Philos Stud (2012) 160:1–29
DOI 10.1007/s11098-012-9906-7

Metaphysics as modeling: the handmaiden's tale

L. A. Paul

Published online: 27 April 2012

Abstract Critics of contemporary metaphysics argue that it attempts to do the hard work of science from the ease of the armchair. Physics, not metaphysics, tells us about the fundamental facts of the world, and empirical psychology is best placed to reveal the content of our concepts about the world. Exploring and understanding the world through metaphysical reflection is obsolete. In this paper, I will show why this critique of metaphysics fails, arguing that metaphysical methods used to make claims about the world are similar to scientific methods used to make claims about the world, but that the subjects of metaphysics are not the subjects of science. Those who argue that metaphysics uses a problematic methodology to make claims about subjects better covered by natural science get the situation exactly the wrong way around: metaphysics has a distinctive subject matter, not a distinctive methodology. The questions metaphysicians address are different from those of scientists, but the methods employed to develop and select theories are similar. In the first section of the paper, I will describe the sort of subject matter that metaphysics tends to engage with. In the second section of the paper, I will show how metaphysical theories are classes of models and discuss the roles of experience, common sense and thought experiments in the construction and evaluation of such models. Finally, in the last section I will discuss the way these methodological points help us to understand the metaphysical project. Getting the right account of the metaphysical method allows us to better understand the relationship between science and metaphysics, to explain why doing metaphysics successfully involves having a range of different theories (instead of consensus on a single theory), to understand the role of thought experiments involving fictional worlds, and to situate metaphysical realism in a scientifically realist context.

L. A. Paul (✉)
Department of Philosophy, University of North Carolina, 212 Caldwell Hall, Campus Box CB 3125, Chapel Hill, NC 27599-3125, USA
e-mail: lapaul@unc.edu

Keywords Metaphysics · Methodology · Science · Models · Inference to the best explanation · Intuitions · Common sense · Kant · Theories · Empirical equivalence · Simplicity · Theoretical virtues · Explanation

Why do contemporary metaphysics? Why rely on metaphysical reflection to develop theories about the world? Critics argue that contemporary metaphysics attempts to do the hard work of science from the ease of the armchair.[1] Physics, not metaphysics, tells us about the fundamental nature of the physical world, and empirical psychology is best placed to reveal the content of our concepts about the world. At best, metaphysics is a handmaiden to science.

In this paper, I will show why this sort of "naturalist" critique of metaphysics fails. Metaphysicians are not armchair scientists, nor are they handmaidens.[2] However, although the critique *does* fail, one can understand why some have embraced it. Over the last 30 years, as the influence of logical positivism and ordinary language philosophy waned, work in contemporary metaphysics slowly shifted from analyzing our concepts about the world to a focus on the world itself. To the uninitiated, modern metaphysics, with its use of a priori reasoning to draw conclusions about the nature of the world, can appear obsolete—it can appear to proceed as though we have learned nothing from Hume and Kant. It also can appear, to the uninitiated, as though metaphysicians really are trying to do the work of the scientist from the armchair: as though metaphysicians are trying to use a priori reasoning to study the parts of the world that scientists can or actually do study empirically.

Appearances here are misleading.

Part of the problem is that the shift has occurred so gradually that participants in the debates have not explicitly noted or acknowledged corresponding shifts away from a traditional rationalist methodology. To respond to the naturalist critique, I shall argue that metaphysical methods used to make claims about the world can be similar to scientific methods used to make claims about the world, but that the subjects of metaphysics are not the subjects of science.[3] The use of a priori reasoning does not derive from rationalist assumptions, but from the assumption that such reasoning involves the application of theoretical desiderata that have been proven to be of value by successful theorizing. Those who argue that metaphysics

[1] Many of those who label themselves "naturalists" make this sort of objection. Recently, James Ladyman and Don Ross, with David Spurrett and John Collie (2007). While the book often mischaracterizes metaphysics, it nevertheless raises important methodological questions that need to be addressed.

[2] My subtitle is ironic. For a somewhat ancient context see Atwood (1985), a staple of Women's Studies syllabuses in the 1980s and 1990s.

[3] Although I'll usually write as though the divide between the subjects of science and metaphysics were sharp, it is in reality fuzzy, and moreover, it can move with conceptual changes and scientific revolutions. Also, although my focus is on philosophical reasoning used to draw conclusions about the nature of the external world, there is still an important role for conceptual analysis in philosophy, in particular, when doing the stage-setting for the rest of the philosophical theory. I'll come back to both of these points throughout the ms.

uses a problematic methodology to make claims about subjects better covered by natural science get the situation exactly the wrong way around: metaphysics has a distinctive subject matter, not a distinctive methodology. The questions metaphysicians address are different from those of scientists, but the methods employed to develop and select theories are often relevantly similar. And just as with natural and social-scientific theorizing, as long as we construct and evaluate our theories appropriately, we are justified in inferring conclusions using inference to the best explanation.[4]

In the first section of the paper, I'll describe some of the distinctive sorts of subject matter that metaphysics tends to engage with, and I'll discuss a few particular lines of enquiry that exemplify the metaphysical stance. In the second section of the paper, I'll show how to understand metaphysical theories as models or classes of models. I'll then discuss the roles of experience, common sense, thought experiments and inference to the best metaphysical explanation in the construction and evaluation of such models. Finally, in the last section, I'll discuss the way these methodological points give us a framework for understanding the viability of contemporary metaphysical projects. Developing an account of the metaphysical method allows us to better understand the relationship between science and metaphysics, to explain why doing metaphysics successfully involves having a range of different theories (instead of consensus on a single theory), to understand the role of thought experiments involving fictional worlds, and to situate metaphysical realism in a scientifically realist context.

The view I develop below describes my preferred way of approaching metaphysical questions. It is intended to show how metaphysical enquiry makes sense from a naturalistic point of view, and to provide a context for the contemporary shift away from conceptual analysis and the idea that metaphysics is solely an exercise in a priori reasoning. And although I won't argue the case here, I think it is fairly obvious that many of my methodological claims about metaphysics extend to other philosophical discussions, such as those in philosophy of mind, ethics and metaethics. What I say below is applicable, *modulo* changes of subject, to these types of philosophical project as well: much philosophical theorizing is productively understood as involving modeling and inference to the best philosophical explanation. But I am not claiming that there are no other ways to

[4] This is the place to make two more important caveats. (i) Although, for simplicity, I will speak as though I am making claims about all ways of doing metaphysics, I am not. Instead, I am giving an account of how one can legitimately engage in reasoning to substantive metaphysical conclusions that approximately captures the methodology of many contemporary metaphysical projects. There may be lots of metaphysical projects that don't employ the methodology described here: such projects will need their own defense. In addition, I won't have much to say about conceptual analysis or about the role of philosophy of language in doing metaphysical theorizing about the world, although I think it can be quite useful, especially when fashioning or refining models. (ii) I develop my view using the semantic approach to scientific theorizing, which is the dominant approach for philosophers of science, but my view doesn't require the semantic approach. While the precise details would need to be fleshed out somewhat differently on a different account of scientific theorizing, my central claim, that most metaphysical claims about the world rely on inference to the best explanation, along with most of my subsidiary claims, would remain the same.

do good metaphysics, and by extension I am not claiming that there are no other ways to do good philosophy: there are, but I leave them to others to explore.

1 Subject

My first thesis is that the questions and problems addressed by metaphysicians are often distinct from those addressed by scientists. It is confused to think that metaphysicians simply study the same parts of the world that scientists study. The truth is far more complicated and subtle.

Of course, topics of metaphysical inquiry vary widely. I won't pretend to cover all the subjects of interest or to be making claims about all the metaphysical questions one can ask. Instead, I will develop an account of a central way in which metaphysics differs from that of science by considering some paradigmatic examples. One of the more popular and enduring metaphysical projects in metaphysics concerns the search for fundamental and general truths about the world. The metaphysician engaging in such a search wants to determine the natures of the world, especially the fundamental natures of the world, as part of her enquiry into the nature of things. The metaphysician also looks to discover systematic, general truths; in the context of the investigation of fundamentality, she hopes to discover systematic, general truths concerning fundamental facts. Such fundamental facts often involve natures, types of composition, and primitive distinctions.

An example of such a project is metaphysical category theory, a theory that involves the attempt to determine the ontological categories, or what Peter van Inwagen (forthcoming) describes as the specification of real divisions among things. This project involves the search for the fundamental natures of the world. The fundamental divisions among things concern the most basic ontological categories, or the most basic divisions between the real natures of things: what van Inwagen terms the "high" classes. Less fundamental divisions concern derivative categories, involving divisions made between natures of less fundamental things. In this way, the project limns the categorical structure of the world. Another project that fits our schema tries to determine whether and how less fundamental constituents of the world are built from their metaphysically prior constituents. This project attempts to discover systematic, general truths about how the world is constructed. Such projects can be described as *ontological* projects.

How do these ontological projects interact with the scientific project of determining the fundamental entities of the world? How do they fit with the scientific approach of taking entities like fields or leptons, or properties of charge, spin and mass, to be *physically* fundamental? And how do they fit with the scientific approach of taking macrolevel objects to be constructed from fields or particles? In physics, for example, entities such as fields are often taken to be physically fundamental, and in chemistry, atoms and molecules are taken to be constructed in some sense from attractive forces and smaller particles. How is this sort of project related to the metaphysician's account of the fundamental constituents and construction of the world?

