC. In fact, he mainly situated his philosophical focus
on explanation against C. Hempel's account. For Salmon the "two
grand traditions of scientific explanation" (Salmon, 1989, pp.68–69)
are: the EC, characterized by its focus on logic and laws, according to
which the act of explanation is to show that a phenomenon fits into
a nomic nexus (generally identified with Hempel's covering model of
explanation); and the OC, characterizing causality and explanation as
a causal-mechanical explanation, fitting phenomena into natural pat-
terns and regularities (Salmon, 1984, pp.84–134; 1989, pp.320–330;
Wright and van Eck, 2018).

The OC originated with work of J. Coffa, who had no interest
in the discussion on mechanisms, but it was further elaborated by
W. Salmon, who was directly engaged in formulating the causal-
mechanical account of the OC. Although Salmon himself had a con-
ception of mechanisms which at first glance does not comport with
conceptions of the Glennan, Craver, Bechtel, Darden, Illari, Kästner,

etc., it nevertheless seems that the mechanistic revival is deeply indebted to his philosophical approach (Campaner, 2013). For instance, Salmon's theory already pointed out the crucial role of such notions as production and interaction, the distinction between constitutive and etiological aspects of causal explanation and the usefulness of counterfactuals if interpreted experimentally. Although further nuances of Salmon's view on scientific explanation are not the aim of my examination here, it is essential to emphasize that Salmon's discussion of OC and EC have profoundly influenced the content of the new mechanistic debate on the metaphysics of explanation.

Among proponents of OC can be included W. Salmon, C. Craver, L. Darden, S. Glennan, P. Illari, M. Povich, T. Knuuttila. Let us now focus on core aspects of OC. While L. Darden (2008, p.959) argues that "mechanism is sought to explain how a phenomenon is produced, how some task is carried out, or how the mechanism as a whole behaves", S. Glennan (2002, p.S348) argues that "the explanation lies not in the logical relationship between these descriptions [of the parts of mechanisms] but in the causal relationships between the parts of the mechanism that produce the behaviour described".C. Craver (2007, p.22) asserts that "the explanandum is the release of one or more quanta of neurotransmitters in the synaptic cleft. The explanans is the mechanism linking the influx of Ca^{2+} into the axon terminal". In another place, C. Craver suggests that "all higher-level causes are fully explained by constitutive mechanisms" (Craver, 2007, p.548).What seems to be common to the above claims is that scientific explanations, conceived in an OC manner, are mechanisms existing in the world. Thus, these explanations are not constituted by sentences, diagrams, models, but by fully objective worldly facts. The most explicit advocate of OC is C. Craver. He defends it in the following words:

> Conceived ontically, however, *the term explanation refers to an objective portion of the causal structure of the world, to the set of factors that produce, underlie, or are otherwise responsible for a phenomenon. Ontic explanations are not texts; they are full-bodied things.* They are not true or false. They are not more or less abstract. They are not more or less complete. They consist in all and only the relevant features of the mechanism in question. There is no question of ontic explanations being "right" or "wrong," or "good" or "bad." They just are (Craver, 2014, p.40 italics added).

The crux of the problem, clearly expressed in the quote above, consists in the fact that some advocates of OC begin with the distinction between representations and worldly mechanisms that are represented, but then they claim that the term "explanation" refers to both the depiction of the things in the world and to things in the world. However, talking about ions in the world and talking about representations of ions in the world is not the same thing. In fact, "what our understanding proceeds 'through' are the representations and models of those entities and activities and the ratiocinative procedures thereon—not the activities and entities themselves" (Wright, 2015, p.26). In other words, identifying explanations with the causes themselves is not only not self-evident (Wright and van Eck, 2018), but confusing. The source of such confusion seems to stem from the attempt to sanction the dependence of OC on how the world is (Craver, 2014). But such a dependence is merely postulated. In reality, there is no conception of explanation that denies this sort of dependency. Any view of scientific explanation that takes explanations to be directed at or about anything at all will be compatible with such commitment. Explanations are about the world, and thus dependent on how it is. This is hardly the special feature of the OC and does not in any way distinguish ontic from non-ontic conceptions. H. de Regt (2017, p.24)

rightly argues that "Salmon's distinction is misleading: explanations, including Salmon's causal-mechanical ones, are always epistemic and not ontic, in the sense that they are items of knowledge". Any explanation seems thus to be an epistemic item or an argument in the broad sense.

Among the main defenders of EC we can find W. Bechtel, B. Sheredos, C. Wright, A. Bokulich, A. Levy, M. Nathan, D. van Eck, R. Frigg, H. de Regt. For proponents of EC, MExs are not things existing in the world but something that reports facts about things in the world. For instance, the tools of EC are descriptions, texts, diagrams or models that provide understanding on how mechanisms are responsible for certain phenomena. Wright and Bechtel (2007, p.51) rightly argue that "explaining refers to ratiocinative practice governed by certain norms that cognizers engage in to make the world more intelligible; the non-cognizant world does not itself so engage". Bechtel and Abrahamsen (2005, p.425) echo the previous claim, emphasizing that:

> it is crucial to note that *offering an explanation is still an epistemic activity and that the mechanism in nature does not directly perform the explanatory work. Providing explanations, including mechanistic explanations, is essentially a cognitive activity*. This is particularly obvious when one considers incorrect mechanistic explanations—in such a case one has still appealed to a mechanism, but not one operative in nature [. . .] Thus, *since explanation is itself an epistemic activity, what figures in it are not the mechanisms in the world, but representations* of them. [italics added]

The EC stresses the fact that explanations are cognitive activities. For this reason, they highlight the ways in which science tries to grasp the *explanandum*. The object of explanation is never a mechanism

simpliciter, but the *explanandum* is always embedded within a broader explanatory context. The latter can be understood as the arrangement of instruments, scientific concepts and models, skills and activities of scientists engaged in the research programme aimed at explaining certain phenomena.

Considering the impressive development of the literature and studies dedicated to the modeling view of science, mainly are trying to answer the question of how to understand, provide, and evaluate scientific theories, laws, statements, and models (Meheus and Nickles, 2009; Frigg, 2022). One might think that the debate between OC vs EC has been settled. In the last decade, however, the debate has shifted from the question "what is an explanation" to the *querelle* on ontic and epistemic norms and constraints on good MEx (Illari, 2013). Such a shift means that philosophers are focused on the question about the kinds of norms and constraints that guide MEx. There is the consensus that scientific explanation is the epistemic phenomenon under which agents develop hypotheses or models and reason under assumptions in very specific contexts. Thus, the concern is not about what scientific explanations are, but what the criteria of good scientific explanations are. We now enter into the intra-epistemic debate on norms and constraints of MEx.

3. Norms and constraints

From the beginning, it is important to distinguish a "conception of explanation", understood as a view about what explanations are, from an "account of explanation", conceived of as a view about how explanations work (Bokulich, 2016, p.263). For instance, one can reject the OC, but at the same time endorse that many explanations are causal. It

is very useful to distinguish between norms and normative constraints when looking at philosophical accounts of explanation. I will apply the same notions that L. Kästner and P. Haueis used in their recent paper (Kästner and Haueis, 2021). According to them norms can be "understood as general instructions of how to search for mechanism and how to construct good mechanistic explanatory texts" (Kästner and Haueis, 2021, p.1638). Furthermore, "ontic and epistemic norms can be achieved by using specific normative constraints. Different such constraints are the determinates of the determinable epistemic norm of intelligibility or the ontic norms of accuracy and completeness, respectively" (Kästner and Haueis, 2021, p.1638). Briefly put, while norms work as the general instructions for successful mechanistic inquiry, the normative constraints determinate the latter by limiting the search space for mechanisms in different ways.

Kästner and Haueis give the example of an ontic norm in the instruction to describe the causal structure of a mechanism. In the case of an epistemic norm, they point out the need to increase the intelligibility of the *explanandum*. How can one justify the importance of such ontic or epistemic norms? For instance, Craver stresses "the fact that an explanation that contains more relevant detail about the responsible ontic structures are more likely, all things equal, to be able to answer more questions about how the system will behave in a variety of circumstances than is a model that does not aim at getting the ontic structures that underlie the phenomenon right" (Craver, 2014, p.41). In other words, according to Craver, when we follow the ontic norm of parsing the causal structure of the phenomena, we are far more likely to provide a *bona fide* understanding of the *explanandum*. In the case of epistemic norms, Sheredos argues that generality and systematicity are two prototypical epistemic norms, since they make "intelligible any explanation's scope [i.e., explanations in a commu-

nicative, textual and cognitive sense], unifying explanatory practices, and facilitating research and testing by delineating a category of cases to which any explanation is presumed applicable" (Sheredos, 2016, p.933).

Let us now focus more specifically on how ontic and epistemic norms can be determined through the use of ontic or epistemic normative constraints. In the case of ontic constraints, the fundamental ones are: 1) spatial or temporal constraints, 2) the mechanism-to-model-mapping (3M) constraint. Since "the entities and activities in a mechanism are organized spatially, temporally and actively such that they produce the phenomenon" (Craver and Darden, 2013, p.20) it is crucial to regard the spatial and temporal organization of mechanisms. The first one consists in locations, sizes, shapes, and orientations of components; while the second one pertains to the orders, rates, and durations of stages. Both spatial and temporal constraints are then crucial for the identification of parts and activities of mechanisms. Apart from the spatiotemporal constraints, one may opt for determination of ontic norms by the use of the 3M constraint. D. Kaplan and C. Craver specify the 3M requirement in the following words:

> (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism (Kaplan and Craver, 2011, p.611).

This quote suggests that explanation should rely on ontic constraints to help in providing a good causal explanation. That it is not to be fulfilled *simpliciter* is obvious, if one considers that explanatory accounts are strongly idealized and contain falsehoods (Potochnik, 2017). Be-

ing aware of the puzzling character of idealization or the explanatory incompleteness of models, Craver and Kaplan, offer a more nuanced view of the completeness and accuracy constraints. They call it "the ontic notion of Salmon-completeness" and define this norm as follows: "The Salmon-complete constitutive mechanism for *P* versus *P*` is the set of all and only the factors constitutively relevant to *P* versus *P*`" (Craver and Kaplan, 2020, p.300). This ontic norm does not imply that a model of phenomena has to be complete. Not "all details are necessary", but only those explanatorily relevant. The ontic norm of Salmon-completeness itself points out that scientific explanation should be precise about a given *explanandum* via expressing it in contrastive terms (*P* versus *P*`). However, as formulated so, the ontic sense of including in explanation everything that makes a difference to the precise phenomenon in question, does not fulfil "ontic" within the OC. In the latter case "ontic" was referring to the fact that the things in the world do account for phenomena. The ontic norm of Salmon-completeness implies that only some details are necessary for explaining the phenomenon in the broader class of explanatory relevance. This means that ontic constraints play their explanatory role if they are referred to the proper class of epistemic relevance. Thus, employing the ontic norm of Salmon-completeness pushes us to adopt the non-ontic conception of explanation.

