Metaphysical "Models" and the Problematic Comparison with Science: Diagrams, Graphs, and Vectors between Communication and Experimentation

Abstract. An interesting idea in recent meta-metaphysical discussions proposes that metaphysics is importantly similar to science in the way both employ models in their investigations. Here I review this proposal in light of developments in the metaphysical literature on causation to argue that the putative similarity is more limited than suggested. Metaphysicians use a variety of representations in their work, but few of those representations qualify as models in the richer sense that applies to what scientists do: diagrams and graphs in metaphysics play a purely communicative role, whereas, in science, modeling is closely tied to experimentation. While this difference does not by itself delegitimize metaphysics, it does undermine the claim for legitimacy on the grounds of similarity with science—which is a problematic move to begin with.

Keywords. Scientific Modeling · Metaphysics of Causation · Diagrams · Communication · Experimentation · Metaphilosophy.

[* endnotes indicated by numbers in brackets]

1 Introduction

In a recently published paper, L. A. Paul (2012) takes up the idea, proposed a few years earlier by Peter Godfrey-Smith (2006), that metaphysics and science are importantly similar in the way both use models in their investigations. When faced with a phenomenon that is complex and little understood, scientists often resort to building *models* such as concrete replicas, mathematical equations, computer simulations, etc., in order to facilitate their understanding of the phenomenon of interest. The idea then is that (much of) metaphysics is done in the same way because metaphysicians use models to investigate their objects of study—even if, as both authors agree, metaphysicians differ from scientists in being generally unaware of this aspect of their own practice. This idea capitalizes on two important (and fashionable!) niches in contemporary philosophy, bringing together, on the one hand, discussions about scientific modeling in philosophy of science, with, on the other hand, the flourishing of metaphysics alongside metaphilosophical discussions concerning the value and viability of metaphysics. Yet, I suspect that the analogy is not warranted. Before I explain why (section 3), let us look more closely at the idea itself (this section) and then at developments in the metaphysics of causation (section 2) that will serve as a helpful illustration.

Paul (2012) starts out by taking issue with the criticism that metaphysics employs ill-suited methodology to address questions belonging to the domain of science. She suggests that not only do science and metaphysics study different things, but, curiously, they do so through very similar methodology, namely by constructing models. First, in her view, metaphysics and science do not have overlapping jurisdiction because science is actually built upon metaphysical foundations: "Metaphysics tries to tell us what laws, naturalness, properties, objects, persistence, and causal relations fundamentally *are*, in terms of natures" whereas "Science assumes that we have a pretheoretical grasp on these natures" and "tries to discover *which* entities there are or how these are exemplified" (Paul 2012: 6, italics original). Second, Paul holds, metaphysics and science coincide in their reliance on model-building in theorizing, with the main difference between the two being the objects to which each applies this methodology: scientists build "models of the world" and metaphysicians build "models of features of the world," but "scientific and metaphysical modeling are methodological peas in a pod" (Paul 2012: 9).

This idea that metaphysics relies on modeling is one that Paul draws directly from Godfrey-Smith (2006). Unlike Godfrey-Smith, however, Paul explicitly assumes the semantic view to be the norm among philosophers of science. She summarizes modeling within the semantic approach as consisting in developing *descriptions of mathematical structures* which, "when interpreted, are the models that are the theory" (Paul 2012: 10). From this perspective, she claims: "a metaphysical theory can be understood as a class of models, where the models are composed of logical, modal and other relations relating variables that represent n-adic properties, objects, and other entities" (Paul 2012: 12). At the same time, Paul distinguishes her favored "metaphysical theorizing about the world" from mere "conceptual analysis" (Paul 2012: 14), possibly suggesting that she is interested in models as models of phenomena rather than, in this technical sense, as models of theory; still, it is unclear whether she actually takes metaphysical models within "metaphysical theorizing about the world" to be models *of the world* or to be models *of metaphysical theories about the world* in the technical sense implied by the semantic view.