Despite initial appearances to the contrary, the different approaches are not in tension, for the ontological account involves features of the world that are metaphysically prior to those of the scientific account. The ontological account describes the metaphysically prior categories and constituents of the physically fundamental entities, and in this sense describes features of the world that are more fundamental than those of natural science. For example, when a fundamental physics takes fields to be the most physically fundamental entities of the world, an ontology will take the theory of the world a step further, by describing the ontological categories of the fundamental constituents of these fields. The substrate-attribute theorist will take fields and particles to be substances with properties, so will take both fields and particles to be substratum-attribute constructions from members of the fundamental categories of substance and attribute, while the bundle theorist will take fields to be bundles of property intensities and location properties, which are members of the fundamental category of property. And so on—the ontological account will take the properties postulated by the scientific theory, such as properties of charge, spin and mass, and ask if they are fundamentally universals or fundamentally tropes (or fundamentally something else).

We see the same relationship between different theories of the construction of entities. An ontological theory of parts and wholes (a mereology) of physical objects describes more fundamental and more general constructional principles than physics or chemistry does, for it gives general principles that govern all the physical objects with parts, including microparticles, atoms and molecules. For example, chemistry may tell us that the physical structure of a polycarbonate is causally created by arranging elements a certain way, and that its physical parts consist of these arrangements of elements and the attractive forces between them. Mereology contributes the additional claim that the molecule *just is* (say), the mereological fusion of its arranged parts (the elements and the attractive forces). The polycarbonate molecule is created by this mereological fusion, but not in a *causal* sense. Rather, it is created in the compositional or ontological sense: it exists when the parts arranged in the right way exist. So the metaphysics tells us what it *is* to be a sum or physical object composed of these structured arrangements of parts, and thus tells us how the physical object is metaphysically constructed (composed) from its parts. In contrast, chemistry tells us what some of the parts and the arrangements of the parts are for different kinds of molecules, and it also tells us how to causally manipulate the world in order to bring such arrangements into existence. Chemistry tells us how molecules are physically or causally constructed from attractive forces and smaller particles.

The examples bring out how metaphysics involves questions about features of the world that are prior to those described by science. One distinction we can invoke to help with the intuitive difference between science and metaphysics is that often metaphysics is concerned to identify the real natures of the world while science is concerned to discover the range of *instances* of these natures. In the examples above, the metaphysics gives the general and systematic story of what the categories are and what the method of composition is. The story from science is more specific: it provides a list or a catalogue of entities that are the actual members of these categories, and a specific story about how causing instances of different kinds and

individuals involve different arrangements of physical properties and objects. Roughly, metaphysics gives us categories (natures), while science gives us members of categories (instances of natures). The categories are metaphysically prior to the members of the categories.

Once we start to look, we see this structure everywhere. Metaphysics investigates, for example, the natures of laws, naturalness, causation, persistence, and properties. Science assumes that we have a pretheoretical grasp on these natures, and then investigates the instances of these natures: it tells us which laws obtain in the natural world, which natural kinds there are and how they are ordered, which other properties and relations are actually instantiated, which objects persist, and what causes what (and how). Metaphysics tries to tell us what laws, naturalness, properties, objects, persistence, and causal relations fundamentally *are*, in terms of natures, and science tries to discover *which* entities there are or how these natures are exemplified.[5]

The fact that the subject matter of metaphysics can be ontologically prior to the subject matter of science is reflected in the fact that many concepts of metaphysics are conceptually prior to the concepts of science. For example, in order to tell us about members of the categories, science relies pretheoretically on basic metaphysical concepts of categories. There is no way to make sense of the central concepts of classical field theory or quantum chromodynamics without using a concept of property. There is no way to make sense of the concept of mechanism in organic chemistry without using a concept of causation. There is no way to make sense of the central concepts deployed by biological representations of the citric acid cycle without using a concept of persistence. In such cases, we start with the metaphysical concepts as the conditions under which we understand the scientific concepts.

This brings out the deeper reason for why it is a mistake to think of metaphysics as a handmaiden to science. It is a mistake to think that one should first study science and then use it as the guide to one's metaphysical conclusions. Science does try to construct an overall picture of the world. But scientific theorizing usually *uncritically assumes* the very organizing principles and deep general truths that metaphysics is concerned to prescriptively develop and understand. For example, as the examples above show, scientists will observe macrolevel entities bringing about effects, and draw conclusions from these effects relying upon pretheoretical metaphysical assumptions about persistence and causation. It is simply naïve to think that the only appropriate role for metaphysics is to take on these uncritical assumptions about the nature of persistence and causation or to merely draw out the implications of such naïve, uncritically assumed scientific suppositions.

Instead, the metaphysician should be concerned to prescriptively develop and understand the prior, deep, and general truths about the fundamental natures of the world used to organize and understand the rest of the world. Science still acts as a constraint upon metaphysics—the metaphysician should want her theory of the whole world to be consistent with accepted scientific theories of the world—but it

[5] As Eric Schliesser argues, this arrangement may be, at least in part, the result of what he calls "Newton's Challenge" to systematic metaphysics. See Schliesser (2011).

should not preemptively define the role or concepts of metaphysics. That would give us an understanding of reality that is exactly the wrong way around.

There is a complication: each metaphysical theory is drawn partly from experience, but it also organizes experience, and thus its success cannot be understood independently from its interpretation. Since the world is, in a sense, seen through the lens of the metaphysical theory, any metaphysical theory is confirmable only indirectly, and as part of a whole package. Choice of a "best" metaphysical theory is based on how well the theory makes sense of the entire world, from ordinary experience to scientific experience and theory, but the way the theory interprets experience is not independent of the theory itself. Alternative theories describe and understand the world in different ways, and the different theories can be seen as different accounts of the way fundamental natures form the structure of the entire world, and so should be understood as different "lenses" through which one views the world.

This point is in some sense a Kantian one, although I reject the idea that our fundamental, systematizing concepts of the nature and structure of the world are mere a priori constructions or must be somehow prior to (or bound up in a strictly Kantian way with) worldly experience. Instead, I am assuming here that many metaphysical concepts are directly caused by experience and knowledge of the natures and structures of the external world.

Recognizing that we use metaphysical principles to structure and understand the world does not imperil metaphysical realism, for the realist holds that the success of our theories employing these structuring principles and concepts is evidence that they accurately represent the objective natures and categorical structure of the world. The realist holds that our successful concepts and theories track, at least approximately, the categories, natures and structures of the world—at least once we have done our prescriptive metaphysical theorizing. The realist may hold—defeasibly—that features of the world, including fundamental categorical and other features, are among the causes of the formation of our concepts, and to the extent that we can use our concepts to successfully grasp and represent the world, she can infer that the nature and structure that our concepts represent is the true nature and structure of the world.[6]

It's worth explicitly mentioning one especially important sort of general truth that metaphysicians are interested in uncovering: modal truth. One way in which modality is the subject of metaphysics is when it involves an attempt to determine what things in a world must be like in order for a certain metaphysical concept to apply to them. The claim, then, involves necessity, since it is about what needs to be the case in any possible world in which the concept applies. For example, David Lewis claims that "... conceptual analysis is required to reveal what it is that all the actual and possible varieties of causation have in common."[7] Often, the project is reductionist in the sense that it explicates one concept in terms of another, more basic concept.

[6] Psychologists sometimes take certain dispositions to make causal or other judgments to be innate. Such a view is not inconsistent with my view, since we might start with these innate dispositions and then develop our ideas or concepts based on experience. There might also be evolutionary success arguments supporting the inference to the veridicality of some of our innate inclinations.

[7] Lewis (2004).

A different sort of project, the sort of project I'm concerned with in this paper, is not primarily conceptual analysis: the goal is to develop a theory of the world itself, not a theory of the content of our concepts. This sort of project may take its claims about the world to be contingently true in the actual world and worlds relevantly similar to our world; for example, one might hold that it is a contingent fact about the world that its fundamental categories are categories of substance and attribute, or that composition is only contingently restricted. On the other hand, it may make stronger modal claims. For example, one might develop a theory of composition as necessarily unrestricted (necessarily, any collection of entities fuses to make a whole) or a theory making claims about the natures or essences of things, such as a theory defending the view that objects have their origins essentially. The modal commitments of each of these theories are part of the overall package of the view: for example, the claim that it is contingent that the world has fundamental categories of substance and attribute is weaker and correspondingly more plausible than the claim that it is necessary, and the view that composition is only contingently restricted is easier to defend than the claim that the restriction of composition is necessary. A theory that makes stronger modal claims will need to compensate by exhibiting other virtues. I will return to this point in Sects. 2.2 and 2.4 below.

Now for a few caveats. Although I have been arguing as though there were a clear line between metaphysics and science, the reality is much blurrier. The vague boundary between science and metaphysics provides a rich opportunity for philosophers of science, who can draw on their scientific expertise to evaluate the plausibility (or otherwise) of metaphysical theories that bump up against the domain of the empirical. One place where the subject matter of metaphysics and science can connect is where metaphysicians study matters that, while currently ignored by science, are in principle empirically observable. Perhaps there are entities with characteristics that are empirically inaccessible right now, but could be made accessible with radically different instrumental techniques—techniques that we cannot even imagine at present. We might now class these entities under the purview of metaphysics, but when science is complete, they will be (correctly) classed as the subject matter of natural science.