In the case of epistemic constraints, the following are crucial: 1) heuristic strategies of decomposition and localization, 2) abstraction and idealization. In the first case, Bechtel and Richardson (2010) rightly note that such heuristic strategies can be seen as basically compatible with the above-mentioned ontic spatiotemporal constraints. The main point of these strategies is to approximate the behaviour of the system based on the interaction of working parts of mechanisms. These strategies need, at the same time, the integration of functional

and structural descriptions of mechanisms and the mapping of activities to working parts (as suggests, e.g., the 3M constraint). This shows that epistemic and ontic constraints can and should be combined in mechanistic inquiry. The strength of decomposition and localization is that they facilitate "an increasingly realistic representation of the explanatory domain, even when the initial representation is seriously distorted" (Bechtel and Richardson, 2010, p.8).

The case of abstraction and idealization strategies suggests that there is no straightforward mechanism-model-mapping in the case of scientific explanation (Parker, 2020). If we use the metaphor of the map, one rather should say that scientists offer the atlas of "explanatory maps" when dealing with *explananda*. In fact, the relation between different mechanisms and the MEx representing stuff in the world may be further illuminated by applying such a metaphor. R. Giere, when discussing the issue of representation, points out that maps represent spatial regions from particular perspectives determined by various human interests (Giere, 1999, pp.81–82). According to him, the operative notion to describe the relationship between models and the world, is not the truth, but rather the similarity or fit, between the model and the world. The map analogy shows that maps are always partial and that they are always maps of something. Map makers and map readers, using interpretative rules, must be able to understand what certain maps represent and how to understand conventions used to prepare the map. The model-based understanding of scientific theorizing proposed by Giere means that scientists generate models using principles, specific conditions, and focus on relevant aspects and relevant degrees of similarity between the model and the target (Giere, 2004).

Craver, well aware of the model-based character of MEx, while defending OC, wants to avoid the problem of idealization by stressing

that "terms like 'true', 'idealized', and 'abstract' apply to representations or models. They do not apply to the ontic structures they represent" (Craver, 2014, p.50). He further argues that the problem of idealization is not a philosophical problem of explanation but of reference. According to him, "the very idea of an idealized model of an explanation commits one, at least implicitly, to the existence of an ontic explanation against which the model can be evaluated" (Craver, 2014, p.50). Contrary to Craver, I think that the very idea of an idealized model faces us with the problem of how such models convey explanatory information. Moving the issue of the idealization from the philosophical domain of explanation to that of reference, would not be a solution to the problem of explanation. When looking at the history of science, we find cases where one can achieve understanding without having a true representation of facts. For instance, one can evoke Maxwell's fictional vortex model treating light as electromagnetic radiation (de Regt, 2015), showing that not only the veridical representations can be explanatory (Bokulich, 2016). It is true that "the goal of building an explanatory text is not to provide the illusion of understanding but rather to provide bona fide understanding" (Craver, 2014, p.49). However, the commitment to a truthful account conveying explanatory information works more as an explanatory ideal in scientific practice than the matter of facts. An *explanandum* is the sum of observational data and concepts, models, hypotheses, etc. Both empirical data and scientific conceptualization have their own limits and only describe a part of reality.

Whether science obtains an explanation of reality or the representations of reality is the question that should not receive yes/no answer if we want to further explore how scientific explanations and objects of scientific investigation are formulated. Following J. Bickle I argue that there is no straightforward mapping of mechanisms onto the world in

the case of scientific explanation. J. Bickle (2008) enumerates four key principles in explaining neural mechanisms. These are principles of observation (occurrences of the hypothesized mechanism are strongly correlated with the occurrences of the behaviors used as experimental measures), negative alteration (intervening to decrease activity of the hypothesized mechanisms must reliably decrease the behaviors used as experimental measures), positive alteration (intervening to increase activity of the hypothesized mechanisms must reliably increase the behaviors used as experimental measures), and integration (the hypothesis about the causal nexus that produces the behaviors used as experimental measures must be connected with as much experimental data as is available about the hypothesized mechanism). These convergent four principles show that the relationships between phenomena and entities' activities are to be discovered and described in a piecemeal way rather than merely "given to us" (whatever the latter claim would mean). These relationships are rather "inferred based on correlations between changes in monitored behaviours or effects that are taken to be indicative of changes in these phenomena and activities. [. . .] This makes explanation fundamentally epistemic" (van Eck, 2015, p.15).

The MEx basically involves, on the one hand, conveying an explanatory and empirical understanding of how entities and activities are organized in the production of the phenomenon to be explained. The explanatory understanding can be expressed as a kind of understanding-why, while the empirical understanding means that we are dealing with domains of empirical inquiry (Khalifa, 2017, pp.1–3). On the other hand, mechanistic understanding results are intimately connected to expertise in the specific scientific field or research programme. In fact, understanding results from the cooperative scientific enterprise performed by many scientists, using a plurality of methods

and experimental practices. I concur with Khalifa that representing mechanisms or causal structures may be treated as one of the local constraints that should be satisfied in addition to the global constraints (such as, the *explanandum* is approximately true, the *explanans* makes a difference to the *explanandum*, the *explanans* satisfies our reasonable ontological requirements) placed on the explanation. This focus on representational processes conveying understanding does not imply a purely psychologistic view on scientific explanation, but rather explores the space of complex methodological and experimental aspects present within the scientific endeavour. It is beyond the aim of this paper to treat this issue in more detail, but it is certainly worth of being further explored.

It is now time to take stock. First of all, in this section I focused on the ontic and epistemic norms and constraints of good MEx, which apply to non-ontic conceptions of explanation. This entering into the intra-epistemic debate has shown the model-based character of MEx and the importance of mechanistic understanding if one is dealing with the specific accounts of explanation. In my analysis of EC and critics of OC, I followed the convention that "sentences are strings of visual or audible symbols that express propositions; propositions are the abstract entities that carry the meaning of the sentences; facts are concrete things in the world and, unlike sentences or propositions, are not capable of bearing truth or falsity" (Khalifa, 2017, p.148). Secondly, the current debate on conceptions of explanation has shifted from what explanations are towards a discussion on the norms and constraints of MEx. Arguing for normative constraints of explanation was called by some authors "the normative turn" (Sheredos, 2016, p.921; Wright and van Eck, 2018, p.1023). According to these authors, with whom I concur, this debate shows that the OC after the normative

turn is abandoned (Bokulich, Sheredos, van Eck, Wright) or at least "in retreat" (Potochnik, 2018). In what follows I want to shed further light on the consequences of "the normative turn".

	Ontic conception (OC)	Epistemic conception (EC)
Main representants	W. Salmon, C.F. Craver, L. Darden, S. Glennan, P. Illari, M. Povich, T. Knuuttila	W. Bechtel, C. Wright, B. Sheredos, A. Bokulich, A. Levy, M. Nathan, D. van Eck, R. Frigg, H. de Regt
Core idea	the causal structure of the world, that is, the entities and activities and the organization by which they produce the phenomenon of interest	an intelligible model of the activities, entities and their organization that scientists can understand, manipulate, and communicate, in order to move ahead in the research program
Vehicles of explanation	full-bodied things; neither representations nor texts these things can be: causes, mechanisms in the world, facts, events, set of factors	representations of mechanisms in the world; such as internal mental representations or external to the cognitive agent (diagrams, linguistic descriptions, mathematical equations, physical models, etc.)
Norms	truth	explanation aims at understanding ontic norms: the instruction to describe the causal structure of the mechanism epistemic norms: to increase intelligibility, generality, systematicity, integration
Constraints	all and only the relevant features of the mechanism in question	ontic constraints: accuracy and completeness, spatial and temporal, mechanism-to-model-mapping (3M) epistemic constraints: heuristic strategies of decomposition and localization, abstraction and idealization

4. What is left after "normative turn"?

P. Illari (2013, pp.248–252) has tried to argue that both ontic and epistemic constraints should be recognized and reconciled since both of them are essential. "Without the first constraint, we are not explaining the production of a phenomenon by a mechanism; without the second, we do not achieve the understanding essential to explanation" (Illari, 2013, p.250). Her reading of ontic and epistemic constraints clarifies that the goal of science to reveal the causal structure of the world and the goal of science to achieve a communicable understanding are both necessary. This sort of clarification is unproblematic. However, when Illari tries to spell out differences between Craver's (OC) and Bechtel's (EC) accounts, she discusses the case of prioritising one norm over the other. I agree that we may grant priority to the ontic or epistemic norms, but such a move should not be guided by *a priori* principles which we employ in scientific reasoning. Moreover, according to her account, both norms work together in order to generate a successful MEx. Without the first one we cannot describe the (causal) structure of the world, while without the second one we cannot build a model of the activities, entities and their organization (Illari, 2013, p.250). Again, I wholeheartedly agree with the relevance of ontic and epistemic norms and normative constraints for building MEx in an integrative way. Let us now focus on how to provide such an integration.

L. Kästner and P. Haueis (2021) fill the gap left by Illari, since they convincingly show how ontic and epistemic norms work together through mechanistic inquiry as a whole. First of all, they emphasize that mechanistic discovery typically starts with characterizing phenomena via different epistemic activities (e.g., modeling, experimenting) in which scientists perform various epistemic operations

(e.g., injecting a current in a neuron) to track different activities and parts of the mechanism which is investigated. Secondly, they argue that epistemic activities containing different models, skills and instruments in the process of discovery do not stand in opposition to the realism of the entities and activities constituting mechanisms actually investigated. In other words, there are interacting multiple dimensions in the elaboration of scientific explanation, rather than a clear-cut alternative between ontic and epistemic norms and constraints. It stems from their analysis that it remains crucial to trace out different stages of the discovery process, with some phases related to the ontic point of view, and others more oriented to epistemic aspects. What is the most interesting aspect of their paper (Kästner and Haueis, 2021)(Kästner and Haueis, 2021) is that they aim to show how ontic constraints may directly or indirectly constrain some epistemic activities; and vice versa, i.e., they show how epistemic norms may help in choosing the accuracy or completeness ontic norms when dealing with empirical findings that conflict with the currently most plausible models of a mechanism.

Kästner and Haueis, in a similar way to Illari, have bracketed the metaphysical issue of explanation from their discussion, by claiming that mechanistic inquiry is both ontically and epistemically constrained. They have gone further than Illari, since apart from integrating both sets of norms and constraints they have shown how ontic constraints guide mechanism discovery from the bottom up and how epistemic constraints help with anomaly resolution. It seems that the main solution to the controversy about the priority of certain norms or constraints comes from considering both of them from a reconciliatory and diachronic perspective. That is, in scientific practice fulfilling both epistemic and ontic norms "requires a kind of dynamic and temporally extended zig-zag between distinct explanatory prac-

tices" (Sheredos, 2016, p.943). In fact, Illari concludes her paper by noting that any successful disentangling of our theories, ontically and epistemically constrained, "will need to look at what is happening over time, rather than at a single time" (Illari, 2013, p.254). "Our good mechanistic explanations are always the result of a struggle to satisfy both ontic and epistemic constraints" (Illari, 2013, p.254). It is the apt correlation between the methods of observation/analysis (epistemic aspect) and the data itself (ontic aspect), that brings about the success of MEx. The ontic and the epistemic norms and constraints are neither alternatives nor directly reconcilable in a simple way. They express different moments of the explanatory strategies and complementary aspects of explanations aimed at the *explananda*. Only the integration of both aspects serves explanatory purposes.

The simplicity of the last claim is a bit striking, considering how long lasting the metaphysical debate on what scientific explanations are and the establishment of criteria for good scientific explanations. Although these debates were evolving in a parallel way, it is important to not conflate them. What gave rise to the debate between OC vs EC, as I suggested at the end of the previous section, was adopting the language suggesting that stuff in the world performs explanatory acts. But causal mechanisms are concrete things in the world and, unlike sentences, propositions, models, etc., are not capable of bearing truth or falsity. Finally, in order to further argue for the more nuanced view on the reconciliation and integration of various norms and constraints, I want to bring up the problem of the boundaries and levels of mechanisms.

