Philos Stud (2012) 160:97–113
DOI 10.1007/s11098-012-9913-8

Metaphysics and the philosophical imagination

Peter Godfrey-Smith

Published online: 13 April 2012

Abstract Methods and goals in philosophy are discussed by first describing an ideal, and then looking at how the ideal might be approached. David Lewis's work in metaphysics is critically examined and compared to analogous work by Mackie and Carnap. Some large-scale philosophical systematic work, especially in metaphysics, is best treated as model-building, in a sense of that term that draws on the philosophy of science. Models are constructed in a way that involves deliberate simplification, or other imaginative modification of reality, in order to make relationships visible or problems tractable.

Keywords Metaphysics · Model-building · Supervenience

1 Introduction

What is it to give a philosophical analysis of causation? Or of truth, knowledge, or moral goodness? The aim of this paper is to look at methods and goals in philosophy. Not all kinds of philosophy will be covered. The focus will be on work with a particular kind of motivation—work organized around a concept, piece of language, or apparent feature of the world that is important to us, but which raises problems of some kind. Not all philosophy is organized in this way; some just dives into reality at a promising spot and tries to describe how things are. Here I am concerned with work guided by "classic problems" of the kind listed above.

The paper has the following structure. I first describe an *ideal*—the shape that a complete solution could take in an area like this. I then describe ways of approaching the ideal, given the information and tools available to us. Some of

P. Godfrey-Smith (✉)
Philosophy Program, The Graduate Center, City University of New York (CUNY), 365 Fifth Ave., New York, NY 10016, USA
e-mail: p.godfreysmith@gmail.com

David Lewis's work in metaphysics, his work on "Humean Supervenience", is critically examined in the light of the ideal. After this I argue that the best way of treating some large-scale theoretical systems in philosophy is different from the way they are usually treated. I see these systems as exercises in *model-building*, in a specific sense of that term that draws on the philosophy of science. Models, in this sense, are constructed in a way that involves deliberate simplification, or other imaginative modification of reality, in order to make some relationships visible or problems tractable. The claim is not that all philosophy, or all metaphysics, is or should be model-building in this sense. It is a strategy that helps with certain kinds of problems, yields a particular kind of understanding, and has costs and benefits.

2 An ideal

One way to ask how to approach topics like causation, truth, and knowledge is to ask what an ideal account would look like in a case like this—an account that was both complete and well-organized. I suggest that such an account would have four elements; it would involve the completion of four different projects. I will label these projects 1, 2, 3, and 4, and also refer to facts of type 1, 2, and so on. A fact of type 1 (for example) is a fact of the sort that is being inquired about in project 1.

Project 1 gives a description of some features of our thought and talk. If our task is an understanding of knowledge or causation, this first set of facts will include a description of how the term "know" or "cause" is used, along with the term's relatives. It will also include a description of some aspects of our thinking. This can be thought of as the description of a "concept", but I understand that word in a broad way, without commitment to any psychological theory of concepts, and mostly as a gesture towards *whatever* is going on inside, guiding usage, when someone uses one of these terms and its relatives. Project 1 also includes a description of aspects of our thinking that may not be closely tied to language use. In the case of causation, for example, the entire psychology of causal thinking is relevant.[1] Project 1 also covers the goals people have when they use the term in question, the intended contribution that this side of our thought and talk make to our lives. This psychological and linguistic account includes all people, not just philosophers, and it notes cultural variation and change over time. So project 1 is concerned with all the facts about the *use and practice* surrounding the problematic term.