A related sort of connection concerns indirect ties between metaphysical theories and scientific theories with established empirical consequences. For example, many scientific theories with empirical content require concepts of naturalness, persistence, causation and identity. Scientific theories also make assumptions about the indiscernibility of identicals, about counterfactual dependence, probability, and sometimes even about the existence of physically undetectable mental properties. Even if different scientific theories are empirically equivalent, differing only in implicitly adopting different metaphysical concepts or making different implicit assumptions about categories or other features of the world, metaphysicians need to have a sense of what sorts of assumptions and tradeoffs are being made.

Another connection between metaphysics and science is more direct. There are places, especially in the highly theoretical realm of fundamental physics, where sophisticated scientific theses will cut directly against metaphysical ones, especially if the scientific thesis in question fits with established theory or enjoys indirect

empirical support. Here there is danger for the scientifically naïve metaphysician, and metaphysically informed work in general philosophy of science and philosophy of physics plays an important role in the refinement and development of metaphysical theories that involve such assumptions. Examples of places where knowledge of contemporary scientific theories bumps up against the metaphysics include theories of space and time, mereological theories based on certain assumptions about the nature of spacetime, intrinsicness, and discussions of what sort of causation is physically possible.[8]

Finally, there can be significant revisions to metaphysical theories in times of scientific revolution, especially because such times usually involve major conceptual revolutions. Revision of central metaphysical and scientific concepts can have ripple effects through a wide range of theories of the world. I will return to this last point in the closing section.

2 Method

My second thesis is that we can understand the methods employed by metaphysicians to be very similar, *modulo* the change in subject matter, to the methods employed by scientists. Both fields are interested in discovering truths about entities or features of the world that are sometimes observable, but are often unobservable, indirectly confirmable, and abstract. (Such entities include objects, properties, relations, or what-have-you.) Both fields rely on a priori reasoning in addition to a posteriori reasoning.

We can theorize about the world using models, that is, by constructing representations of the world, and metaphysical theorizing is no exception. Scientific theorizing is often understood in terms of the construction of models of the world, and scientific theories about the nature of features of the world may be understood as models of features of the world. Metaphysical theories about the nature of features of the world may also be understood as models of features of the world. Both fields can be understood as relying on modeling to develop and defend theories, and both use a priori reasoning to infer to the best explanation and to choose between empirical equivalents. On this view, the most important differences between the scientific method and the metaphysical method derive merely from the difference in subject matter and the resultant difference in the role they give to ordinary experience.

2.1 Scientific modeling[9]

To be clearer about how scientific and metaphysical modeling are methodological peas in a pod, let's look closely at the "semantic view," a widely accepted account

[8] See Maudlin (2007), Albert (1996), Loewer (2004), Ney (forthcoming), Monton (2012), Healey (1991) and Paul (2012) for relevant discussion.

[9] I'm especially indebted here to conversation with Peter Godfrey Smith and Michael Strevens.

of scientific theorizing. On this approach, theories of the world are sets or families of models, where the models are usually mathematical relations or some other sort of suitably structured entity. Stephen Downes describes the view thus: "[s]cientific theories consist of families of (mathematical) models including empirical models and sets of hypotheses stating the connections between the empirical models and empirical systems. Empirical models are models that specifically purport to have relations to an empirical system."[10] To the extent that the theories are successful as representations of features of the world, their models are isomorphic in some relevant sense to those features of the world. The models, as things that represent a bit of the world, provide a kind of semantics for a theory. As Godfrey-Smith puts it, ... "[a] model [on the semantic approach], basically, is a set of objects (and relations between them) that functions as an *interpreting structure* for a set of sentences. A model is used as something for a set of sentences [the theory] to be *true of*."[11]

In summary: one develops a description of a theory, in part by developing a description of some mathematical structures; these mathematical structures, when interpreted, are the models that are the theory. The theory is true just in case it has a model that is isomorphic to the relevant features of the world, including (but not limited to) the structures that can be described in experimental and measurement reports.

State space models function as exemplars of the semantic approach. Take a sample state space model to be an abstract entity. This entity has a structure that can be described by a set of equations. We can take values of state variables to pick out the coordinates of a point in the abstract state space, and use initial values plus the equations to determine an evolution of the values of the state variables over a stretch of time. If our abstract entity is relevantly isomorphic to an actual dynamical system, it represents that system. In this way, we can use a mathematical model to represent the evolution of an actual dynamical system, and we can use a class of such models to describe and represent a range of similar dynamical systems. This is the way a theory can be understood to be a family or class of interpreted mathematical models of some feature of the world.

For another example, consider Lloyd's (1988) account of the way to understand the semantic theoretic treatment of population growth in ecological theory. A logistic equation, a derivative describing a pattern of population growth ($dx/dt = rx(1-x/k)$, is a statement or mathematical expression that (following Downes 1992) can be thought of as a description of mathematical relations between mathematical objects. The mathematical objects represent things like population density (x) and rate of increase (r). The mathematical objects plus their relations are the model, the mathematical equations and other statements we use to describe and refer to these objects and relations describe the model, and the isomorphisms between the model and growth rates of various populations allow the growth rates to satisfy the model.

[10] Downes (1992, p.143).

[11] Godfrey-Smith (2006b, p. 727).

Now, there is an extra step here that one might be inclined to omit. As the semantic account is presented, we move from a description, which picks out an abstract structure, which in turn represents the world via isomorphism. I am not sure that we always need the abstract structure to actually exist. Why not move directly from a description of an abstract structure to the world itself? In other words, take the description to be a description that purports to limn the structure of a feature or portion of the world itself, rather than limning an abstract structure that in turn must be isomorphic to the world.

In some cases, such a simplification may be welcome. But not always. One reason is because complete models of complex systems don't simply fall from the sky like manna from heaven. For this reason, models that rely on ideal systems, fictional populations and other sorts of "imagined concrete things" are a very important part of actual science, especially in sciences other than physics. Much scientific modeling is an attempt to "gain understanding of a complex real-world system via an understanding of [a] simpler, hypothetical system that resembles it in relevant respects."[12] Godfrey-Smith distinguishes between abstract modeling, i.e., modeling that leaves out features of the world it represents, but where what the model describes is taken to be literally true, from ideal modeling, i.e., modeling that incorporates literally false elements and thus what the model describes is not wholly literally true. Both sorts of modeling are ubiquitous in scientific theory-building and evaluation.[13] As Nersessian (1999) points out, the use of models shades off from using them as agents of conceptual change or as idealizations to using them as a precise representation of complex systems.

So we cannot in general dispense with the abstract structure. Building a class of models and defending it as a theory often involves the construction and evaluation of idealized and abstract models, and for this we need separate abstract entities that are not identical to the bits of the world described by the theory. We need to use abstract structures when we model using idealization and abstraction in order to distinguish clearly between model and world. We can include abstract and ideal models in the class of models that is the theory as long as we specify the ways in which they represent and do not represent, or we can simply use these models to generate a final, different class that excludes the hypothetical models.

Once we have our classes of models, how do we decide between competing theories, each of which purport to represent the same feature of the world? Assuming that the competing scientific theories are approximately empirically equivalent, or at least empirically acceptable, selection of a theory over its competitors is determined by a mix of desiderata, including its overall explanatory value, which is evaluated in part by its simplicity, elegance, and fit with already accepted theories, intuitions and assumptions.[14] This is one place where a priori

[12] Godfrey-Smith (2006b, p. 726). According to Godfrey-Smith, some of this sort of modeling is not captured by the semantic approach. I am not taking a stand on this controversial issue.

[13] Godfrey-Smith (2006b). Also see Martin Thompson-Jones, "Models and the Semantic View" and "Idealization and Abstraction: A Framework", both unpublished, for excellent discussions of these topics.

[14] For example, Strevens (2008) discusses the way that unification accounts of explanation include simplicity as a desideratum for the model.

reasoning and inference to the best explanation play an important role. After a theory is selected from the mix as providing the best explanation, if one is a scientific realist, its class of models is supposed to give us the truth about the nature and structure of certain features of the world: i.e., we accept the theory as a representation of these features of the world.

Now, in many cases, this picture is somewhat idealized, since in science even *approximate* empirical equivalence is very rare. Nevertheless, the importance of the role of inference to the best explanation as a means to grasp scientific truths about the nature of the world is well confirmed, at least when we look past the context of fundamental physics to wider scientific contexts, for example, to the context of evolutionary biology.[15]

2.2 Metaphysical modeling

In metaphysics as with science, we can also understand theories of the world as built by developing models. Such theories may include the construction of models involving idealization, abstraction and hypothetical systems, as well as more precise and complete models of complex features of the world. There are obvious parallels, for example, between the use of thought experiments as hypothetical, ideal and abstract models of features of the world, and the development of logics as precise models of features of the world. Once the models are developed, just as in science, theories are compared with respect to the elegance, simplicity and explanatory virtues of their models, and theories are chosen over their competitors using inference to the best explanation. On this way of understanding theorizing about the world, much of metaphysics, like much of science, proceeds by model-building.