The criteria for individuating the boundaries between entities, activities, and mechanisms themselves are the most difficult problems to solve (Kaiser, 2017). There are different principles that may be of help in carving the mechanisms and their components, such as the

individuation of natural boundaries of biological objects (e.g., the cell membrane, the skin, the chain of mountains that borders a specific ecosystem) (Darden, 2008) or deciding upon the strength of interactions (e.g., interactions among parts are generally conceived of as stronger than interactions between parts and environment) (Wimsatt, 1974). Even if one applied these principles, it is possible to get different results, since decompositions of a mechanistic system into parts depends on the explanatory context. For instance, the human body has many systems which are responsible for the various activities of the body, such as the cardiovascular, the respiratory, the nervous, the endocrine, the muscular-skeletal system, etc. The difficulty is that the part decompositions generated by the phenomena will play out in the body in different overlapping ways, e.g., the arteries and veins of the cardiovascular system are also involved in the respiratory system (Kaiser, 2017, pp.37–38). Hence, how we "cut nature at its joints" depends upon the circumstances in different contexts of explanation.

In this context, L. Kästner (2018), rightly notes that the mechanistic talk of levels does not seem to be a satisfactory way of defining what is at the same level in terms of local composition relations. She proposes instead an approach of epistemic perspectives, largely indebted to Giere's (2006) approach, which I have already mentioned. Kästner characterizes these perspectives along five dimensions (i.e., resolution—different temporal or spatial scale; specificity—what kinds of things can be detected from certain a perspective; point of view—depending on the background theories and taxonomies assumed within a certain perspective; sensitivity—with respect to specific factors; scope—allowing for investigating different portions or aspects of phenomena depending on various methodological con-

straints) and stresses that the characterization and individuation of them is primary an empirical affair. She explicates the advantage of such an approach in the following words:

> Once the relations between them are worked out, epistemic perspectives characterized along dimensions [. . .] give us the means to locate observed entities and activities at 'the same level' or 'different levels', a way of distinguishing different kinds of dependency relations (such as causal and constitutive relations in mechanisms), and a solid foundation for integrating multiple scientific observations into complex multi-perspective mechanistic explanations. We can thus piece together the mechanism mosaic while avoiding the problems associated with local, compositionally related, levels of mechanisms in this context (Kästner, 2018, pp.77–78).

I am very sympathetic to Kästner's view that both the decomposition of mechanisms and carving their levels deeply depend upon contextual elements (Woodward, 2008, pp.217–220). In fact, "there is an inherent perspectival aspect" (Darden, 2008, p.960) in the case of levels or boundaries of mechanisms. Such perspectivalism, however, does not necessitate "arbitrary choices in individuating phenomena and mechanisms" (Darden, 2008, p.960), but evidences the search for context-sensitive considerations which are the best from the point of view of specific explanatory purposes. "That different models and observations depend on our purposes does not imply, however, that they do not show anything real. They just emphasize different features, carve out different aspects, or provide different filters" (Kästner, 2018, p.76). I think that MEx presents "a more sophisticated picture of the relation between realistic representation and explanatory understanding" (de Regt, 2015, p.3795), as mentioned in the previous section. On the one hand, epistemic perspectives on explanations show that

scientific knowledge is inextricably bounded to the modeler's knowledge, employed tools and pragmatic concepts. On the other hand, the perspectival approach does not merely trigger unlimited explanatory pluralism. Rather, admitting many epistemic perspectives is useful in the function of finding the best explanation in a certain context.

Analysing the MEx through epistemic perspectives sheds further light on the relevance of ontic and epistemic norms and constraints for building MEx in an integrative way. It shows that mechanisms in the world are not doing the explanatory job, but on the contrary, investigating a phenomenon from multiple different perspectives is what builds the MEx. Such an epistemic conception of MEx does not entail neglecting the objectively real character of causal mechanisms, but forces us to admit the partial and perspectival character of explanatory work.

5. A dual ontic-epistemic approach

The aforementioned characterization of boundaries, levels or causal mechanisms fits the EC very well, i.e., boundaries, levels and causal mechanisms constitute idealized representations of processes. In the case of causal mechanisms, D. Nicholson (2012, p.160) rightly argues that "explanations always presuppose a context that specifies what is to be explained and how much detail will suffice for a satisfying answer, [. . .] it is this very epistemic context that determines how causal mechanisms are individuated and what details are featured in them". If this is so, it seems that the term "mechanism" does not necessarily refer to worldly causal mechanisms. What does it refer to then?

When speaking about mechanisms, it is quite intuitive to think about mechanisms "out there in the world" and MExs or models that depict them. However, as Nathan (Nathan, 2021, pp.171–172) notes, the term "mechanism" is ambiguous and may refer to both aspects, that is, to things in the world and to their representations. Since MExs represent entities and activities in the world, it may seem to be quite obvious that claims about representations are claims about the mechanisms in the world. In fact, such a blending together of two distinct claims may stem from the OC erroneous assumption that "ontic explanations are not texts; they are full-bodied things" (Craver, 2014, p.40). The ambiguity of the term "mechanism" would be then an additional evidence of this initial error. Although the distinction between mechanisms and mechanistic models is at the core of mechanistic literature (Glennan, 2005), the free employment of the term "mechanism" to both entities and their representations probably has further added to the confusion that stuff in the world performs explanatory acts. One of the consequences of the aforementioned epistemic perspectival approach would be the claim that real mechanisms are not just the represented mechanisms.

If mechanisms "out in the world" should be kept distinct from the represented mechanisms, what about the use of the term "mechanism"? Should it be abandoned? According to Nathan (2021, pp.162–190), NMP has the merit of having shown how the concept of mechanism always has a two-sided nature: on the one hand, mechanisms are "out in the world", on the other hand, they are a model-theoretic construction and a mode of scientific representation. Mechanisms are central for constructing a scientific theory, although they are "black boxes". This means that mechanisms work as placeholders—frames or difference-makers—in a causal explanation represented in a model (Nathan, 2021, p.133). For Nathan, the practice of black-

boxing consists of three phases: 1) the framing stage of specifying the *explanandum*, 2) the difference-making stage of providing a causal explanation of the *explanandum*, 3) the representation stage which determines how the difference-makers should be portrayed via the abstraction and idealization strategies. Mechanistic models are distinguished by their underlying structure (e.g., real physical entities representationally related to some abstract systems) and their interpretational capability. Both actual reference and imagined elements are part of the model, and the model's interpretation must explain their relationship and how the model represents phenomena.

It is important to emphasize that representations in scientific practice are not only involved in the *explanans*, but are also the *explanandum*, which corresponds to Nathan's first stage of black-boxing. This is so because scientists explain the phenomenon-as-represented. In other words, the explanation is always conceptualized within the particular explanatory context (Bokulich, 2011). The phenomenon-as-represented means a representation shared by a community of researchers, not just the subjective state of cognition or particular linguistic or conceptual description. That these explanatory choices are not arbitrary is due to the fact that "the world constrains which representations are, or are not, going to be adequate for a given explanatory context" (Bokulich, 2018, p.802). Such a conception of explanation recognizes the presence and joint working of both ontic and epistemic norms and constraints.

My colleagues and I have discussed the problem of the definition of species, suggesting that the concept of species is better understood within a dual ontic-epistemic approach (Marcacci, Oleksowicz and Conti, 2023). We have chosen this concept as the case study, since it plays various roles in biological investigations. On the one hand, it is based on certain ontic assumptions about the properties of species

and the cause of an individual's belonging to a species; on the other hand, it describes this membership as a consequence of domain specific methodological operations and conceptual assumptions. First of all, we argued that the concept of causal mechanism is not a decisive argument for realism about natural kinds or species. In fact, both the consideration of underlying causal processes and explanatory interests play an indispensable role in recent approaches to natural kinds. Secondly, in agreement with Kästner's view on the possible variety of ontological commitments of epistemic perspective approach, we contend that the perspectival character of the concept of mechanism is not a decisive argument for nominalism about natural kinds or species. Thirdly, the application of MEx entails theory-dependent pluralism about natural kinds or species.

Let us briefly further comment on these three points. The last one means that the explanatory pluralist stance is the viable option from the mechanistic point of view. The interest-relativity of scientific classificatory categories, such as species, entails that various epistemic strategies and constraints remain essential and irreducible features of any biological explanation in case of species. Referring to the first and the second point, we argued that neither realism nor nominalism about natural kinds or species is the solution to the problem. The viable solution, if species are to be grasped within an ontic/epistemic approach, can be formulated in the following form:

> Close attention to the various accounts of species shows that they play important and distinct roles within the sciences. On the one hand, they work as metaphysical posits; on the other, as explanatory postulates. In the former case, the objectivity of species is what grounds the objectivity of explanations, and sound explanations of species require us to identify the relevant factors at work in evolutionary processes. As explanatory postulates, they play a specific role restricted to the context of

> a particular theoretical framework or model (Nathan, 2023).
> In this case, species are essentially preliminary hypotheses or
> theoretical units awaiting to be replaced by more perspicuous
> explanatory elements (Marcacci, Oleksowicz and Conti, 2023,
> p.12).

This distinction between species *de re* and explanatory species provides the conceptual resources to rethink the presence of ontic and epistemic norms and constraints in the long-standing debate on natural kinds and the notion of species. This distinction parallels that one between mechanisms out in the world and explanatory mechanisms. Both these distinctions help us to note that while proponents of OC would argue that species *de re* or mechanisms *de re* do the explanatory job, proponents of EC argue that this is not the case. The inconclusiveness, as it may seem, of the debate on boundaries or levels of mechanisms, or on the definition of species can be expressed through the following philosophical maxim: "distilling metaphysical implications from scientific explanations requires close attention to explanatory practice" (Love and Nathan, 2015, p.773). If contextual aspects in the explanations are unavoidable, then we should moderate the ontological implications drawn from our models. Not because there is "nothing out there", but because there is in fact more than we expect to be there. The epistemic-relativity of our categories, explanations and models can be compatible with their objectivity. The mechanistic strategy is the illuminating one, since it pays attention to how epistemic and ontic norms and constraints are intertwined within scientific explanation, and how such an objectivity may be achieved.

Conclusions

NMP has the merit of having shown how the concept of mechanism as a tool for representation has always a two-sided nature: on the one hand, mechanisms are things in the world, on the other hand, they are model-theoretic constructions. Mechanisms are central for constructing a scientific theory, although they are theoretical placeholders, that is, indicators of what a theory is built around. If one neglects this feature of the term "mechanism", one will not understand the importance of integrating ontic and epistemic norms and constraints within MEx.

Today the crux of the debate is the interplay between different norms and constraints in providing MEx. It seems that the scale is tipped towards the proponents of EC. The OC remains under attack for the set of reasons previously discussed: the stuff in the world does not perform explanatory acts; the crucial role of general principles in scientific reasoning; scientific reasoning relies heavily on abstraction/idealization/generalization; in scientific practice, we deal rather with highly idealized representations of causal processes rather than with real mechanisms; we do not have direct access to the phenomenon, but we deal with phenomena-as-represented which are previously conceptualized; the currently dominant modelling strategy of science evidences the crucial role of abstraction and idealization, etc. Briefly put, there is no scientific explanation without intentional agents who try to decipher things out there in the world. However, this does not imply that there are no things "out in the world".