This is an unnecessary confusion. Recognizing this as a potential danger, Godfrey-Smith (2006) had already proposed to brush aside technical senses of "model" (such as that in model theory) and simply to utilize the term as it applies to scientific *practice*. This is the sense I employed in the beginning of the paper, a sense that encompasses scientists' construction of a variety of tools such as physical structures (e.g., the San Francisco Bay dam model), computer simulations (e.g., Schelling's segregation model), and mathematical equations (e.g., the Lotka-Volterra predator-prey model).[1] What models like these have in common is their utility for "surrogate reasoning" (Suarez 2004) and "indirect theoretical investigation of a real-world phenomenon" (Weisberg 2007: 4).[2] It is this sense of "models" that Godfrey-Smith has in mind when he suggests that metaphysics relies on modeling. The comparison between metaphysics and science is thus a very specific claim: "The relation I am here asserting between a scientific activity and a philosophical one is not merely one of analogy” (Godfrey-Smith 2006: 9); rather, “The products of systematic metaphysics can function as models, be useful as models, despite being surrounded by the ideology of direct theorizing” (Godfrey-Smith 2006: 10). Whether or not the semantic view assumed by Paul is a good philosophical account of scientific theorizing, it is just that, a philosophical perspective on science. The idea at hand, remember, is that of comparing metaphysics and science, not metaphysics and *philosophy of science*; accordingly, my evaluation (in section 3) of the claim that metaphysics and science alike use model-based methodologies will take into consideration this broader sense of “models” adopted by Godfrey-Smith rather than Paul's narrower technical sense. But first let us briefly review particular instances of prima facie promising cases of metaphysical modeling from the literature on causation.

2 Approaches to the Metaphysics of Causation

A popular approach to the study of causation consists of using so-called “neuron diagrams” with circles and arrows to represent causal systems. Although originally meant to represent causal relationships within neuron systems (hence the name), the diagrams are now widely used to represent a host of other ordinary and hypothetical causal scenarios, ranging from straightforward to increasingly complicated cases of early and late preemption, prevention and double prevention, etc. Paul (2012) very briefly mentions neuron diagrams as an example of metaphysical modeling, explaining that the diagrams “represent fictional neurons stimulating other neurons, in order to represent complex hypothetical patterns of counterfactual dependence and causation” (Paul 2012: 14). The original representational conventions for neuron diagrams are: shading circles to indicate that the "neuron" is active ("firing"); and lines connecting circles to indicate stimulatory or inhibitory relationships, depending on whether the line meets the next circle with an arrow or a black dot, respectively (Paul & Hall 2013: 9). Subsequent conventions include fat arrows for probabilistic connections (Paul & Hall 2013: 65, Figure 5), bold circles to indicate a "stubborn neuron" that only fires if there are two incoming signals (Paul & Hall: 90, Figure 14), and shading circles in black to indicate stronger outgoing signal and in checkerboard pattern to indicate the deflection of a weaker incoming signal (Paul & Hall 2013: 163, Figure 32), among others. With all of these representational variations, neuron diagrams can be offered in addition to some fictional story that fits the causal scenario of interest, but often a diagram is presented by itself as a sketch of a causal system, with no further fleshing out.

A distinct approach combines causal graphs with structural equations. This approach is presented by Christopher Hitchcock (2007) as a direct alternative to neuron diagrams, but it is important to note that it originally arises in a distinct literature that, instead, focuses on providing a formal account of causal reasoning (more on this later). For the sake of comparison, below I have reproduced two figures from Hitchcock (2007) that exemplify the two approaches. Causal graphs may, at first glance, look similar to neuron diagrams. Yet, an important innovation in this approach is the use of structural equations to specify causal relationships—in contrast with embedding information in the neuron diagrams themselves through additional stylistic conventions. This approach starts with variables representing "incompatible events, or states that an object or system can instantiate," and equations are then added to represent the relations between variables (Hitchcock 2007: 76-77).