While I am not wedded to the semantic approach, I shall formulate my view of metaphysics as modeling in terms of it, since it is the dominant approach to scientific modeling. Other ways of thinking of metaphysical theorizing as modeling may do the job just as well.

Employing the semantic approach in the service of metaphysics, 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. (We might also think of conceptual analyses in terms of modeling concepts, but I'll set this aside for now.) For example, a theory of composition can be thought of as a class of models of composition relations between parts of composite objects. Consider a theory of the composition relation such that some *x*s compose a *y* if and only if the activity of the *x*s constitutes a life. The models we can take to be the theory are structures of abstract objects that represent activity-constituting objects standing in necessitation relations to abstract objects that represent composites or wholes of the activity-constituting objects. The theory is a class of (suitably abstract) models, where these models are isomorphic to various instances of the activity-constituting relations between parts and wholes.

[15] Even those who wish to minimize the role of IBE in evolutionary biology can grant the importance of theoretical desiderata like simplicity in reasoning to conclusions about the nature of the world (see, for example, Sober 2008).

For another example, consider a simple counterfactual theory of the causal relation that holds that *c* is a cause of *e* if and only if, had *c* occurred, *e* would not have occurred. Models for the theory are structures that represent events standing in relations of counterfactual dependence (descriptions of these models are descriptions of these structures). If these structures are isomorphic to causal relations in the actual world, the theory represents actual causal relations and gives an account of the nature of actual causation. Of course, as aficionados of the metaphysics of causation know, a truly successful counterfactual analysis of causation will be more complex than this simple model, but the same techniques will apply.

What about idealization and abstraction? Just as in science, idealization and abstraction play a huge role in metaphysical model-building and theory evaluation, in particular, in the use of thought experiments. Many thought experiments are basically models of hypothetical situations: the possible worlds are simply the abstract structures functioning as models in the semantic approach. Possibilia can function as abstractions, that is, as representations of a part of an actual structure but with irrelevant detail removed, and as idealizations, that is, as representations of fictional situations.[16]

Let's start with thought experiments functioning as abstract models. One can consider a thought experiment involving a possible world where the relation, property or object we want to investigate exists, but where irrelevant detail does not—it is abstracted away. Such a possible world can be used to focus attention on the entities that the theory needs to be able to characterize, that is, to focus on the relevant feature of the world. The possible world serves as a model of a portion of a more complex actual structure. Consider a theory of composition as relation R where the ontologist examines possible worlds that have far fewer objects than the actual world, say, only three actual objects. Such worlds are useful because they can isolate the instantiation of relation R and make it salient. Those working on the ontology of causation may consider worlds where there are few extraneous events in order to ensure that the relation they take to be causation is highlighted, or they might look at such worlds simply to illustrate the abstract structure being suggested, since in the real world identifying a case of the existence of the relation may be complex or controversial. For a case of the latter, consider how one who takes causation to be a form of counterfactual dependence might draw on abstract models to clarify the intuitive power of the view and thus motivate the central premise that counterfactual dependence is sufficient for causation. For example, Lewis (1973, 1986) asks us to consider an abstract model in which Socrates' death depends on his drinking the hemlock. Assuming this sequence of events actually happened, we abstract away from the complex details of the actual world to identify and isolate a counterfactual dependence relation between the events of Socrates' death and his drinking the hemlock. Since counterfactual dependence is extrinsically sensitive, it can be extremely useful to move to an abstraction where extrinsic detail is eliminated in order to allow the dependence relation to appear. If we agree that in the abstract model, the instance of counterfactual dependence is an instance of

[16] We also sometimes use thought experiments as tests for models when doing conceptual analysis, and more generally when drawing out what we think we know about some feature of the world.

causation, this lends credence to the philosophical hypothesis that such dependence is sufficient for causation. The thought is that the abstract structure is relevantly isomorphic to the real-world case, even if extrinsic features of the real-world case muck up the real-world dependence facts.

Metaphysicians also rely on idealization.[17] For example, Lewis and others working on causation make use of what they call "neuron diagrams," diagrams that represent fictional neurons stimulating other neurons, in order to represent complex hypothetical patterns of counterfactual dependence and causation. Other projects reply heavily on assessments of idealized possibilia: for example, a possible world with only three simples and seven objects is an idealized world that can be used to explain and discuss the basic idea behind unrestricted composition.[18] (Note the parallel here between this sort of thought experiment and thought experiments involving Maxwell's Demon in thermodynamics.) Sometimes, idealizing models have properties or structure that is supposed to exist in our world as well as in the world of the model, and other times idealizing models may only have properties or structure that could exist if the fictional elements of the model existed and the philosophical theory it concerned were true. For an example of the latter, when Peter van Inwagen considers a hypothetical situation where the sugar in the tea that Alice drinks becomes a part of her, there is idealization: Alice and her sugary tea are fictional, but the description of the composition that Alice, as an organism, exhibits would apply literally if Alice had actually existed.[19] The idea here is that the ideal model represents some of the actual or possible structure of the actual world: for van Inwagen, in both the Alice-world and the actual world, in some sense, *x*s that constitute a life compose an object *y*. Other examples of abstraction and idealization in the philosophical literature abound.

To the untutored, such examples can seem like pointless and unnecessarily complicated flights of fancy. But as the use of abstraction and idealization in actual scientific practice reminds us, they play an essential and eminently respectable role in theory-building. Models are used to focus on salient entities or structures (especially in the case of abstraction), to show how to apply untested theories to possible cases, and most importantly to communicate large amounts of complex information in a simple way. The models are clear and efficient ways to explicate the properties and implications of a philosophical theory and they motivate and illustrate new ideas. Thus they function as important agents in conceptual development and change in addition to functioning as part of a finished theory.

Once we have a developed account of metaphysics in terms of model-building, we can distinguish metaphysical theorizing about the world from the more traditional metaphysical project of conceptual analysis. This is important. It is important in part because it gives us a clear way to understand and evaluate the different sorts of modal claims made in metaphysics.

Recall that the metaphysician who is only developing an account of what our concepts express is only developing an account of what, necessarily, the world must

[17] The line between abstraction and idealization is often a blurry one.

[18] Van Inwagen (2002).

[19] Van Inwagen (1990, p. 94).

be like in order for our concepts to apply. The metaphysical project that develops an account of the external world is different from such a conceptual analysis, and hence its modal claims must be understood differently. Claims that certain features of the external world are necessary are different from claims about what is necessary for our concepts to apply in the world and must be defended differently. One way to defend claims that certain features of the world are necessary situates them in a theory as a whole, and argues that such claims are plausible in light of their contribution to an overall best theory. Another defense involves holding that we have defeasible evidence in favor of the relevant modal fact: perhaps we could argue that we have special epistemic access to the natures of persons and minds, or perhaps science has shown us that water is essentially H_2O, or that ordinary experience cannot be explained unless the world has a certain nature, etc. These defenses are not merely defenses of what our concepts express.

Finally, we must make a clear distinction between the different types of projects in order to be clear about what a theory predicts, and hence what counts as a counterexample to the theory. Note first that a metaphysically interesting model of the nature of the world does not need to rely on claims involving necessity. For example, one might claim (i) that causation is the transfer of conserved quantities, but hold that this is a contingent truth or that composition is restricted in the actual world, but only contingently so.[20] On the other hand, one might claim (ii) that causation is necessarily a relation of counterfactual dependence between events, or that universal composition is necessary. And finally, one might be undertaking (iii) a conceptual analysis of the concept *causation* and hold that the content this concept expresses is one of counterfactual dependence between events, or analyze our concept *composition* such that it only applies to an unrestricted relation. These are different kinds of claims by different kinds of theories, so they make different predictions and are subject to different types of counterexamples. A counterexample to the causal theory of (i) requires that we find an instance of causation in the actual world that is not a transfer of a conserved quantity (for example, a case of causation by omission). A counterexample to the causal theory of (ii) requires that we find some metaphysically possible world with an instance of causation between events that does not exhibit counterfactual dependence (for example, an instance of late preemption[21]). A counterexample to the causal theory of (iii) merely requires that we can coherently conceive of an application of the concept of causation that does not involve events standing in the right relations of counterfactual dependence.

The different types of counterexamples set different standards for thought experiments. Thought experiments involving wizards and magic are totally irrelevant when considering theories of type (i), for here we are interested in a certain sort of physical possibility. But they are totally relevant when considering theories of type (iii), for here we only need conceptual possibility. And when considering theories of type (ii), we are interested in metaphysical possibility, so

[20] For the claim that composition is contingent, see Cameron (2007).

[21] See chapter three of Hall and Paul's *Causation: A User's Guide* (2013) for a discussion and explication of preemption.

thought experiments describing possibilia that have wizards and magic might be relevant, depending on how they are designed. Clarity about this issue is essential.[22]

2.3 The role of experience

When constructing metaphysical models, ordinary experience has a privileged role in the sense that it provides a baseline or a starting point. Sometimes the metaphysician develops a model that takes a part of ordinary experience as an irrefutable given. But what is more common is for a metaphysician to privilege ordinary experience in the sense of relying on it as an initial, but *defeasible*, guide to the nature of the world. Such a metaphysician starts with the defeasible assumption that the relevant feature of the world is as it seems to us, given ordinary experience.