What are then the main conclusions? First of all, the ontic features of the world do not alone settle all questions about the adequacy of an explanation, but the latter is rather settled by an evaluation of our explanatory practices and the features of our models. An increased

awareness of the presence of multiple dimensions within scientific explanations (e.g., ontic and epistemic norms and constraints involved in the elaboration of the concept of mechanistic boundaries or levels, or the definition of species), can aid in comprehending distinct strategies employed in the various sciences, and help to understand how they cooperate to adequately account for complex phenomena. This brings us to the last but not least conclusion.

The ontic and the epistemic norms and constraints express different moments of the explanatory procedures and two complementary aspects of specific accounts of explanations aimed at the *explananda*. The latter fact implies that only the integration of both aspects serves explanatory purposes. My analysis of the ontic/epistemic debate shows that one cannot simply read off truths, or the truth about what the world is like. At the same time, it does not imply that every claim is equivalent. On the contrary, via the use of our models, concepts, and theories, we have only a piecemeal formulation of partial knowledge about reality. Dependence on how the world is will be a commitment entailed by any good conception and account of explanation.

Bibliography

Andersen, H., 2014a. A field guide to mechanisms: Part I. *Philosophy Compass*, 9(4), pp.274–283. https://doi.org/10.1111/phc3.12119.

Andersen, H., 2014b. A field guide to mechanisms: Part II. *Philosophy Compass*, 9(4), pp.284–293. https://doi.org/10.1111/phc3.12118.

Bechtel, W. and Abrahamsen, A., 2005. Explanation: a mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), pp.421–441. https://doi.org/10.1016/j.shpsc.2005.03.010.

Bechtel, W. and Richardson, R.C., 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press.

Bickle, J., 2008. Real reduction in real neuroscience: metascience, not philosophy of science (and certainly not metaphysics!). In: J. Hohwy and J. Kallestrup, eds. *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Oxford: Oxford University Press, pp.34–51. https://doi.org/10.1093/acprof:oso.

Bokulich, A., 2011. How scientific models can explain. *Synthese*, 180(1), pp.33–45. https://doi.org/10.1007/s11229-009-9565-1.

Bokulich, A., 2016. Fiction as a vehicle for truth: Moving beyond the ontic conception. *The Monist*, 99(3), pp.260–279. https://doi.org/10.1093/monist/onw004.

Bokulich, A., 2018. Representing and explaining: The eikonic conception of scientific explanation. *Philosophy of Science*, 85(5), pp.793–805. https://doi.org/10.1086/699693.

Campaner, R., 2013. Mechanistic and Neo-mechanistic accounts of causation: how Salmon already got (much of) it right. *Metatheoria*, 3(2), pp.81–98.

Craver, C.F., 2006. When mechanistic models explain. *Synthese*, 153(3), pp.355–376. https://doi.org/10.1007/s11229-006-9097-x.

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

Craver, C.F., 2014. The Ontic Account of Scientific Explanation. In: M.I. Kaiser, O.R. Scholz, D. Plenge and A. Hüttemann, eds. *Explanation in the Special Sciences: The Case of Biology and History*. Vol. 367, *Synthese Library*. Dordrecht: Springer Netherlands, pp.27–52. https://doi.org/10.1007/978-94-007-7563-3_2.

Craver, C.F. and Darden, L., 2013. *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago, IL: University of Chicago Press. Available at: <https://press.uchicago.edu/ucp/books/book/chicago/I/bo16123713.html> [visited on 29 September 2023].

Craver, C.F. and Kaplan, D.M., 2020. Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, 71(1), pp.287–319. https://doi.org/10.1093/bjps/axy015.

Darden, L., 2008. Thinking again about biological mechanisms. *Philosophy of Science*, 75(5), pp.958–969. https://doi.org/10.1086/594538.

van Eck, D., 2015. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Axiomathes*, 25, pp.5–22. https://doi.org/10.1007/s10516-014-9243-x.

Frigg, R., 2022. *Models and Theories: A Philosophical Inquiry*. London: Routledge. https://doi.org/10.4324/9781003285106.

Giere, R.N., 1999. *Science Without Laws*, *Science and its conceptual foundations*. Chicago: University of Chicago Press.

Giere, R.N., 2004. How models are used to represent reality. *Philosophy of Science*, 71(5), pp.742–752. https://doi.org/10.1086/425063.

Giere, R.N., 2006. *Scientific Perspectivism*. Chicago, IL: University of Chicago Press.

Glennan, S., 2002. Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), pp.342–353. https://doi.org/10.1086/341857.

Glennan, S., 2005. Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. Mechanisms in biology, 36(2), pp.443–464. https://doi.org/10.1016/j.shpsc.2005.03.011.

Illari, P., 2013. Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, 78(S2), pp.237–255. https://doi.org/10.1007/s10670-013-9511-y.

Kaiser, M.I., 2017. The components and boundaries of mechanisms. In: S. Glennan and P.M. Illari, eds. *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, *Routledge handbooks in philosophy*. London; New York: Routledge, pp.116–130.

Kaplan, D.M. and Craver, C.F., 2011. The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78(4), pp.601–627. https://doi.org/10.1086/661755.

Kästner, L., 2018. Integrating mechanistic explanations through epistemic perspectives. *Studies in History and Philosophy of Science Part A*, 68, pp.68–79. https://doi.org/10.1016/j.shpsa.2018.01.011.

Kästner, L. and Haueis, P., 2021. Discovering patterns: On the norms of mechanistic inquiry. *Erkenntnis*, 86(6), pp.1635–1660. https://doi.org/10.1007/s10670-019-00174-7.

Khalifa, K., 2017. *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108164276.

Levy, A., 2013. Three kinds of new mechanism. *Biology & Philosophy*, 28(1), pp.99–114. https://doi.org/10.1007/s10539-012-9337-z.

Levy, A., 2014. Machine-likeness and explanation by decomposition. *Philosophers' Imprint*, 14(6), pp.1–15.

Love, A.C. and Nathan, M.J., 2015. The idealization of causation in mechanistic explanation. *Philosophy of Science*, 82(5), pp.761–774. https://doi.org/10.1086/683263.

Machamer, P., Darden, L. and Craver, C.F., 2000. Thinking about mechanisms. *Philosophy of Science*, 67(1), pp.1–25. https://doi.org/10.1086/392759.

Marcacci, F., Oleksowicz, M. and Conti, A., 2023. Ontic and epistemic differentiation: Mechanistic problems for microbiology and biology. *Foundations of Science*. https://doi.org/10.1007/s10699-023-09918-9.

Meheus, J. and Nickles, T., eds., 2009. *Models of Discovery and Creativity*. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-3421-2.

Nathan, M.J., 2021. *Black Boxes: How Science Turns Ignorance Into Knowledge*. Oxford, New York: Oxford University Press.

Nathan, M.J., 2023. Causation vs. causal explanation: Which is more fundamental? *Foundations of Science*, 28(1), pp.441–454. https://doi.org/10.1007/s10699-020-09672-2.

Nicholson, D.J., 2012. The concept of mechanism in biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1), pp.152–163. https://doi.org/10.1016/j.shpsc.2011.05.014.

Parker, W.S., 2020. Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, 87(3), pp.457–477. https://doi.org/10.1086/708691.

Potochnik, A., 2017. *Idealization and the Aims of Science*. Chicago, IL: University of Chicago Press. Available at: <https://press.uchicago.edu/ucp/books/book/chicago/I/bo27128726.html> [visited on 27 October 2023].

Potochnik, A., 2018. Eight Other Questions about Explanation. In: A. Reutlinger and J. Saatsi, eds. *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*. Oxford: Oxford University Press, pp.53–73. https://doi.org/10.1093/oso/9780198777946.003.0004.

de Regt, H.W., 2015. Scientific understanding: truth or dare? *Synthese*, 192(12), pp.3781–3797. https://doi.org/10.1007/s11229-014-0538-7.

de Regt, H.W., 2017. *Understanding Scientific Understanding*, *Oxford Studies in Philosophy of Science*. Oxford, New York: Oxford University Press.

Salmon, W.C., 1984. Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, (2), pp.293–305. https://doi.org/10.1086/psaprocbienmeetp.1984.2.192510.

Salmon, W.C., 1989. Four decades of scientific explanation. In: P. Kitcher and W.C. Salmon, eds. *Scientific Explanation*, *Minnesota studies in the philosophy of science*, 13. Minneapolis: University of Minnesota Press, pp.3–219. Available at: <http://conservancy.umn.edu/handle/11299/185700> [visited on 4 October 2023].

Sheredos, B., 2016. Re-reconciling the epistemic and ontic views of explanation (or, why the ontic view cannot support norms of generality). *Erkenntnis*, 81(5), pp.919–949. https://doi.org/10.1007/s10670-015-9775-5.

Wimsatt, W.C., 1974. Complexity and Organization. In: K.F. Schaffner and R.S. Cohen, eds. *PSA 1972: Proceedings of the 1972 Biennial Meeting of the Philosophy of Science Association*, *Boston Studies in the Philosophy of Science*, 20. Dordrecht: Springer Netherlands, pp.67–86. https://doi.org/10.1007/978-94-010-2140-1_5.

Woodward, J.F., 2008. Comment: Levels of Explanation and Variable Choice. In: K.S. Kendler and J. Parnas, eds. *Philosophical Issues in Psychiatry: Explanation, Phenomenology, and Nosology*. Baltimore: Johns Hopkins University Press, pp.216–235.

Wright, C., 2015. The ontic conception of scientific explanation. *Studies in History and Philosophy of Science Part A*, 54, pp.20–30. https://doi.org/10.1016/j.shpsa.2015.06.001.

Wright, C. and Bechtel, W., 2007. Mechanisms and psychological explanation. In: P. Thagard, ed. *Philosophy of Psychology and Cognitive Science*. Elsevier, pp.31–79. https://doi.org/10.1016/B978-044451540-7/50019-0.

Wright, C. and van Eck, D., 2018. Reconciling ontic and epistemic constraints on mechanistic explanation, epistemically. *Ergo, an Open Access Journal of Philosophy*, 5(38). https://doi.org/10.3998/ergo.12405314.0005.038.

Book reviews

Recenzje

Exploring the epistemic and ontic conceptions of *Models and Idealizations in Science*

Alejandro Cassini & Juan Redmond (eds.), *Models and Idealizations in Science: Artifactual and Fictional Approaches*, Springer Iternational Publishing, Cham 2021, pp.xv+270.

This book presents an insightful collection of papers that explore scientific modeling, idealizations, and representation from diverse philosophical perspectives. The book's first chapter offers an extensive, well-written introduction to the topic of scientific models, while the last chapter provides a relevant annotated bibliography on the philosophy of models and idealizations in science. The ten remaining chapters consist of previously unpublished articles that explore certain intricacies of both modeling practices and the ontology and epistemology of models. Although each chapter is self-contained and independent, one can find a few recurrent themes, such as analyses of artifactual and fictional approaches to modeling, or examinations of the role of surrogative rea-

soning and (de)idealization in modeling practices.

In what follows I provide a brief overview of these papers, pinpointing how some of their contributions might relate to the debate between ontic and epistemic conceptions of explanation.