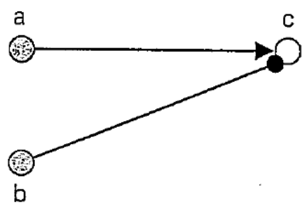

Figure 4.1
A typical neuron diagram. In the neural system depicted, neurons a and b both fire. a stimulates neuron c, while b inhibits c. c does not fire.

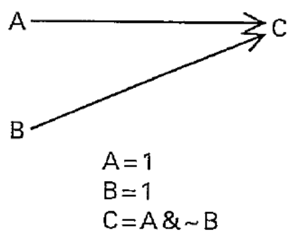

Figure 4.4
The neural system of figure 4.1 is represented by a causal graph and a system of structural equations.

The two approaches reviewed so far are both characterized by Jonathan Schaffer (2014) as forms of a "network model" of causation. Among the chief differences between the two, Schaffer indicates that while in the first approach the connections between neurons represent actual causation, in the second approach connections represent "lines of *possible* causal influence, and do not entail any *actual* causation between the actual values of their variables" (Schaffer 2014, emphasis added). This distinctive representational feature captures a difference concerning the goals each of the two approaches serves, a point that will be relevant for our discussion in section 3. For now, however, an important similarity between the two brings us to our third and final approach. Both neuron diagrams and causal graphs/structural equations assume a conterfactual theory of causation; from the competing perspective of realism about causal powers and dispositions, Stephen Mumford and Rani Lill Anjum (2011) propose that vectors provide more adequate representation. In simple one-dimensional cases, the vector approach starts from a blank "quality space" divided by a vertical line that represents the system's starting point and that demarcates the two sides representing two properties (or a property on one side and its negation on the other); horizontal arrows are then added to represent powers tending toward each of the two sides. For illustration, consider Figure 5 below (from Mumford & Anjum 2011: 61), where, if we take F and G to represent the temperature in a room (with F = cold and G = hot, although specific temperature values can be used as well), then arrows *a* through *f* represent distinct causes tending toward each outcome (such as *a* for the air conditioner being on, and *f* for a window being open and letting outside hot air in). As Mumford & Anjum propose, in cases similar to the one just described, powers can be added to represent the resultant power—determining, i.e., the actual temperature resulting from the dispositional-causal contribution of each factor (see Figure 6 from Mumford & Anjum 2011: 63). As for more complex cases, where powers tend to multiple relevant properties or states, a three-dimensional quality space can be used (see Figure 12, from Mumford & Anjum 2011: 70).