The privileging of ordinary experience in metaphysics connects with the generality of the subject of metaphysics, with the fact that metaphysics usually focuses on structural features like ontological categories while science usually focuses on instances of the categories, with the role of empirical facts in the evaluation of metaphysical theories, and with the way that metaphysical models are indirectly confirmed.

Why is such privileging legitimate? First, because of the distinctive subject matter of metaphysics. The metaphysical projects we are exploring are concerned, in the first instance, to discover *general* truths or features of the world, that is, truths that are supposed to hold across all levels of experience, from the macroscopic level to the microscopic level. Hence, an examination of the ordinary macroscopic objects and properties of ordinary experience is often a useful starting point. If objects are constructed from substrates and attributes, then this is true for partridges as well as for particles. If causation is a matter of counterfactual dependence between events, then this is true for wars as well as for ion exchanges. And so on.

This means that the metaphysician is perfectly justified in focusing on the ordinary objects and properties of experience rather than the extraordinary objects and properties of science, at least in the first instance. One's observational experience of causation is no more informative when it occurs as the result of observing tracks made in a cloud chamber than when it occurs by observing tracks made on a white carpet. One's observational experience of what properties are is no more informative when one derives it from examining the colors of a slice of material observed with a microscope as when one derives it from examining the colors of the varieties of ice cream at Thomas Sweet's. Use of an electron microscope or an infrared spectrometer will not give us any more empirical information about the nature of composition than we can derive from everyday experience. The metaphysician is, here, assuming that the extraordinary micro- and macro-level objects of the sciences are covered by the same metaphysical regularities as the ordinary middle-sized objects of experience. If, as an enquiry is developed, it turns out that ordinary middle-sized objects and their properties cannot serve as exemplars in this way, then, of course, the metaphysician must refine her thesis accordingly.[23]

[22] This is something that contemporary metaphysicians could be a bit more careful about.

[23] It is also important to note that science relies on ordinary experience as well, to the extent that it relies on observation and experiment—including observation of effects of unobserved entities. Scientists still

A different reason for thinking that the metaphysician is perfectly justified in focusing on the ordinary to the exclusion of the extraordinary derives from the fact, noted above, about how the entities of metaphysics structure the entities of the rest of the world. Recall that scientific reasoning must *assume*, in some inexplicit sense, metaphysical facts about the structure of the world in order to proceed. Science has an enormous amount of empirical content, but it is able to have this content in part because of its pre-theoretical metaphysical assumptions, which allow it, at least initially, to interpret observational evidence more or less at face value. And, as I already noted, the ontological priority of the metaphysical is often reflected in a parallel conceptual priority: once we assume we know what causation, properties and objects are, we can pick out instances of causation, recognize kinds of objects, and detect patterns of property instantiation across time. Then we can develop theories about the electrons, organic molecules, and energy transfers of science. Since our original epistemic access to this worldly structure comes from ordinary experience, it is perfectly reasonable to start with experience and develop our metaphysics from there, for we look through the same metaphysical lens when we use our unaided eyes to detect macrolevel objects as we do when we use instruments to detect microlevel objects.

Of course, in every case, after drawing on experience to develop a theory, in evaluating it we need to look back at the natural science just in case our ordinary experience of the world conflicts with what our best natural science says about the world. If it does conflict, then often the assumptions based on ordinary experience should be rejected. This connects to one of the ways in which metaphysics and science interact, and brings out one of the ways in which metaphysics must take ordinary experience to be defeasible.

Another, less well-recognized way in which ordinary experience is defeasible derives from the fact that empirical work in psychology and linguistics suggests that perceptual experience and natural languages are in an important sense constructed via the human cognitive and linguistic apparatus. Experience and language have a variety of contingent features due to historical environmental contingencies and contingencies of our physical cognitive and linguistic capabilities. Metaphysicians need to be sensitive to this in order to distinguish between features of our experience and language that reflect the nature of the world and features of our experience and language that merely reflect the way that we respond to the world. Normally, only the former should be used as a clue to worldly structure and natures. The metaphysician must not be naïve about her interpretation of perceptual and linguistic experience. I and Alvin Goldman have argued elsewhere (Goldman 2007; Paul 2010a, b) that this is a point that needs more attention.

The defeasible role of ordinary experience in metaphysics contrasts sharply with the importance of observational evidence for scientific theories. This is because the features of the world described by metaphysics are not manipulable or testable the way the features of the world described by science are. There isn't the faintest

Footnote 23 continued
rely on *experience* to tell them about the theoretical or unobserved, even if the source of such experience requires heavy interpretation.

glimmer of an idea of what sort of instrument (and much less of an idea of how to build one) we could use to detect the presence of numbers, or the presence of composition, or of necessity, or the category of properties (e.g., there is no direct empirical test of whether a property is an instance of a universal or is a trope).

That this is the case is no surprise, for to create an instrument to detect such facts requires that these facts cause a distinctive sensory experience detectable by an instrument recording their presence. But we don't have this sort of sensory experience. The only sort of sensory experience we have is consistent, at least up to a point, with many different ways of modeling the metaphysical structure of the world. And thus, instruments such as cloud chambers, microscopes and spectrometers give us information about particular instances of causation or composition or properties, and about particular ways parts may be arranged or about new kinds of properties, but not about what composition is or what properties are.

For this reason, the ability of a metaphysical model to accommodate ordinary experience at most contributes to a kind of indirect confirmation of the theory. Ordinary experience does constrain metaphysics, but only because experience is taken to be one defeasible clue among many to the fundamental and general facts about the world. And as I've been emphasizing, ordinary experience isn't the only clue, since metaphysics ultimately engages in a kind of back-and-forth interaction with science. The metaphysical realist's theory of the fundamental natures of the world is indirectly confirmed by its success as a theory that fits with ordinary experience and by how well it fits with other well-accepted theories, including empirically confirmed scientific theories. But "fit" is not just one-way, for the metaphysician may challenge the pre-theoretic metaphysical assumptions that led to the scientific conclusions, and may argue that ordinary experience is to be re-interpreted in a way consistent with her preferred metaphysical model, or may argue that ordinary experience should be understood as being related to the manifest image, while metaphysics aims to capture reality.

The point about the indirect confirmation of metaphysical theories is familiar and general, even if it is not always explicitly recognized. For example, the realist thesis that there are objectively natural properties or primitive distinctions between qualitative kinds is not confirmable via direct observation, but receives indirect support from the fact that ordinary and scientific judgments, along with successful theories of laws, counterfactuals, and reference rely on naturalness. The realist thesis that there exist sets or their functional equivalents is not confirmable via direct observation, rather, it is indirectly confirmed to the extent that mathematics and physics require sets or their functional equivalents. The realist thesis that pluralities of parts compose to make a whole is not confirmable via direct observation, rather, it is indirectly confirmed in part by our ordinary experience of a world that includes persons and other sums.

2.4 Use of a priori reasoning

Metaphysicians rely heavily on a priori reasoning to develop theories of the world. Has this reliance on the a priori been discredited by the success of empirical science?

No. The use of a priori reasoning has not been discredited by the success of science—quite the opposite. For scientific theorizing itself, even empirically-based science, relies on a priori reasoning involving simplicity, elegance and explanatory strength. Such considerations play an important role in the development of successful scientific theories, and the use of the a priori in metaphysics is similar to the use of the a priori in science. In neither case is a priori reasoning used to investigate directly observable or testable features of the world. And at important points in theory development, each sort of theorizing relies on a priori reasoning based on the evaluation of theoretical virtues such as simplicity, strength, elegance and the like. Finally, in both sorts of theorizing, one thing that can justify the use of a priori reasoning is that they employ inference to the best explanation based on the idea that theories that maximize simplicity, strength, elegance, and other theoretical virtues are more likely to be true. The a priori elements of the method used by metaphysicians are often just part of the standard arsenal of tools employed by any theorist of the unobservable, the indirectly confirmable, and the abstract.

The main difference between modeling in science and in metaphysics lies in *how much* of the theory involves the unobservable, the indirectly confirmable, and the abstract; and hence in *how many* different, competing models may maximize the theoretical virtues while doing an adequate job of saving the phenomena. For practical reasons, scientific theories are more constrained than metaphysical ones, for, given the subject matter of most scientific theories, it is harder to construct an empirically adequate theory. The practical difference between scientific theorizing and metaphysical theorizing means that metaphysicians engaging in modeling may have many more acceptable theories to choose from. But this does not undermine the legitimacy of a priori theorizing.

Now that we see how modeling can be part of metaphysical theory construction, we can also see more clearly how the use of the a priori to make claims about the nature of the external world can be justified. Recall that the semantic approach takes a theory to be a class of abstract models (in the case of scientific theories, often mathematical relations) that are satisfied by the parts of the world the theory is about. If the theory is successful (true), the class of models is isomorphic in the intended sense to the parts of the world it represents.

As a theory or class of models is developed, and before it can be taken to represent the world, it is tested for empirical adequacy. In experimental science, much weight is given to the success or failure of a theory's empirical predictions, especially if those predictions involve new or surprising results. If a theory predicts that y will occur if x occurs, and when x occurs, y in fact occurs, this result contributes to the confirmation of the theory. Setting aside issues about the nature and extent of such confirmation, it is clear that experiments of this form play an important role in the acceptance or rejection of many scientific theories.