Natalia Carrillo & Tarja Knuuttila's chapter examines idealization in scientific modeling. A big focal point of the debate on idealization concerns whether idealizations are beneficial or an epistemic deficiency to overcome. If one believes, as proponents of the ontic conception often do, that explanatory texts representing the ontic explanations need to be complete and accurate, then idealizations are something to be eventually replaced—whether by de-idealizing or by textually representing the 'actual explanation'. Here, de-idealization would not only be achievable, but would constitute a worthy goal. Those who adopt the epistemic conception tend to find intrinsic value in idealizations; for they allow scientists to identify the proper level of abstraction, enable selecting relevant factors, or are fundamental towards generalizing or towards unifying.

Nevertheless, both positions view idealizations as deliberate misrepresentations or distortions. Car-

rillo & Knuuttila reject this view by adopting an artifactual approach. In this approach models are epistemic artefacts, and idealization is a set of assumptions that align different representational tools to construct a model aiming to answer research questions. This approach highlights the difficulty in disentangling epistemic benefits and deficiencies in the model and challenges both the idea of easy de-idealization and of idealizations simply being distortions.

Sympathisers of the epistemic conception might find a strong argument here in that idealization makes the model possible in the first place (Cassini and Redmond, 2021, p.57), for example, by enabling the application of mathematical and computational tools. However, their analysis runs deeper by noting that labeling idealizations 'distortions' assumes one has enough knowledge about the target phenomenon, a flaw which may be present in both sides of the debate. Furthermore, talk about distortions obscures an important dimension of scientific modeling: exploring the possible (how phenomena could be produced) rather than actual. Here, the ontic conception presents a clear disadvantage, as it does not have the

resources to explicate such explanatory practices.

Mauricio Suárez & Agnes Bolinska's chapter apply communication theory to analyze the informational content of scientific models. They argue that models can be seen as communication channels, transmitting information about their targets, whereas idealizations and abstractions can be likened to sources of informational noise and equivocation. The authors argue that this analogy can clarify certain modeling practices—for instance, shedding light on the trade-offs involved in minimizing idealization and minimizing abstraction. In this analogy, the explanandum phenomenon is the informational source, whereas the model is the courier that codifies the information.

Surprisingly, the chapter does not discuss machine learning, despite the numerous parallels between their account and machine learning techniques. The authors appear to have missed an opportunity to establish a more fruitful analogy, especially since machine learning models are often explicitly used as scientific models. For instance, the encoder-decoder architecture shares many similarities with the examples discussed in the chapter and is often

employed in physics' simulations. In these contexts, concepts like loss functions, noise, or the dimensionality of encoder/decoder can provide more meaningful positive analogies for the topic at hand.

Nonetheless, the idea of quantifying idealization in terms of noise may be of interest to defenders of the ontic conception, as it measures how 'far away' a complete and accurate explanatory text might be. Detractors might however note that this overlooks the fact that idealizations often enhance the representational relationship between the model and target by sharpening the focus on what is of interest and 'carving the world' at the right seams.

Staying within the analogy, ontic proponents might see a pure signal devoid of noise or equivocation as a worthy goal (notwithstanding that a completely faithful representation is not equivalent to an ontic explanation). However, it is doubtful whether scientific modeling is actually concerned with trying to capture a pure signal (i.e., to represent faithfully, whatever that may mean). Much like Borge's perfect map, such a model would likely be of little use.

Alejandro Cassini's chapter discusses de-idealization in scientific models, emphasizing that its bene-

fits and drawbacks depend on the model's aim. He argues that de-idealization should enhance explanations, predictions, or model effectiveness, rather than seek faithfulness. While he frames the debate in terms of (non-)representationalists and (anti-)realists, he raises points that are relevant to the epistemic/ontic debate. For instance, some models cannot be de-idealized due to their holistic nature (e.g., certain meteorological models), idealization might be irreversible in minimal models, and it is undesirable to de-idealize in models whose purpose is mathematical tractability.

José Diez's chapter outlines a monist account of modeling. His account posits that scientific models are ensembles of entities and their relations, with some entities intended to stand for those in the target system. This includes a contextual constraint determining the required degree of accuracy for a given purpose. The account consists of explaining the conditions for performing a representation and analyzing the success or adequacy of an existing representation. Supposedly this provides a unified account of scientific modeling by providing necessary and 'substantive' conditions without relying on strong fictionalist elements. However, it is

unclear whether this account actually solves, as the author claims, the problem of representation ('in virtue of what does the target [successfully] represent the model?'). The proposed account seems descriptive as to certain success conditions for when one can claim there has been (successful) representation, but it does not answer why the representation was successful.

Roman Frigg & James Nguyen's chapter defends the fictionalist view of scientific models, which takes them to be analogous to characters and places in literary fiction. Their main argument lies in showing how several 'myths' often used to discredit this view are incorrect, highlighting that it is possible to combine the fiction view with an account of scientific representation. From the perspective of the ontic/epistemic debate, the first and third 'myths' are particularly interesting. The first myth suggests that the fiction view regards scientific products as falsehoods, while the third implies that the fiction view opposes representation. The first myth is tackled by separating two notions of fiction: infidelity and imagination. The fiction view supports the latter: models prescribe certain things to be imagined without committing to the truth status

model components. Thus, the fiction view conceives of fictions as tools for learning truths about the world. The third myth is tackled by underscoring that the fiction view primarily concerns the ontology of scientific models, not their representational content. Several approaches are then suggested to combine the fiction view with different accounts of scientific representation.

Many scientific developments can be traced back to fictional uses of the imagination. For the fictionalist, viewing models as fictions affords creative freedoms when investigating certain scenarios. While some might argue that this process should simply be seen as a simple heuristic which allows one to grasp the ontic explanation out there, there is also an argument to be made that fictions are part and parcel of the modeling endeavor.

Fiora Salis' chapter proposes an integrated fiction view for the ontology of theoretical models that combines insights from the fictional and artifactual perspectives. In the integration view, theoretical models are human-made artifacts, capable of serving different functions in various contexts while being analogous to fictional stories. These models are complex objects, consisting of model descriptions and proposi-

tional content. Model descriptions, which include linguistic and mathematical symbols prescribing specific imaginings, act as concrete representational tools and serve as props in a game of make-believe. Scientists build models by selecting (and interpreting) the model descriptions that best serve certain purposes and contexts. However, these imaginings do not imply the existence of any fictional entities. Model content is determined by model descriptions in collaboration with principles of generation.

By placing imagination at the center of the modeling process, the integration view resolves issues that challenge the artifactual view (2021, p.173), such as explaining model building and development, attributing concrete properties to model systems, clarifying the notion of representation of so-called representational tools, and addressing how scientists engage in model-world comparisons. Similarly, and partly by how model descriptions and content are separated, this approach also solves several problems which afflict the fictional view, such as the non-existence of models, the unclear relationship between model descriptions and imaginary systems, difficulties in scientists sharing the same imaginings, and issues with resemblance between imaginary systems and their targets. Model descriptions and content exist and can, therefore, stand in relation. Moreover, model descriptions serve as props that, through their prescriptions to imagine, enable and constrain an agent's imaginings and allow them to share said imaginings.

I found the integrated fiction account to be a noteworthy one that addresses several important issues in the literature. The idea of combining the artefact and the fiction view, while simple, is well executed and makes for a useful tool for the philosopher of science interested in modeling.

Manuel García-Carpintero's chapter posits that utterances about fictional entities and scientific models involve figurative language with clear truth-conditions. He applies this perspective to debates in semantics, specifically addressing supervaluationist models of indeterminacy.

Otávio Bueno's chapter presents a structural account of scientific representation, arguing that reification of structures as abstract entities is unnecessary. Instead, four different strategies are proposed: adopting a modal-structural interpretation of set theory, reconstructing relevant mathematics using second-

order logic, resisting the need for a metaphysical interpretation of set theory, and employing ontologically neutral quantifiers when quantifying over sets.

Juan Redmond's chapter presents an inferential conception of scientific representation addressing the question of how are models used to represent the world. He rejects the idea of correspondence between a model and its target, emphasizing the importance of how users use models through interactive and dynamic processes. This dialogic approach calls into question whether there can be an 'accurate and complete' (textual) explanation irrespective of the uses and users of a model.

Andrés Rivadulla's chapter advocates for an instrumentalist approach to theoretical models in the physical sciences, emphasizing their utility as tools for explaining and predicting phenomena rather than as representations. In his view, theoretical models are idealized constructions that facilitate calculations, explanations, and predictions in scientific inquiry. A key observation is that incompatible theoretical models exist for the same phenomena, emphasizing their use as tools rather than faithful representations. While this observation is perfectly compatible with an epistemic conception,

ontic proponents require additional effort to accommodate this observation. Here, they either have to resolve the puzzle of how there can be two or more valid ontic explanations for the same phenomenon, to show that in fact these models are targeting different phenomena, or to show that one (or more) of the incompatible explanations is incorrect.

Overall, this book is a valuable resource for philosophers of science, proficiently investigating topics such as modeling, the fiction view, and the artifactual view, as well as the role of de-idealization in scientific modeling. Although it was not originally focused on ontic and epistemic conceptions, it proves valuable for the debate by including several hidden gems that can be used to highlight problems with the ontic conception. This includes difficulties in accommodating various scientific practices (such as idealization and how-possibly modeling), the fact that there can be several possible explanations for the same phenomenon, or the fact that many explanations often target general, sometimes idealized, phenomena.

Kristian Campbell
González Barman
University of Ghent

Cassini, A. and Redmond, J., eds., 2021. *Models and Idealizations in Science: Artifactual and Fictional Approaches*. Vol. 50, *Logic, Epistemology, and the Unity of Science*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-65802-1.

Mechanisms 'all the way down'?

Mechanistic Explanations in Physics and Beyond, Brigitte Falkenburg and Gregor Schiemann, editors, *European Studies in Philosophy of Science*, Vol. 11, Springer International Publishing, Cham 2019, pp.220

The reviewed volume includes contributions presented at a conference in Dortmund, Germany, organized in 2016 by the *Académie Internationale de Philosophie des Sciences*, *The German Research Foundation*, and the *Technical University of Dortmund*. Most of the materials come from the conference presentations and gravitate around a central theme: applying the 'new mechanistic philosophy' to physics, where mechanistic explanations and models have not been typically used (although some contributors in Part III tackle the applications of mechanisms to economics, medicine, computer science, or tangentially, biology).[1]

Most contributions shed new light on some traditional topics in the philosophy of science, such as explanation, reductionism/emergentism, levels of reality, discovery, models, and so on. Specifically, as potential unifying themes, most authors delve into one or more of the following questions and issues:

(1) Are mechanistic explanations or models adequate somewhere else than in 'special science' (biology, neuroscience, cognitive science, psychology, where they originated) especially in 'fundamental science' (mainly physics)? Are these explanations better in physics than other types of explanation (nomological, mathematical, etc.)?

(2) If the concepts of mechanism and causation are closely interspersed, how do we generalize mechanistic philosophy to physics, where causation is not omnipresent and nomological explanations, models, and theories abound?

(3) Is the world composed of mechanisms 'all the way down' (up to and including the 'most fundamental entities in the world': quarks,

---

[1] Unless otherwise specified, this review refers to the pagination of the reviewed (Falkenburg and Schiemann, 2019) volume.

strings, spin networks, or whatever is 'down there')?