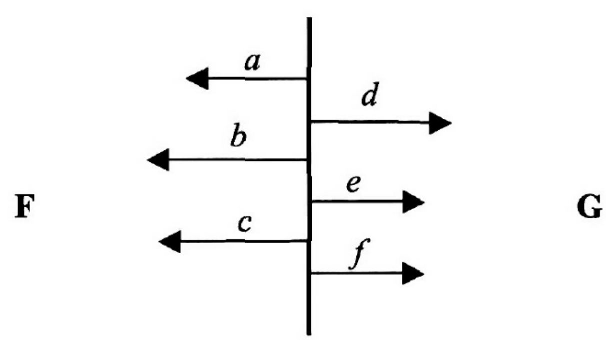

Figure 5: Multiple powers at work

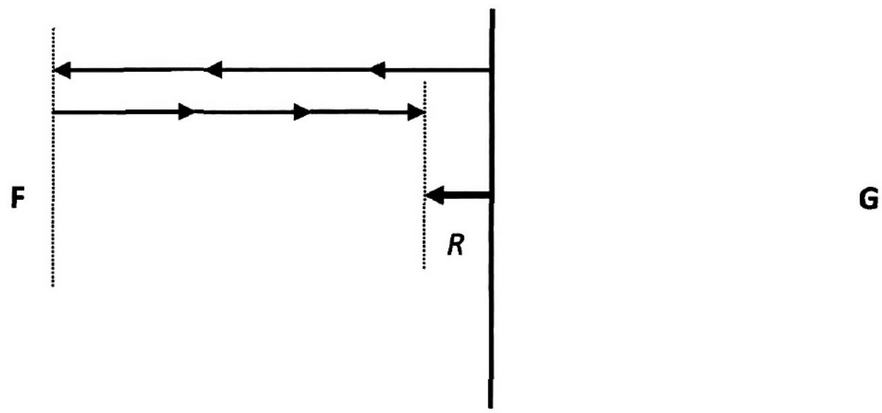

Figure 6: Vector addition for calculation of the resultant R

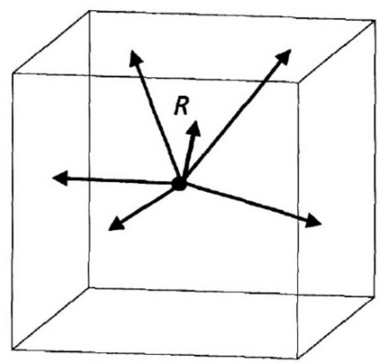

Figure 12: Powers within a three-dimensional quality space

Mumford and Anjum apply this vector approach to a number of cases that, they argue, are more easily accommodated from a dispositionalist perspective than within a counterfactual theory of causation, such as causation "of absence" and "by absence," and overdetermination. For them, vectors demonstrate the causal contribution of different powers, showing "how the extent of a power makes a difference" as well as "how countervailing powers make a difference," and applying not only to simple scenarios, but also to "multi-dimensional, probabilistic and nonlinear cases" (Mumford & Anjum 2011: 78).

3 Communication and Experimentation

Having briefly sampled a few of the representational devices used in contemporary metaphysics of causation, let us return to the question driving this paper. It is, again, the idea that metaphysics and science are importantly similar in the way both use models in their investigations. I think this idea poses a problematic comparison, for reasons which I now proceed to address.

Think, first, of how scientists use pictures and diagrams when presenting their research and findings to peers. Figures and graphs projected on a screen at a conference or published in a paper provide an insight into a scientist's work that might only cumbersomely be expressed in words.[3] But this is arguably a different use of representations than what takes place in the usual context of model-based research. Photographs and blueprints of the San Francisco Bay dam model, for example, arguably fulfill different purposes than what the actual model accomplishes when operating. This is the distinction between representations that play a communicative role and the "manipulative" or "experimental" use of models. Now consider my own use of figures in the previous section. The examples I provided of neuron diagrams, causal graphs with structural equations, and vectors served merely to illustrate three approaches to the metaphysics of causation. This falls within a communicative use of representations: I complemented verbally-encoded details with further information specified diagrammatically. The important question to ask is whether the way metaphysicians use those "models" is substantially different from my merely communicative use—whether, i.e., they use those as resources that can be manipulated and experimented upon for "indirect theoretical investigation" (Weisberg 2007: 4), as is characteristic of scientific models.

Metaphysical "models" play, for the most part, a merely communicative role. Starting with neuron diagrams: Paul and Hall affirm that they represent "a complex situation clearly and forcefully, allowing the reader to take in at a glance its central causal characteristics," "laying out most of the central issues (…) in a clear and immediately graspable way"(Paul & Hall 2013: 10). This sounds a lot like using images merely to represent an idea that could otherwise be communicated with words, even if with greater difficulty. Paul indeed seems to accept this point, as she affirms models to be used "most importantly to communicate large amounts of complex information in a simple way," further adding that "models are clear and efficient ways to explicate the properties and implications of a philosophical theory and they motivate and illustrate new ideas" (Paul 2012: 14).