In metaphysics, empirical adequacy is more complicated. Metaphysical theories often involve hypotheses about fundamental and general ways of structuring the world, and thus are only indirectly confirmed by their overall success in making sense of ordinary and scientific experience and language. Metaphysical theories tend to be of features of the world that do not reveal their nature through physical manipulation, and the indirect nature of their confirmation means that candidate

theories can be hard to knock off the list. That said, metaphysical theories still generate predictions, and we can evaluate them. Such predictions involve further use of thought experiments.

For example, one way to test a theory involves considering the actual world or perhaps close possible worlds involving fictional, yet physically possible situations, to see how well the theory does in giving an overall explanation of the nature of such a world. Another way to test a theory is to look for possible worlds that seem to contradict the theory. If a theory makes claims that extend past the actual world, if, for example, it holds that, necessarily, composition of an object y occurs *only* when there exist xs that constitute a life, then that theory predicts that no possible world includes ys without also including xs that constitute a life. To test the theory, we consider whether there are, or are not, such possible worlds. So some sorts of thought experiments function as abstract or ideal models in the way I described in §2.2, while others function as tests for theories.

Already, simply by allowing thought experiments to function as tests for philosophical predictions, we have a significant role for the a priori. But we can't always rely on this sort of reasoning to give us access to the truth. The trouble is that it can be difficult to ascertain the truth about which worlds are metaphysically possible. While we seem to be able to draw on inferences based on possibilia that are very similar to our world to determine how well the theory explains features of the actual world, many considerations involving more exotic possibilia are highly speculative. Examples involving modal claims about singletons and equivalence relations, for example, might be easy to assess. But once we move past these relatively conservative sorts of speculations, things get more controversial. I might argue that there exist worlds where there are computers and chairs composed from simples, and van Inwagen can argue that I am mistaken. At this point, we are at a stalemate. We have conflicting empirically adequate theories of composition.

If one rejects, as I do, the existence of some special faculty that will allow us to reliably evaluate counterfactuals concerning substantive metaphysical suppositions about distant possible worlds, we seem to be stuck with an unlovely proliferation of empirically adequate theories with no way to choose between them. And the problem extends to many metaphysical debates. It seems that, for any position in a metaphysical debate, any bizarre, fantastical view, as long as it can preserve appearances, is a viable candidate.[24]

Well, things are not as bad as all that. First, as I noted above, many interesting and important metaphysical theories are concerned with the actual world and its near relatives, not with all possible worlds, nor even with all metaphysically possible worlds. But even if we wish to defend more controversial claims involving metaphysical possibility and metaphysical necessity, we are still in very good shape. It is true that we will always have many candidate theories, and that few theories will be eliminated using empirical techniques. But what's so bad about this? It is the fate of philosophy to have many too many options. As Russell put it, "while

[24] This is too negative: I think there is some consensus about which descriptions are descriptions of possible worlds and which ones don't refer (or refer to impossible worlds). But there is enough debate over crucial cases to make the worry stick.

diminishing our feeling of certainty as to what things are, [philosophy] greatly increases our knowledge as to what they may be; it removes the somewhat arrogant dogmatism of those who have never traveled into the region of liberating doubt...".[25]

Only those who wish that philosophy were more like science or mathematics should feel aggrieved. As Russell points out, it is of significant philosophical value to develop theories about what there is, even if we cannot prove that a theory is true or confirm it with a wide range of empirical results. That's just not what most of philosophy, especially metaphysics, involves.

Moreover, even for those with a secret yearning to give up philosophy and become scientists or mathematicians, it is simply not true that just any empirically adequate metaphysical theory is a viable theory. Metaphysical theories, like other theories, compete with respect to their theoretical qualities as well as their empirical adequacy. We use theoretical desiderata as guides to truth in metaphysics just as we use such desiderata as guides to truth in science, since the method is fundamentally the same even when the subject matter is different. In both cases, an acceptable theory must maximize theoretical virtues while being empirically adequate. And this winnows down the field.

The theoretical desiderata we use to choose a theory include simplicity, explanatory power, fertility, elegance, etc., and are guides to overall explanatory power and support inference to the truth of theory. A scientific realist should take such desiderata to be truth-conducive, since it is hard to see how such desiderata can lead us to truth if they are merely or even mainly pragmatic virtues. If such theoretical desiderata are truth conducive in science, they are also truth conducive in metaphysics (and in mathematics, and in other areas). The main point I want to make here is that if the method can lead us to closer to the truth in science, it can lead us closer to the truth in metaphysics.

Now, relying on theoretical desiderata is not infallible: we can make mistakes. But it is perfectly rational to rely on theoretical desiderata to prefer a theory over its empirically adequate competitors, on the grounds that in an important sense it is more likely to be true, and one can do this even while being a fallibilist. In science, the amount of empirical data usually helps to restrain the size of the class of empirically adequate competitors enough to allow the application of theoretical desiderata to winnow down the field to one or maybe just a few theories. The remaining contenders exhibit some mix of theoretical virtues as well as empirical adequacy, and one can either prefer one mix of virtues to another or remain undecided.

One might argue that scientific theories are different from metaphysical theories because scientific theories involve empirical evidence, while metaphysical theories do not (apart from ordinary experience and apart from evidence that is derivative in the sense that it is drawn directly from science). The suggestion is that simplicity and other theoretical desiderata are truth conducive for empirically based theories, but not for metaphysical theories, and so the success of science does indeed confirm the use of theoretical desiderata, but only because science has empirical, confirmable features.

[25] Russell (1988), p. 157.

But while it is true that the empirical, confirmable features of scientific theories have allowed us to confirm the value of theoretical desiderata for theorizing, if such features are truth conducive in the case of science, they should be truth conducive more generally. That is, if simplicity and other theoretical desiderata are truth conducive in scientific theorizing, they are truth conducive in metaphysical theorizing. This is a central part of my thesis: if we accept inference to the best explanation in ordinary reasoning and in scientific theorizing, we should accept it in metaphysical theorizing.[26]

In ontology, because of the large size of the class of empirically adequate competitors, it is rare to have the application of theoretical desiderata winnow down the field to a single theory. There are usually a number of remaining competitors, each of which exhibit some combination of theoretical virtues combined with varying ways of accommodating the basic characteristics that are supposed to compose the empirical data. A bonus of this situation, not to be underestimated, is the value of epistemic diversity or disagreement: having different acceptable theories in competition with each other can contribute to the depth and quality of our overall ontological account of the world.

Because theories of the ontological contain very little observable content, for any ontological entity we consider, we may find ourselves with a host of empirically acceptable theories about it, and with little or no means of developing additional empirical techniques to eliminate candidates. Hence, the choice between ontological theories is determined largely by theoretical desiderata, given that each theory is empirically adequate. As I noted, this narrows the field. But it rarely narrows it to one "best" option. While choosing one theory over its competitors based on a particular balancing of theoretical virtues is perfectly rational, we will almost always have a range of alternative theories to choose from: this is because there are many different ways of balancing respective theoretical virtues and hence many different ways of arriving at reflective equilibrium.

When trying to achieve such equilibrium, tradeoffs between maximizing theoretical virtues (e.g., simplicity and elegance) and preserving commonsense beliefs are common, but tradeoffs that sacrifice central commonsense tenets only rarely convince. For example, in the debate over whether mereological composition occurs, nihilists argue that for any simples y, y can be arranged x-wise, but there is no sum or composite that is x. Thus, there may be simples arranged Finbarr-wise, but there is no sum of these simples that is Finbarr. The view entails that there are no persons, cats, rocks or stars. Understood in terms of models, nihilists are defending a model of our ordinary beliefs, concepts and language that maximizes simplicity and ontological parsimony at the expense of our commonsense interpretations of what we mean when, e.g., we say that "Finbarr ate the strawberry." The nihilist reinterprets this claim to mean something like *Some simples arranged Finbarr-wise ate*[27] *the simples arranged strawberry-wise*. Their defense of their model depends on valuing ontological parsimony over commonsense interpretations of ordinary language and other desiderata. Since the nihilist

[26] For a different view, see Ladyman's contribution to this volume.

[27] Actually they'll need something else here, since simples don't eat, exactly.

solution also entails solutions to related problems involving composite entities (e.g., there is no problem of material constitution since there are no entities to stand in the material constitution relation) they may also claim that their view has great explanatory value.[28]

Theories developed in opposition to nihilism will stress the value of minimizing such radical reinterpretations of our commonsense language and beliefs, and will sacrifice a measure of ontological parsimony to do so. (Different theories sacrifice parsimony to different degrees.) Note that all the theories, including that of the nihilist, are empirically adequate to the extent that they allow us to go on making assertions in ordinary language and having ordinary beliefs: it's just that those ordinary assertions and beliefs, according to some models, need revisionary accounts of what is needed for them to be true. Our ordinary concepts, beliefs and language are part of the empirical data to the extent that we need to preserve their assertibility if we wish to preserve empirical adequacy, and we accept that they capture what we *ordinarily* say, think, etc. In some cases we might reject even the data itself (e.g., deny assertibility) if a model proved to be of great value otherwise, but convincing cases are as rare in metaphysics as they are in science.[29]

Thus, I view metaphysical theorizing as a process—one that is sometimes messy—involving model-building, interpretation, evaluation and trade-offs, and I do not regard the absence of a firm consensus about many metaphysical topics to be a particularly surprising or even an especially negative feature of the state of the field. The messiness is just what we'd expect if we actually look at how the creative process of model-building works (as Nersessian (1999) shows). Moreover, while I agree with Williamson (2007) that philosophy has moved past its linguistic and conceptual turns, and I also agree that philosophical reasoning needs to give a central role to the evaluation of counterfactuals and other types of thought experiments, this role for counterfactuals is most productively embedded in a larger picture of philosophy (here, metaphysics) as modeling. Counterfactuals are exceedingly useful tools for model-building, especially for abstract and fictional model-building, and for developing and assessing the predictions of a theory, and it is modeling and inference to the best explanation that form the core of the metaphysical method I am defending.[30] So I make no claim for any special faculty

[28] This is not to say that the nihilists don't create new problems for themselves. For example, they need to work out the complex semantic details of their new interpretation of the reference of our language.