(4) Does the mechanistic world-view as developed during and right after the Scientific Revolution (Descartes, Newton, Leibniz, Huygens, Kant, etc.) overlap with the explanations and models of the 'new mechanistic philosophy' (Salmon, Bechtel, Glennan, Machamer et al., etc.)?

The reviewer and the reader will find several original and new ideas and proposals to address these questions.

As most readers know, the 'new mechanistic philosophy' (NMP) originated in general philosophy of science, specifically in the literature on explanations, models, and scientific progress. Conventionally, Salmon's works from the 1980s are taken as a starting point of the NMP (Salmon, 1984; 1989). NMP (Glennan, 2017; Illari and Glennan, 2017), also called in this volume the 'New Mechanism', is the constellation of ideas in the philosophy of science, epistemology, and metaphysics that addresses (1)–(4) and the like. However, NMP has been inspired by the practice of special sciences (biology, cognitive science, neuroscience, psychology, sociology, economics, computational science, engineering, etc.), and most case studies in the literature originate here, not in physics or chemistry. Before the 2016 conference in Dortmund, only a handful of authors had addressed (1) (Illari and Williamson, 2011; Kuhlmann and Glennan, 2014). The *Routledge Handbook of Mechanisms and Mechanical Philosophy*, which was published after the conference, includes no less than fourteen contributions in a section called 'Disciplinary perspectives of mechanisms' (Illari and Glennan, 2017). However, only one of its contributions makes any substantive reference to physics (Kuhlmann, 2017). As expected, the present volume fills this missing gap in the literature on mechanisms.

The philosophy of science literature on scientific explanation, models, and discovery has been dominated for at least four decades by approaches in the same vein as NMP. Many characterizations of mechanisms are available, and most would include concepts from metaphysics and epistemology: entities, activity, organization, parts, levels, structures, function, and so forth (Illari and Glennan, 2017). Above all, mechanistic philosophy explains the behavior of a composite system by

reference to (i) its parts and (ii) the interaction among these parts. However, given the breadth of the contributions and the attempts to generalize existing approaches to mechanisms over disciplinary boundaries, the volume needs a uniform definition with some chances of generalizing to 'physics and beyond'.

Rather than delving into the definitional complexity of mechanisms, most authors in this volume consider Glennan's delineation of 'minimal mechanism' as its working definition: 'A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon' (Illari and Glennan, 2017, p.17).

According to the mechanistic philosophy, the most successful explanations in special sciences, as enumerated in (1), are mechanistic, and special sciences are in the business of discovering mechanisms. This thesis about science's success and progress is sometimes associated with a different claim about the world: the world consists of mechanisms. The strong metaphysical commitment to mechanism is that the world is composed of mechanisms. This volume addresses whether there are mechanisms 'all the way down'. The difference between the methodological questions (1) and (2), the more metaphysical (and ontological) question (3), and the historical (4) pervades most contributions to this volume. Interestingly, the authors do not visit the potential difference between the ontic and the epistemic concepts of mechanisms *in physics*, which has been debated in the literature for decades (Salmon, 1984; Glennan, 2002; Wright, 2012; Illari, 2013).

A word of caution is in order here. One must acknowledge that scientific practice and informal language used to communicate science do *not* help this endeavor. The term 'mechanism' is explicitly and extensively used in all areas of physics and chemistry. The presence of a word in a discipline is illusory, nevertheless. Do terms such as 'mechanism' have the same meaning in physics, chemistry, 'special sciences', or beyond science, and into history, art, or religious studies? A related question is looming in the background: is NMP historically contiguous with the terms 'mechanism' or 'machine' as used during or right after the Scientific Revolution? The worry that we overuse or overreach some terms to include 'mechanisms', 'energy', 'entropy', 'information', and 'complexity' over disciplinary boundaries is genuine.

It is trite to say that many areas of physics use the suffix '-mechanics'. Falkenburg lists almost a dozen areas of physics where the term 'mechanism' is used extensively: generation of turbulences within fluids; tsunami generation by submarine mass flows; generation of turbulences within the accretion disk of astrophysical objects such as active galactic nuclei, quasars, or black holes; pulsation of stars and giant planets; transport, propagation, or diffusion of charged particles or photons; slowing down or acceleration of charged particles; interactions of particles in particle physics, including the Higgs 'mechanism'; the mechanism of decoherence in condensed matter physics. Falkenburg argues that NMP does not apply well or smoothly to most cases (2019, pp.84–85). How do we escape the linguistic trap and find genuine and relevant generalizations of mechanistic thinking in physics?

In the first contribution to the volume, S. Psillos & S. Ioannidis differentiate metaphysical claims, *e.g.* answers to (3), from methodological claims about scientific practice: explanations, models, and prediction as answers to (1) or (2). The former relates to concepts in metaphysics such as causation, part-whole relation, or levels of reality: the most natural gambit is to relate

mechanisms to causation as strongly as possible. The latter is a thesis about how science advances by providing explanations and models of the world. Science advances by discovering new mechanisms, stipulating new mechanistic explanations, or creating new mechanistic models. This is dubbed 'the mechanistic methodology'. Psillos & Ioannidis claim that addressing (4) is essential and that philosophers during the Scientific Revolution were interested in the stronger claim, the metaphysical mechanistic thesis. Descartes had developed a methodological thesis of continuity according to which properties of the invisible world are similar to properties of the visible world. Hence, one can assume that gravity is identical enough to the mechanisms governing the flow of liquids (whirlpools). Descartes and other so-called 'Old Mechanists' suggest that the world operates like 'machines' (as artifacts we are very familiar with, as we created them). If continuity is correct, there are mechanisms all the way down to gravity and beyond gravity. The controversy was alive during the 17th century: Newton denied that mechanistic explanations extended to gravity and preferred a more law-based explanation of gravity. Nevertheless, Huygens, Leibniz, among others, did not buy the idea that

gravity follows a law, but demanded that physics return to the mechanistic explanation of forces. Newtonian mathematical or non-causal explanations were the alternative to the Old Mechanistic metaphysics.

Reflecting on NMP, Psillos & Ioannidis think of mechanisms as an elaborate (theoretical) way of speaking of causal pathways in the world. This reduces the mechanisms to theoretical ways of talking about causation. Each scientific field uses a specific language to describe these causal pathways; in some disciplines (biology, genetics, neuroscience), this language must include mechanistic terms. However, the scientists decide what language describes "causal pathways" and not the metaphysicians. As this is a matter of language and pragmatism, a metaphysical commitment is unnecessary. The two authors conclude that the NMP is too metaphysically loaded, similar to the 'Old Mechanist Philosophy' (OMP). As a line of criticism, Psillos and Ioannidis conclude that according to the more plausible methodological mechanism thesis, mechanisms can be replaced or displaced in any scientific discipline as needed.

In a more historical vein, the co-editor (G. Schiemann) argues that one way to answer (4) is to assume that the ontological commitment to mechanisms during and after the Scientific Revolution was more preëminent, although the NMP inherited partially a peculiar ontological commitment from OMP. He takes the ontological commitment of the physicists and philosophers during the Scientific Revolution as either monistic or dualistic, based on their commitment to the existence of matter and forces. Newton qualifies as a dualist in this sense. 'The early modern pair of concepts of matter and force is structurally related as regards the contrast it draws to the contemporary conceptual pair of entity and activity' (2019, p.43). Schiemann extends this monism-dualism distinction to contemporary mechanistic philosophy: Glennan qualifies as a monist, as activities and organizations are properties of parts. In the NMP camp, Machamer et al. are dualists because entities and activities are fundamentally different. Another touching point between the OMP and the NMP is a commitment to levels of reality (as Schiemann argues, both the contemporary monists and the dualists need this division).

Importantly, D. Dieks' contribution confronts questions (1)–(4) head-on. Starting from the contrast between Descartes' mechanis-

tic worldview and Newton's mathematical formalism of physics, Dieks emphasizes some problems that Newtonians had with concepts such as 'time,' 'becoming,' or causation, which were needed in explaining phenomena beyond the simple kinematics. However, these concepts are not mechanistic in nature. Moreover, the Lagrangian and Hamiltonian approaches to mechanics use mathematical and non-causal concepts (energy, momentum, phase space, symmetries, etc.) that are not immediately compatible with OMP.

Maxwell's equations for electromagnetism were initially formulated as mechanical properties of a medium (the ether), like Descartes' vortices. Still, their contemporary interpretation has no mechanical medium, and the most likely interpretation is Lagrangian (as noted by Poincaré). Mechanical models of electromagnetism are possible, and they come in handy as long as one is committed to the existence of the ether. The model with ether also assumes an infinite number of point charges as one can choose the number of particles interacting to be anything, even infinite. Dieks suggests that a field interpretation of electromagnetism taints a mechanistic interpretation of Maxwell's equation and sides with Poincaré, who thinks

that mechanistic models are underdetermined in this case.

As other authors in this volume suggest, there is a strong connection between causation and the mechanistic worldview. Although causation in classical electromagnetism can be accommodated (given some caveats), Dieks claims that causation must be local to operate in the minimal model definition of Glennan: activities, organization, and interaction must be localized. Although this may work in classical electromagnetism, action-at-a-distance is most likely incompatible with the mechanistic worldview. The structure of the quantum space of a simple system (composed of two particles only) is too rich and far too non-local for a mechanistic view at this level. The superposition principle, which is a core element of quantum mechanics, vindicates the view that the state of a composite quantum system cannot be reduced to the states of its parts. Knowledge of all the properties of the two parts is not enough to determine the state of the whole. Dieks uses here an idea promoted by Glennan and Kuhlmann: decoherence masks the quantum properties of systems and gives us the illusion of semi-classical behavior that can be captured by a mecha-

nistic view—i.e., where causation and mechanistic models may work. Dieks argues against Kuhlmann and Glennan's caveat that restricting in practice quantum models to cases of decoherence does not make these models plausible: 'So there are features of reality, detectable in principle, that show that the literal content of the ontological claims of the mechanistic explanation strategy is false' (2019, p.60). For Dieks, mechanical models in quantum systems may work as idealized and simplified views about the world (conditionalized on the strength and peculiarity of decoherence), but they are not enough to ontologically commit quantum mechanics to the mechanistic worldview.

If the description of the world proffered by quantum mechanics is non-local and holistic, then there is still a critical role that the mechanical model plays: the way we understand the world. Dieks yields that mechanical models in quantum physics or even classical theories (electromagnetism) provide the conceptual grip needed to understand the world better (2019, p.63). Dieks contrasts mathematical explanations with mechanical ones and concludes they are both valid. However, in quantum contexts, non-mechanistic explanations are more accurate and

truthful, although mechanistic models (given the decoherence caveat) can help us better understand quantum systems. This is nevertheless a good reason to suspect that mechanistic explanations cannot be fundamental in physics.

In the other co-editor's (B. Falkenburg) own contribution to the volume—one of the longest—there are some affirmative answers to questions (1), (2), (3), and (4), and some counterexamples to (1) and (2). B. Falkenburg distinguishes between the NMP and OMP but focuses on what levels of reality are used in mechanistic explanations. She is open to the idea that despite differences, the OMP ideas can be generalized to the practice of physics of the 21st century. She argues that Descartes and Kant used a multi-level mechanical model of the universe: in his astronomy lectures, Kant used levels of description based on the size of celestial objects and their life and becoming.