Writing on creativity and heuristics in philosophy, Alan Hájek (forthcoming) suggests that neuron diagrams and causal graphs both afford the discovery of counterexamples through simple "trial and error." But, at least when it comes to neuron diagrams, there is room for doubt. First, because, as already pointed out, advocates of neuron diagrams mainly mean them to illustrate or exemplify a causal system. Second, however, because even when critics have offered particular neuron diagrams as counterexamples to a particular theory, advocates have reacted by adding novel conventions to accommodate those criticisms. This is indeed Hitchcock's complaint, namely that, because new representational conventions can always be introduced ad hoc, “neuron diagrams cannot delimit the logically possible structures a priori” (Hitchcock 2007: 81).[4]

The causal graph approach seems more directly related to the manipulationist (rather than merely communicative) role of models in science, which brings us to a question alluded to twice earlier in the paper. Causal graphs and structural equations come from an “interventionist” tradition that focuses on aiding causal reasoning by identifying possible empirical manipulations of variables and their likely effects. In the words of one pioneer, “causal modeling” with graphs and equations thus serves to “determine mathematically what effects an intervention might have, what measurements are appropriate for control of confounding, how to exploit measurements that lie on the causal pathways, how to trade one set of measurements for another, and how to estimate the probability that one event was the actual cause of another” (Pearl 2000: xiv). Paul and Hall acknowledge the utility of the causal graph approach, but caution their readers that the success of the approach depends “on the accuracy of the original representation by the model of the causal system” (Paul & Hall 2013: 18), which they take to be no small matter since in many cases “we only have access to part of the structure of events” (Paul & Hall 2013: 35). Still, the fact that the causal graphs and structural equations approach requires the careful prior specification of the causal structure of the case of interest is a feature that makes it more similar to scientific models than neuron diagrams are. This stems from the interventionist orientation of the advocates of this approach, who indeed intend this type of modeling to inform causal reasoning in scientific practice.

The third approach, using vectors to represent causation in a dispositionalist perspective, has not yet been subject to comparable scrutiny as the first two approaches; thus, I cannot indicate any counterexample obtained via direct manipulation of the vectors. Yet, it seems that the representational resources of vectors allow them to play more than a merely communicative role, as is demonstrated in Mumford and Anjum's use of additions. But this should come as no surprise, since the authors recognize that they are borrowing the heuristics from physics. To say that this type of metaphysical "model" is similar to scientific models, therefore, is misleading: this is indeed a scientific model put to a metaphysical use.

These differences between metaphysical "models" and scientific models and the communicative and manipulationist roles they can play are the main reason I think the comparison between modeling in metaphysics and in science is problematic.[5] But there are at least two further reasons to question that idea. First, an important insight from naturalistic approaches in the philosophy of science is that science arose through refinements of ordinary human cognitive abilities and practices.[6] As such, the use of pictorial representations to communicate ideas simply and forcefully is not exclusive of science, but is found in much of ordinary experience. The use of images solely to illustrate a metaphysical point or theory thus cannot be used to compare metaphysics to science any more than to daily experience. A more convincing comparison with science depends on metaphysical "models" fulfilling a manipulative role.[7] As we have seen, however, that occurs only to a limited extent, and perhaps exclusively when metaphysics is already intentionally paired up with science—when, e.g., it is intended to support interventionist causal reasoning (causal graphs) or when it directly borrows scientific tools (vectors).

A final reason to be suspicious of this comparison between modeling in metaphysics and in science concerns the idea of comparing the two at all. Remember that the idea at hand was proposed in the context of a meta-metaphysical discussion about the value and viability of metaphysics. While I understand the metaphysician's desire to support her discipline, it seems that attempting to affirm the legitimacy of metaphysics by showing ways in which it is similar to science implies that it has a secondary type of legitimacy, one that is derivative of the legitimacy of science. The fact that the aspect chosen for the comparison in this case—modeling—is dubious does not help. But comparisons in general seem to be the wrong foot to start off on, as they only further reinforce the impression of science's superiority over metaphysics.

Notes

1. Weisberg (2013) reviews some of these models and argues that the three categories I mentioned—concrete, mathematical and computational—encompass most of the modeling done in contemporary science.tational—encompass most of the modeling done in contemporary science.