[29] In ethics, we use things like "killing babies is bad" as empirical data. In some cases, this can be rejected, but only at great cost to the attractiveness of a theory. There might be some cases where a reinterpretation of the empirical data might result in reinterpretation of general claims about empirical regularities without violating empirical adequacy. In this sort of case the empirical data are *refined*, not rejected. The view that killing a single baby is not bad if it saves a billion babies can be seen as a refinement of the general claim that killing babies is bad just as the view that in certain circumstances effects can occur simultaneously with their causes is a refinement of the general claim that causes precede their effects.

[30] This allows for the flexibility in designing possible worlds that Ichikawa and Jarvis (2009) points out we need in philosophical theorizing, and should also accommodate imagining, supposing, and conceiving in all its glorious variations. See Ichikawa and Jarvis (2009). I am also sympathetic to Jenkins' (2008) arguments against Williamson: I agree with her view that modal epistemology is not reducible to counterfactual epistemology. Jenkins (2008).

to assess the truth-value of counterfactuals. We don't need to make ourselves handmaidens to linguistics any more than we need to make ourselves handmaidens to science.

3 The contribution of metaphysics

Now we can fit the pieces together to see how doing metaphysics can be understood as a contribution to our overall understanding of the world.

As I argued above, one line of inquiry in metaphysics holds that we start with ordinary experience, defeasibly holding that it gives us a small amount of a posteriori knowledge. For example, we perceive certain basic properties such as cohesiveness and continuity, and we may use that information to construct a theory of how objects persist.[31] When a metaphysician wants to give an account of the nature of macrolevel objects and properties, she rightly draws on our ordinary experience of macrolevel objects and properties and our ordinary language describing such objects and properties to begin to develop a systematic and general theory of the world.

This a posteriori starting point of this sort of metaphysical inquiry shows how Hume undermined himself with his empiricist skepticism. Recall Hume's skeptical despair in the *Treatise*, where he begins to.

> "fancy [him]self in the most deplorable condition imaginable, inviron'd with the deepest darkness, and utterly depriv'd of the use of every member and faculty. Most fortunately it happens, that since reason is incapable of dispelling these clouds, nature herself suffices to that purpose, and cures me of this philosophical melancholy and delirium, either by relaxing this bent of mind, or by some avocation, and lively impression of my senses, which obliterate all these chimeras. I dine, I play a game of back-gammon, I converse, and am merry with my friends; and when after three or four hour's amusement, I wou'd return to these speculations, they appear so cold, and strain'd, and ridiculous, that I cannot find in my heart to enter into them any farther."[32]

From my metaphysician's point of view, Hume has indeed argued himself into a depressing corner. It is all the more depressing because he is failing to adopt the most plausible strategy for understanding the world, which is to believe that our perceptions of properties suggesting persistence, objects, causation and the like are *prima facie* correct (at least), and proceeding with our philosophical work on this defeasible assumption. The "lively impression" of Hume's senses is exactly what he should be relying on for philosophical insight! Hume's overblown skepticism is a reaction to the overblown claims about necessity and the like made by traditional rationalists. A more moderate stance is called for on all sides. A metaphysician need make no claim for the *necessity* of the veridicality of our ordinary perceptions, rather, she may give perceptual impressions the same initial defeasible or prima facie credence as the scientist gives to empirical results.

[31] See, for example, Xu and Carey (1996).

[32] David Hume, *A Treatise of Human Nature*, §1.4.7.

So our metaphysician, like the scientist, can hold that our everyday experiences seem to give us a small amount of a posteriori knowledge of the features of the world. Our basic metaphysical concepts and beliefs about these features arise from this everyday knowledge, and these concepts and beliefs are the starting point for further work. This further work follows the model-based methods of theory development and assessment that we see in much successful scientific practice. The various theories are then compared with respect to their empirical adequacy and theoretical virtues. To the extent that one can endorse the realist view that scientific theories are true (and that we can infer truth from successful explanation), one endorses the thesis that maximizing theoretical desiderata brings one closer to the truth. To the extent that the naturalist endorses the thesis that maximizing theoretical desiderata brings one closer to the truth, the naturalist can endorse the view that doing metaphysics, and philosophy more generally, is a rational and reasonable way to try to discover fundamental and general truths about the world. (Hence, as I argued above, the success of science indirectly confirms metaphysical theories that maximize the theoretical virtues, and if maximizing the theoretical virtues leads to better scientific theories, it should lead to better theories more generally.)

Now, while it seems right, at least at present, to hold that categories or entities such as objecthood, persistence, composition, causation and others belong squarely in the realm of the metaphysical, the metaphysician must admit that it is possible that at least some of these things are actually in the realm of the empirical—it's just that we have not yet discovered them. Perhaps science will eventually succeed metaphysics.

Metaphysical theories exploring parts of the world that are in principle accessible to scientists could be taken as describing *toy* models of the empirical facts, where such models represent ways the world might be, given the information we have to date.[33] These models can be compared in terms of elegance, simplicity, empirical adequacy (to the extent that empirical facts are known) and consistency with contemporary science, but should not be adopted as true, that is, we should not infer their truth based on the sort of explanation they give. This is because a toy model of this sort is not providing the *best* explanation. It is merely giving us *an* explanation. The best explanation should include any as-yet unobserved data, and so the model explicitly leaves open the possibility that it fails to adequately represent parts of the world that have not yet been discovered. Science and empirical discoveries will ultimately determine which, if any, of the toy models provided by metaphysicians should be given the status of a true theory of the world.

In the case of toy modeling, one must note that the metaphysician is not studying what scientists already study, but is studying what scientists *could* study. For this reason, such models are valuable even if they are just toys, which is brought out by

[33] Here I agree with Godfrey-Smith (2006a). However, Godfrey-Smith also argues that *all* of metaphysics should be thought of as toy modeling, and this I am not sympathetic to. While some metaphysics might involve toy modeling, most of it doesn't. I see no reason to adopt this sort of antirealism about metaphysics, unless one wants to hold that all of the scientific models that incorporate metaphysical assumptions about causation, laws, persistence and the like are toy models as well. The antirealist or instrumentalist might well hold this. But the realist should demur.

the way that toy modeling plays a distinguished role in the early stages of most kinds of sciences. To name a few that rely most obviously on toy modeling, consider evolutionary biology, economics and rational choice theory, sociology, psychology and cognitive science.[34] The contribution of the metaphysician here would be to highlight the interest of the empirical question by discussing it, to provide hypothetical models that could in principle inform the investigations of scientists, and, perhaps most importantly from a philosophical point of view, provide ways of understanding the world that outrun what science has managed to accomplish thus far. This last task includes the job of providing toy models of parts of the world that science has neglected, but also includes the job of tying the models to our everyday understanding and experience of the world.

There is another way that science might eventually encroach upon metaphysics: perhaps we have not yet discovered the change in conceptual stance we need to make in order to measure categories such as objecthood, persistence, composition, and causation, and developments in science may show us the way. But the real import of this possibility must be understood clearly. For work in science to overhaul our current concepts of basic metaphysical features, there would have to be a significant conceptual revolution.

The best way to see how a basic metaphysical concept could be changed in interaction with science is to look at a historical example.[35] One of the most famous cases is that of space, and our concept of it as absolute or Euclidean. Kant famously argued that we had a priori knowledge of space as absolute and Euclidean. Even more famously, Einstein ushered in a scientific and conceptual revolution when he realized that by rejecting a Euclidean treatment of space in favor of Minkowski spacetime and the Riemannian theory of manifolds, he could provide an elegant relativistic treatment of mechanics. Thus, a conceptual revolution with respect to some basic metaphysical concepts ushered in a scientific revolution and successfully established spacetime and motion as empirically detectable entities.[36] Other sorts of deep conceptual revisions of basic metaphysical concepts would surely have very great impacts as well, since each such revision would entail fundamental changes in one's understanding of the physical world.

To understand how fundamental and radical such conceptual changes are, consider Friedman's (2001) discussion of how Einstein formulated his theory of general relativity.

> "[Einstein] focussed, to begin with, on the already well-known and well-established empirical fact that gravitational and inertial mass are equal, so that all bodies 'fall' with the same acceleration in a gravitational field. Einstein

[34] I'm indebted to suggestions from Peter Godfrey-Smith here.

[35] Friedman's *Dynamics of Reason* is a lovely study of some of these sorts of cases (although Friedman uses the examples to defend the neoKantian idea that science depends on conventionally chosen constitutive frameworks). See Friedman (2001).