In NMP, we also need the so-called 'levels of description'—more precisely, level-based decomposition and recomposition of systems. Falkenburg argues that this is similar to Galileo's and Newton's proposed method of analysis and synthesis and is imported successfully into contemporary neuroscience and bi-

ology. Nevertheless, this multi-level mechanistic view declined sharply in the 20th century. Falkenburg takes clues from the literature on scientific practice of the 21st century and concludes that the mechanistic view can be generalized, despite the last century's rift, to the practice of the 21st century. Consequently, contrary to Dieks's suggestion, mechanisms can be generalized to fields, as classical fields admit a causal interpretation (Salmon, 1984, p.239; referred in Falkenburg and Schiemann, 2019, p.72). What is needed to generalize mechanisms in contemporary physics? In physics, it is often possible to express the causal processes underlying a mechanism in the precise terms of laws of physical dynamics (2019, p.73).

The familiar methods of top-down and bottom-up modeling in biology and neuroscience are for Falkenburg methods that generalize the views of Descartes, Newton, and Kant and illustrate the need for multi-level analysis. Their ideas are mirrored in the commitment to the existence of levels in Machamer, Darden and Craver (2000).

Mechanical models can be heuristically successful. Although false, they explain when they are used at a higher level. Falkenburg takes the kinetic theory of gases as a successful bottom-up model that explains a lot of thermodynamic phenomena without being true. The kinetic theory is a bottom-up model, albeit likely a false model. Unlike Dieks, Falkenburg considers locality constraints (discussed in detail in Bechtel and Glennan) too restrictive: the parts do not need to be localized (2019, p.81). We can delocalize parts of the system and generalize them to fields or non-localized entities (cf. Dieks). For Falkenburg, the part-whole relation does not need to be restricted to spatial or temporal localized domains: they can be generalized such that field interactions or superpositions of quantum subsystems qualify as part-whole relations.

Although we do not have a well-established concept of causality in physics and elsewhere, "it is possible to generalize the notion of a mechanism in an unconventional way" (2019, p.82). The parts are now idealized mathematical entities (she follows here a suggestion of Malisoff, a biochemist from the 1940s). We trivially replace real objects in the world with mathematical placeholders (point particles, infinite distances, etc.).

However, given the linguistic usage of the word 'mechanisms' in physics, Falkenburg argues that the

minimal mechanism description offered by Glennan or Salmon is inadequate in most cases. What about quantum systems? Here, the two short sections (§5.3.3.3 and §5.4.2) in which Falkenburg tries to address mechanisms in quantum physics are not satisfactorily developed. In her cursory note, Falkenburg suggests that a sum rule is all we need to connect parts of the system to the whole. Unlike Glennan and Kuhlmann or Dieks, Falkenburg's strategy to generalize mechanisms to quantum systems does not need decoherence. We only need a conventional 'sum' rule (2019, p.83). These last sections of the paper, although promissory, are underdeveloped (and see the reviewer's perspective on the overall missing points of this volume). The idea of a sum rule is not developed at all (the reviewer reminds the reader that there is a plurality of sum rules in quantum mechanics with somewhat different meanings).

In a different note, in addressing (1) and (2), M. Ghins tries to amend the standard view on mechanisms of Glennan and Machamer by revisiting the original mechanistic approach of Salmon and Dowe. The key concept used is that of causal laws identified by a formal criterion: they must contain the time derivative of a relevant quantity. In

this sense, causation is not an informal concept anymore but has a more elaborate definition. This, in return, may solve two problems of the mechanistic view: the regression 'difficulty' and the 'circularity problem'.

In short, Ghins is poised to solve the circularity and the bottom-up problem of NMP. If mechanisms are causal and designed to explain causation, we have a regress problem: the lower-level causation present in a mechanism needs a deeper mechanism to explain it, and so forth. Then we face the question: where do we 'bottom-out' mechanisms? Where do we stop with mechanistic explanations? (2019, p.99) Where the mechanisms end, the laws of nature must govern without being explained mechanistically. Ghins believes that Glennan and other mechanistic philosophers must admit that fundamental laws are explained neither causally nor mechanistically. Then, are mechanistic explanations genuinely fundamental?

Given these issues, Ghins is willing to give up the minimal mechanisms and return to the notion that mechanisms are elaborate descriptions of causal processes. In adding to Salmon's original idea of a mechanism, Dowe took the causal pro-

cess as a transmission of an invariant quantity. 'A causal interaction is an intersection of worldlines that involves the exchange of a conserved quantity' (2019, p.103). Salmon and Dowe avoided introducing the idea of laws into the discussion of mechanisms, but for Ghins, objects are stable worldlines that *need* laws of nature, more precisely causal laws of nature. A mechanism in Ghins is defined as a complex system of *nomic* causal interactions that explains the behavior of the system as changes of some relevant properties in time. Moreover, a law is causal only if it contains a time derivate (2019, p.106). In Ghins or Salmon's modified view, mechanisms depend on fundamental causal laws. The temporal variation of properties characterizes the behavior of mechanisms.

The reviewer notices a couple of problems here. If some laws are causal, the immediate question is whether non-causal laws play any explanatory role in this variant of the mechanistic philosophy. If there are non-causal laws, are they explanatorily idle with respect to the mechanism? Do they play any role? It would seem so. For example, symmetry conservation laws are not causal, but do they play a role in mechanistic philosophy? And if symmetry and other types of con-

servation laws should play a role in mechanisms, then this aspect needs some elaboration in Ghins' account.

The following contributions emphasize levels, as both the methodological and metaphysical elements of NMP. In the chapter by M. Buzzoni, levels are taken to be crucially perspectival, i.e., relative to a context, a point of view, and the goals of scientists. This new framework is helpful in clarifying various types of intertheoretical relationships. Levels are relative to a choice of theoretical perspective, and 'the question concerning the sameness of mechanism or level cannot be answered without a perspectival approach' (2019, p.118). Buzzoni lines up with those who emphasized the context-relativity of mechanisms (Potochnik, Pâslaru, Rueger, McGivern, etc.) and somehow against Craver, who believes more in an ontic concept of mechanistic explanation that is more-or-less context-independent. Buzzoni offers a few examples of relationships among knowledge claims in science: 'strong relations' (when two theories contribute together to understanding) and 'weak relations' (when two theories compete and overlap with respect to evidence) being the most relevant. Buzzoni does not show how these intertheoretical

relationships clearly affect the debates on the nature of mechanisms. One suggestion would be to connect more explicitly with the epistemic account of mechanistic explanations. The whole discussion on intertheoretical relations would benefit from an example drawn from the mechanistic literature. It is also unclear how different mechanistic explanations based on different levels and perspectives compete or complement each other.

The chapter by H. Lenk moves the discussion on mechanisms into a different conceptual landscape: the cognitive process of interpreting mechanisms, called here 'schematizations' and 'interpretative constructs.' They are mental/cognitive higher-level constructs resulting from interpretative processes at the representational level. Lenk aims to offer a meta-theoretical and methodological approach to NMP. This connects to the standard difference between mechanism schemas and mechanisms sketches in Machamer et al. (2000). The approach is inspired by both Kant's forms of judgment and Cassirer's levels of interpretation, and applies directly to how mechanisms are schemas interpreting causation. Lenk uses Woodward's (2013) attempt to limit the applicability of NMP and avoid the

limitless scope of mechanistic explanations. This does partially address the question (3) by explicitly stating distinctive levels and meta-levels in a cross-disciplinary sense. Lenk urges the New Mechanists to think about how lower-level mechanisms relate to those on a higher level and how these are interpretative schemas of causal processes.

The reviewer is not yet convinced that such a general idea of multi-level interpretation can relate directly to the link we need to draw between mechanisms in the special science and physics as suggested in (2) and (3). Lenk does not immediately bring in case studies of interpretation schema that could connect physics to special sciences.

J. Faye critically discusses the hierarchy and multi-level of reality needed in NMP. For pragmatic reasons, a multi-level reality worldview, *e.g.* what Oppenheim & Putnam (1958) suggested in the 1950s, can be justified. But is the metaphysical assumption of multilevel reality suspect? One problem with the vertical view of reality is the causal impotence of higher-order causes and the logical inconsistency of downward causation. Neither reductionism nor emergentism can solve some of the puzzles of the vertical view. As this vertical per-

spective is questioned, Faye revisits some alternatives. Most notably, the horizontal view of reality is one in which systems exist at the same level but have categorical properties and dispositions. Dispositions are relational properties of systems that 'cannot be merely grounded in the intrinsic properties of the categorical basis' (2019, p.177). A manifested disposition is an 'extrinsic property of the system brought into existence by its interaction with an environment E' (2019, p.179). Therefore, in Faye's horizontal perspective, the interaction with an environment is sufficient to include NMP *sans* the multi-level ontology. Faye does not immediately address how physics can use mechanistic models even in this horizontal framework. The only example—the flock of starlings—does not belong to physics directly so it is unclear how it illustrates NMP's generalization to physics (if any).

M. Kuhlmann introduces the idea of what's called 'econophysics', a putative discipline that would seek to import models from statistical physics into economics. For example, ferromagnets and financial markets act similarly. Kuhlmann restricts this analysis to a methodological pluralism embedded in an ontological reductionism. They both illustrate the macroscopic

behavior of a system based on the interaction of the micro-components. Many details of the microscopic components do not matter in both cases. Most critics of econophysics have argued that the similarities between an Ising model in physics and what happens with financial markets are insufficient to build models in economics based on models in condensed matter physics. Kuhlmann thinks it is more appropriate to employ mechanisms in both disciplines rather than mathematical models: the mechanistic account of explanation is premised on the 'interactive organization (between the parts of the mechanism) that does the explaining. And sometimes it is not all the details of the interactive organization that matter but just some structural details of it' (2019, p.195). The lack of a renormalization group or scale invariance in the financial market are such details that do not matter. What is then a structural detail? The suggestion is to move from causal explanation accounts to structural accounts and, hence, structural mechanisms. If one stays at the structural level, 'there is a common mechanism in diverse systems such as ferromagnets and financial markets' (2019, p.198). The promised structural notion of mechanism is unfortunately underdeveloped in this material. One sugges-

tion is that model building is more important for structural mechanisms than explanation.

Another problem is that Kuhlmann does not immediately relate structures in structural mechanisms to mathematical structures governing the behavior of the two systems. This would be an exciting add-on to his perspective to recognize the importance of mathematical similarities in the structures of the two systems. Last, we also have computational models in two disciplines that use similar numerical simulations to obtain predictions from one area Ising model to financial markets. In what sense are these models related to NMP? It is a question worth pursuing (see a possible connection in Humphreys, 2019).

The last contribution by V. Fano, P. Graziani, M. Tagliaferry, & G. Tarozzi addresses how to relate a physical system to a given computation. When we have an abstract model of a computation, how exactly can we implement it in physical systems? Fano et al. offer an alternative to Piccinini's (2015) view about realization (*aka* implementation) and adopt a more or less standard 'mapping' strategy between a Turing machine's states and a physical system's states. Fano et al. depart from Piccinini's standard

mechanistic approach and limit implementations by physical laws and presumably avoid some standard objections by Putnam and Kripke (2019, pp.217–218). As in the case of the previous contribution, it is not immediately clear whether Fano *et al.* actually generalize or employ any of NMP's assumptions in their approach.