2. Godfrey-Smith frames modeling in terms of an "indirect approach to representing complex or unknown processes in the real world" (Godfrey-Smith 2006: 7, emphasis added). I" (Godfrey-Smith 2006: 7, emphasis added). I think it is misleading to treat the model-target relationship as fundamentally representational, and I favor instead an artefactual view of models—but that is a question I address in another paper.

3. From a classical cognitivist perspective, Jill Larkin and Herbert Simon (1987) argue that diagrammatic representations have computational advantages over sentential representations, the former often making for easier recognition processes than the latter. But there are constraints, as Paul Thagard affirms in his book The Cognitive Science of Science: "External representations such as diagrams and equations on paper can also be useful in creative activities, as long as they interface with the internal mental representations that enable people to interact with the world" (Thagard 2012: 112).

4. As evidence that this complaint is warranted, note that, in addition to the new stylistic rules mentioned before, Paul and Hall themselves recognize that in one example for ate preemption

they changed conventions from interpreting the diagram as representing a causal system over time to representing it at a discrete time t (Paul & Hal 2013: 99, Figure 18).

5. Although my analysis has focused on causation, it seems quite plausible that the same conclusions apply to a number of other metaphysical topics I have not been able to address here. An interesting example concerns the representations used in the metaphysical literature on composition and on levels of organization, which also find their way in more speculative parts of scientific theorizing; my suspicion is that pictorial representations of part-whole relations and of hierarchical levels play only a communicative role rather than affording the sort of indirect experimentation and manipulation we see in scientific modeling.

6. Nancy Nersessian calls this the "continuum hypothesis," i.e., that there is a continuity between the cognitive resources and practices in ordinary and in scientific problem-solving (e.g., Nersessian 1992 and 2002).

7. And even then, the comparison might as well be with ordinary problem-solving, which, some argue, at times relies on a manipulationist use of physical representations as well (see previous note).

References

1. Godfrey-Smith, P. (2006) Theories and models in metaphysics. *Harvard Review of Philosophy*, 14(2006), 4–19.

2. Hájek, A. (forthcoming) "Heuristics for Philosophical Creativity", in *The Philosophy of Creativity*, ed. Elliot Samuel Paul and Scott Barry Kaufman, Oxford University Press.

3. Hitchcock, C. (2007) What's Wrong with Neuron Diagrams? in J. Campbell, M. O'Rourke, and D. Shier (eds.) *Causation and Explanation* (Cambridge MA: MIT Press), pp. 69-92.

4. Larkin, J. H., & Simon, H. (1987) Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* 11, 65-99.

5. Mumford, S., & Anjum, R. L. (2011) Spoils to the Vector: How to Model Causes if you are a Realist about Powers. *The Monist*, vol 94, no 1, pp 54–80.

6. Nersessian (2002) The cognitive basis of model-based reasoning in science. In Carruthers, P., Stich, S. & Siegal, M. (eds.) *The Cognitive Basis of Science*. 133–153. Cambridge University Press.

7. Nersessian (1992) How do scientists think? Capturing the dynamics of conceptual change in science. In Giere, R. N. (ed.) *Cognitive Models of Science*. University of Minnesota Press. Minneapolis, MN. 3–45.

8. Paul, L. A. (2012) Metaphysics as modeling: the handmaiden's tale. *Philosophical Studies* 160:1–29. DOI 10.1007/s11098-012-9906-7

9. Paul, L. A., & Hall, N. (2013) *Causation: A user's guide*. Oxford University Press.

10. Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

11. Schaffer, J. (2014) "The Metaphysics of Causation", The Stanford Encyclopedia of Philosophy (Summer 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2014/entries/causation-metaphysics/>.

12. Suarez, M. (2004) An inferential conception of scientific representation. *Philosophy of Science*, 71 (5). 767-779.

13. Thagard, P. (2012) *The cognitive science of science: explanation, discovery, and conceptual change*. MIT Press.

14. Weisberg, M. (2007) Who is a Modeler? *British Journal for Philosophy of Science*, 58, 207-233.

15. Weisberg, M. (2013) *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.