[36] One thing I've been glossing is whether there is still a basic ontological concept of space (or time, etc.) underlying the empirical concepts of absolute space and relative space. I am inclined to think there is. If there is, then truly revising the concept of space would require an even deeper conceptual revolution. One might use this point against the thesis that when science is complete we will have shown that everything is a member of the empirical realm.

> then leapt from this well-established empirical fact to the bold 'heuristic' principle that gravitation and inertia are the very same phenomenon. He proceeded, on this basis, to construct models of gravitational fields from 'inertial fields' (generated by non-inertial reference frames wherein inertial forces, such as centrifugal or Coriolis forces, arise) in a special relativistic space–time, and he eventually saw that non-Euclidean geometries are thereby associated with gravitational fields. The final step was to take the four dimensional space–time metric, first introduced by Minkowski in the case of a special relativistic flat space–time, as our representative of the gravitational field, and to describe the variations in the space–time curvature associated with this metric by the Einsteinian field equations. The result was Einstein's general theory of relativity, only completed, after a long struggle, in 1915–16."[37]

The fact that such deep conceptual changes can occur when basic concepts are reclassified or revised does nothing but strengthen the position of the defender of metaphysics. In response to philosophers of science and others who argue that metaphysicians should merely look to science, especially current physics, for their results, the metaphysician can reiterate that the work of normal science (i.e., the everyday work of filling in the details of what is in what category, such as what the actually instantiated properties are), need not be the main domain of interest for contemporary metaphysicians. More importantly, the metaphysician can agree that metaphysics should pay (and historically *has* paid) very great attention to conceptual and scientific revolutions, since it is one way to discover new facts about metaphysical natures.

When we look at the history of the Einsteinian revolution, we see that there was in fact quite a bit of engagement between philosophers and scientists during the development of the new concepts. Scientists like Einstein, Mach, Helmholtz and Poincare engaged in philosophical discussions and used philosophical means to develop their views. Metaphysicians and philosophers of science in turn paid close attention to the scientific developments, and metaphysicians to this day respect the fact that facts about spacetime and motion, among other things, have been shown to be empirically determinable. The way that conceptual and scientific revolutions relate also shows how there is a clear role for metaphysics—even if metaphysical theories turn out (in some instances) to be toy models—for such modeling provides just the sort of information that scientists can use when refashioning their basic concepts to construct new theories.

There is a further conclusion to be drawn. Truly revising basic metaphysical concepts in order to develop radically new scientific theories that have revolutionary empirical impact would have an enormous influence on our scientific picture of the world. If, as some may think (although I am skeptical), there really are no deep metaphysical facts about natures after all, then when science is complete, all metaphysical modeling will eventually be discovered to be toy modeling.

[37] Friedman (2001) p. 89–90.

This is a radical result: it entails that science has a great many revolutions before it, since at the moment science relies on a good-sized stock of concepts of metaphysical entities. If we are in this situation, science as we know it is *not* close to a complete understanding of the world or any sort of grand unified theory: rather, there are very many more revolutions to undergo before we can think we are close to a complete picture of the world, and whatever our final theory will be, it will look almost nothing like our best current fundamental theories of the world.

If we do indeed find ourselves in this situation, then there is no shame in the fact that metaphysicians find themselves with a wide variety of acceptable theories for a given entity. And this brings out the epistemic value of diversity—having a range of conceptual options increases the depth and breadth of our understanding of the world, and should ultimately result in more successful final theories. If physics has a long way to go, it will need metaphysics for most of the way. Friedman argues that there is a special role for "the philosophical articulation of what we might call meta-paradigms or meta-frameworks for revolutionary science capable of motivating and sustaining the transition to a new scientific paradigm." Agreed.[38]

Note that while my arguments for the truth-conduciveness of metaphysics will not sway the constructive empiricist—since he doesn't even accept an inference to the belief in the objective existence of electrons, he's certainly not going to accept an inference to the belief in the objective existence of a composition relation—my thesis of the role of basic concepts in scientific theorizing carries some bite even for him. While the empiricist will likely prefer to hold that the organizing concepts we use to make sense of the world are chosen in a conventional or purely pragmatic way, he will agree that once the concepts are in place that they form the lens with which to understand science.[39] Moreover, he will agree that science will likely continue to undergo significant changes as it moves through revolutions. But if he agrees to this, then he can hardly reject the importance of metaphysical modeling, since as history shows, such modeling plays an important role in conceptual change and its attendant scientific theory development. So while the empiricist, unlike the realist, need not accept that metaphysicians are using inference to the best explanation to develop true theories, he must still accept that the practice of metaphysics stands side by side with natural science as an important and legitimate developer of our conceptual schemes. A metaphysician need not ask for more.

Acknowledgements Thanks are due to Robert Adams, Ross Cameron, Richard Healey, Chris Hitchcock, James Ladyman, Jonathan Schaffer, Eric Schliesser, Michael Strevens, Tuomas Tahko, Peter van Inwagen and Peter Godfrey-Smith for discussion.

References

Albert, D. (1996). Elementary quantum metaphysics. In J. Cushing, A. Fine, & S. Goldstein (Eds.), *Bohmian mechanics and quantum theory: An appraisal* (pp. 277–284). Dordrecht: Kluwer.

Atwood, M. (1985). *The handmaid's tale*. Toronto: McClelland and Stewart.

[38] Friedman (2001) p. 44. Truth in advertising compels me to note that I am using this quote for my own ends, since Friedman and I differ markedly on our attitudes towards the a priori and to metaphysics.

[39] Cf. Fraassen (1980).

Cameron, R. (2007). The contingency of composition. *Philosophical Studies, 136*, 99–121.
Downes S. (1992). The importance of models in theorizing: A deflationary semantic approach. In D. Hull, M. Forbes & K. Okrulik (Eds.), *Proceedings of the Philosophy of Science Association*, vol. 1, East lansing, (pp. 142–153).
Friedman, M. (2001). *Dynamics of reason*. Chicago: University of Chicago Press.
Godfrey-Smith, P. (2006a). Theories and models in metaphysics. *Harvard Review of Philosophy, 14*(2006), 4–19.
Godfrey-Smith, P. (2006b). The strategy of model-based science. *Biology and Philosophy, 21*, 725–740.
Goldman, A. (2007). A program for 'Naturalizing' metaphysics, with application to the ontology of events. *The Monist, 90*(3), 457–479.
Hall, N & Paul, L. A. (2013). *Causation: A user's guide*. Oxford: Oxford University Press. (forthcoming).
Healey, R. (1991). Holism and nonseparability. *Journal of Philosophy, 88*(8), 393–421.
Ichikawa, J., & Jarvis, B. (2009). Thought-experiment intuitions and truth in fiction. *Philosophical Studies, 142*(2), 221–246.
Jenkins, C. S. (2008). Modal knowledge, counterfactual knowledge and the role of experience. *Philosophical Quarterly, 58*, 693–701.
Ladyman, J., Ross, D., Spurrett, D., Collier, J., et al. (2007). *Every thing must go: Metaphysics naturalized*. Oxford: Oxford University Press.
Lewis, D. (1973). Causation. *Journal of Philosophy, 70*, 556–567.
Lewis, D. (1986). *Events, philosophical papers, vol 2* (pp. 241–269). New York: Oxford University Press.
Lewis, D. (2004). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (p. 76). Cambridge: MIT Press.
Lloyd, E. (1988). *The structure and confirmation of evolutionary theory*. New York: Greenwood Press.
Loewer, B. (2004). Humean supervenience. In John Carroll (Ed.), *Readings on laws of nature* (pp. 176–206). Pittsburgh: University of Pittsburgh Press.
Maudlin, T. (2007). *The metaphysics within physics*. Oxford: Oxford University Press.
Monton B (2012) *Prolegomena to any future physics-based metaphysics. Oxford studies in philosophy of religion, vol III*. Oxford: Oxford University Press.
Nersessian, N. J. (1999). Model-based reasoning in conceptual change. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 5–22). New York: Kluwer Academic/Plenum Publishers.
Ney, A. (forthcoming). The status of our ordinary three dimensions in a quantum universe. *Noûs*.
Paul, L. A. (2010a). Temporal experience. *Journal of Philosophy, CVII*(7), 333–359.
Paul, L. A. (2010b). New roles for experimental work in metaphysics. *European Review on Philosophy and Psychology, 1*(3), 461–476.
Paul L. A. (2012). The fundamental constituents of the world. *Philosophical Studies*
Russell B. (1988 edition). *The problems of philosophy*, New York: Prometheus Books.
Schliesser, E. (2011). Newton's challenge to philosophy: A programmatic essay. *HOPOS: The Journal of the International Society for the History of Philosophy of Science, 1*, 101–128.
Sober, E. (2008). *Parsimony arguments in science and philosophy: A test case for naturalism*$_{P_{...}}$ Romanell Lecture given to the Central Division meeting of the American Philosophical Association in December 2008.
Strevens, M. (2008). *Depth*. Cambridge: Harvard University Press.
van Fraassen, B. C. (1980). *The scientific image*. Oxford: Clarendon.
van Inwagen, P. (1990). *Material beings*. Ithaca: Cornell University Press.
van Inwagen, P. (2002). The number of things. *Philosophical Issues, 12*, 176–196.
Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.
Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology, 30*, 111–153.