The overall result of this volume is impressive, but the reviewer has several questions looming at the end. The reader can find some clear answers to questions (1), (2), and (4) as well as contributions that do not offer that much with respect to the title and the main aim of the book. For example, the theoretical framework of particle physics, quantum field theory (for example, Feynman diagrams, mechanisms in gauge theories, or Higgs mechanisms) is sporadically mentioned but never directly addressed. Of course, to address the question (3), it would be beneficial to integrate fundamental physics, probably even including attempts to discuss mechanisms in quantum gravity. It is also strange the spacetime theories are barely mentioned in this volume. Dieks' contribution relates immediately to Glennan's and Kuhlmann's papers (Kuhlmann and Glennan, 2014; Kuhlmann, 2017), but a more intri-

cate work would address the sum rule (cursorily discussed by Falkenburg) as well as semi-classical models in quantum mechanics where one can sense various ideas and forms of mechanistic reasoning—*ditto* about areas of modern physics such as cosmology, astrophysics or condensed matter. Finally, there is a feeling that chemistry is mainly ignored in this volume. Materials about mechanisms in fundamental physics (particle physics and perhaps quantum gravity), chemistry, or biochemistry would complete this outstanding volume.

Ioan Muntean

University of Illinois,
Urbana-Champaign

Falkenburg, B. and Schiemann, G., eds., 2019. *Mechanistic Explanations in Physics and Beyond*. Vol. 11, *European Studies in Philosophy of Science*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-10707-9.

Glennan, S., 2002. Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), pp.342–353. https://doi.org/10.1086/341857.

Glennan, S., 2017. *The New Mechanical Philosophy*. New York, NY: Oxford University Press.

Humphreys, P., 2019. Knowledge transfer across scientific disciplines. *Studies in History and Philosophy of Science Part A*, 77, pp.112–119. https://doi.org/10.1016/j.shpsa.2017.11.001.

Illari, P., 2013. Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, 78(S2), pp.237–255. https://doi.org/10.1007/s10670-013-9511-y.

Illari, P. and Glennan, S., eds., 2017. *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. London.

Illari, P.M. and Williamson, J., 2011. What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2(1), pp.119–135. https://doi.org/10.1007/s13194-011-0038-2.

Kuhlmann, M., 2017. Mechanisms in physics. In: P. Illari, S. Glennan, P. Illari and S. Glennan, eds. *The Routledge Handbook of Mechanisms and Mechanical Philosophy*. London, pp.283–296.

Kuhlmann, M. and Glennan, S., 2014. On the relation between quantum mechanical and neo-mechanistic ontologies and explanatory strategies. *European Journal for Philosophy of Science*, 4(3), pp.337–359. https://doi.org/10.1007/s13194-014-0088-3.

Machamer, P., Darden, L. and Craver, C.F., 2000. Thinking about mechanisms. *Philosophy of Science*, 67(1), pp.1–25. https://doi.org/10.1086/392759.

Oppenheim, P. and Putnam, H., 1958. Unity of Science as a Working Hypothesis. In: H. Feigl et al., eds. *Minnesota Studies in the Philosophy of Science*. Vol. 1. Minnesota.

Piccinini, G., 2015. *Physical Computation: A Mechanistic Account*. Oxford: Oxford University Press.

Salmon, W.C., 1984. Scientific explanation: Three basic conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, (2), pp.293–305. https://doi.org/10.1086/psaprocbienmeetp.1984.2.192510.

Salmon, W.C., 1989. Four decades of scientific explanation. In: P. Kitcher and W.C. Salmon, eds. *Scientific Explanation*, *Minnesota studies in the philosophy of science*, 13. Minneapolis: University of Minnesota Press, pp.3–219. Available at: <http://conservancy.umn.edu/handle/11299/185700> [visited on 4 October 2023].

Woodward, J., 2013. Mechanistic explanation: Its scope and limits. *Aristotelian Society: Supplementary Volume*, Supp(87), pp.39–65.

Wright, C., 2012. Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science*, 2(3), pp.375–394. https://doi.org/10.1007/s13194-012-0048-8.

How science tracks understanding

Kevin McCain, *Understanding: How Science Explains the World*. Cambridge University Press, Cambridge 2022, pp.xx + 122.

The aim of McCain's book is an ambitious one: to provide the readers with an understanding of science— i.e., of what science does, of how science works, of what science aims to achieve, and, more generally, of what makes science a successful epistemic endeavor.

The author states in chapter 1 that, at its core, science is about explanation. Somewhat more precisely: the scientific enterprise is an enterprise of formulating, evaluating, and testing explanations for empirical phenomena. Why do scientists care about explanation? What is gained once one has access to an adequate or successful explanation for some phenomenon? Adequate or successful explanations, the author notices, are means to many ends that we value. Some of these ends are practical; other are epistemic. On one hand, a world that we can explain adequately or successfully is a world in which we can act effectively, i.e., in which we can predict what will happen, contribute to make things happen, and prevent

things from happening (at least typically; the author explores multiple ways in which explanation and prediction or successful intervention might come apart in chapter 4). On the other hand, a world that we can explain adequately or successfully is a world that makes sense to us. It is a world that we understand. Adequate or successful explanations, thus, contribute to making the world an intelligible place. The author conceives of these two goals of explanation, the practical and the epistemic, as tightly intertwined. Given that we understand a phenomenon or subject matter, he claims, we will—at least typically—be effective in our practical interaction with it. Understanding a phenomenon gives us some sort of power over it: it turns it into something that we can handle and, to some extent, bend to our will.

But what exactly is an explanation? What does it mean to explain a phenomenon? What is the general structure of an explanation? What does an explanans tell us about an explanandum? These questions are notoriously difficult to answer, which is reflected in the somewhat intimidating variety of models of explanation that has been flourishing in the philosophical literature in the last decades. In chapter 2, Mc-Cain manages to offer a very broad

and yet instructive explication of the term 'explanation'—one that could be deployed as a sort of compass to orient oneself while navigating this variety of models. Explaining (prominently in science, but probably also in everyday life), the author suggests, is a matter of 'tracking dependency relations' (p. 23). Somewhat more precisely: an explanation of a phenomenon φ provides answers to why- and how-questions relevant to φ, and in doing so, it aims at showing what φ depends upon and what depends upon φ (e.g., causally, nomologically...).

The author points out in chapter 3, while zooming in on specific kinds of scientific explanation, that scientists are after scientific *knowledge* of explanations. As standardly conceived in epistemology, knowledge requires, among other things, the fulfillment of a truth-condition. We cannot know that p if p is false. If we know that p, then p is true. If this conception of knowledge holds, then, for us to know that q explains p, it must be that q *correctly* explains p. In other words: an explanation that is known is one that mirrors the facts, that depicts dependency relations actually holding in the real world. The author seems to align to this conception of knowl-edge when he writes: '[w]e can know that a[n] ... explanation is correct *when it is in fact correct* and we have sufficient evidence for believing that it is correct' (McCain, 2022, p.38, emphasis added). And yet, a couple of pages later, the author claims something different. He writes: 'we can know that a[n] ... explanation is correct by possessing evidence that makes the truth of that explanation beyond a reasonable doubt' (McCain, 2022, p.40). This second quote suggests the following: whether we have knowledge of an explanation or not is not a matter of how accurately the explanation mirrors the facts; this rather depends on the evidence that we have in support of the explanation, and probably also on our evidential standards—standards that are not carved in stone, but likely to change over time. These strike me as two quite different conceptions of (scientific) knowledge, that at least *prima facie*, are not easy to reconcile. If the second conception holds, what counts as scientific knowledge at a certain point in time might be overturned at a later point in time. We can have scientific knowledge of explanations that are extremely well-grounded in light of our evidential standards and yet fail to mirror the

facts. This is not possible under the first conception. If we know an explanation, then it must be true.

How could this tension be resolved? Maybe what McCain is telling us is that genuine knowledge of explanations, i.e., knowledge of explanations that are correct, functions at best as a regulative ideal in the scientific endeavor. In scientific practice we have no way of stepping out of our representational systems and comparing them to an independent reality. Thus, we have no infallible way to tell whether we really know, and whether the explanations that we formulated and deploy are indeed corresponding to the facts. The best we can do in practice is hold true or commit ourselves to those explanations that we reasonably judge as acceptable—i.e., to those that perform best in light of our evidential standards. Given that we have sufficient grounds to endorse an explanation, we have scientific knowledge of it. But scientific knowledge is not necessarily genuine knowledge; it can be directed to representational systems that, despite all evidence suggesting the contrary, do not fulfill a truth-requirement. The final chapter of the book, chapter 8, seems to provide at least some support to this reading. In the chapter, the author investigates the role of the inference to the best

explanation in the production of scientific knowledge and he makes it clear that the function of an inference to the best explanation is to help identify not truth, but what is reasonable to endorse in the given epistemic circumstances.

While this reading is certainly in line with the book's overall spirit, I am not sure it fully captures what McCain has in mind. Here and there throughout the book, one gets the impression that truth and genuine knowledge are more for the author than regulative ideals orienting the scientific endeavor. Somewhat more radical realist tendencies shine through the book's pages. Consider, for example, chapter 5: there, the author deals with the question of how explanations are evaluated and explores the role of empirical and theoretical virtues in such evaluation. At the end of the chapter, in what seems like a sort of 'optimistic meta-induction', he claims that in light of how successful we have been in the past by letting empirical and theoretical virtues orient our theory choice, it is reasonable for us to trust that 'such virtues are good indicators of the truth' (McCain, 2022, p.68). Of the truth, then—not of some weaker epistemic desideratum such as reasonable acceptability! In chapter 6, truth peeps out prominently again. While investigat-

ing the role that explanations play in fostering understanding, the author asks: 'Is truth important for understanding or scientific explanations?' (McCain, 2022, p.78). His answer does not leave much room for interpretation: 'Yes, truth is extremely important for both. While a false explanation might be such that if it were true it would provide understanding, genuine understanding requires accurate explanations' (ibidem). Even allegedly inaccurate scientific representations, such as idealized models, or so the author claims in chapter 7, provide us with understanding only insofar as they make us appreciate truths about dependency relations that we would otherwise overlook.

Genuine understanding then, according to the author, requires truth and is gained via explanations that mirror the facts. We need genuine knowledge of explanations in order to understand; a less demanding epistemic state such as scientific knowledge (as clarified above) won't do if it does not guarantee that truth has been reached. Now, McCain clearly does not take understanding to be something that only a final science will achieve. He takes understanding to be instantiated in real-life scientific practice. Our scientific understanding grows, he claims; we make advancements in understanding (see, e.g. McCain,

2022, p.68). But then, he must be endorsing some form of scientific realism. That is, he must believe that science not only targets truth, but has actually achieved it (at least to some extent); he must believe that science has managed to formulate explanations that are not only worthy of being endorsed in the given epistemic circumstances and in light of our evidential standards, but actually correct, i.e., corresponding to reality (at least approximately so).

I think McCain's excellent book would have gained in further clarity and depth if such a realist or optimistic stance towards science would have been not just presupposed and used as an unquestionable basis to build on, but rather put on the table, made explicit, articulated, and defended against alternatives. The book certainly succeeds in its aim: it provides the readers with *an* understanding of science; but as scientific realism is not the only available and viable stance, what is offered is just one way in which science could be understood.

Federica I. Malfatti
University of Innsbruck

McCain, K., 2022. *Understanding: How Science Explains the World*, *Understanding Life*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108997027.