Project 2 is a description of a part of the world—the part that the term in question is apparently used to deal with. We describe this part of the world using whatever language is clear and helpful. In this paper I work within a naturalistic approach to project 2, though this is not necessary for most of the claims I will make. And while science may be the source of the information used in project 2, this project is not one of merely collating and transcribing "raw" scientific work on the topic. Scientists' descriptions of the world may be full of metaphor, language chosen for

[1] See Gopnik and Schulz (2007) for recent work on causal cognition.

convenience and sharpness, and dubious philosophy. The description of the world given in project 2 takes the form of "philosophically processed" science.[2]

Project 3 is a description of how the first and second sets of facts are related. Questions asked here include the following: has the term investigated in project 1 picked out a "natural kind"? Is there some real connection in the world, a connection between events or facts, which we are successfully talking about at least some of the time when we use the term "cause", and which our causal thinking is tracking? Or does our ordinary use of the term "cause" embody a factual error, as Mackie (1977) argued is found in ordinary talk of moral goodness?.[3]

Though an ideal account includes facts of all these kinds, a philosopher might only care about 2. A philosopher can use a term like "cause" just to gesture towards the subject matter they are interested in, without looking closely at how the term is normally used. If, on the other hand, someone is pursuing project 2 with a strong interest in 1 and 3, it might be advisable to avoid using the crucial term from project 1 while working on project 2. While investigating the features of the world that the word "cause" is directed on, it might be best not to describe them using "cause". In some cases this might be very difficult or impossible; there might be no way to say anything significant within project 2 without using the term whose status is at issue. This would not undermine the combination of projects described here, however. If this happens, it should simply be noted, and it would take its place as an interesting type 3 fact.

One way that a picture of roughly this kind has been discussed recently is in terms of "roles" and "occupants" (or similar terminologies. See Lewis 1972; Jackson 1998). The idea is that "platitudes" accepted by ordinary users of a term specify a *role*—the causation role, the belief role, and so on. We look at the actual world to find the *occupant* of the role, if there is one. This is seen as a way of solving the "location problem" for problematic parts of our common-sense picture of the world.

Role-occupant analysis is one way of pursuing the combination of projects described here, but it uses a particular linguistic model which many cases may not fit. For example, a non-cognitivist view of moral language, such as prescriptivism, is a project 1 position. The prescriptivist thinks that an attribution of moral value functions as a special kind of command (Carnap 1937; Hare 1963). If prescriptivism is true, there can be no role-occupant relationship for moral value. This is not because nothing *occupies* the moral value role, but because there *is* no moral value role in the relevant sense; moral talk does not determine a role that may or may not be occupied. In a broader sense of "role", prescriptivists certainly think that moral talk has a role within our lives—there is something that we use this talk to try to do. But this role, in the broader sense, does not involve the specification of a role, a "slot", that might be filled by a worldly occupant that moral talk is about. So prescriptivism of this kind takes us outside the role-occupant approach, but a

[2] Godfrey-Smith (2009), chapter 1, discusses this kind of work in more detail.

[3] Russell (1917) is often described as claiming that our ordinary concept of cause is based on assumptions incompatible with modern physics, and hence should be abandoned. I am not sure from Russell's discussion, however, whether he thought that the errors that infect philosophers' talk of causation also infect ordinary talk, as Mackie did think in the case of morality.

possibility like this does not prevent a 1-2-3 analysis of the kind outlined here. Project 3 would then describe how features of human social life sustain a certain kind of moral language despite the absence of reference and truth—despite the absence of even *attempted* reference and truth. The role-occupant approach assumes referential connections of a certain kind between language and the world. That is one possible connection but not the only one (Blackburn 1993), and the framework described here is intended to admit a wider range of possibilities.

Once we know how a term or concept works, and what some relevant parts of the world are like, we might consider doing things differently. This brings us to project 4, which looks at the possibility of reform of the patterns of use and practice described in project 1. In the case of causation, there might be a better way of talking about the relations of dependence between events. Alternatively, it might be demonstrated that the concept of causation we presently have is as good as anything else we can imagine.

Discussions of reform will be based largely on facts of type 3, but the 4th project also requires input of other kinds. It requires information about the purposes behind our use of the term, and about a range of alternative ways of talking and thinking. Here I will concentrate on the question of purposes. Without information of that kind, we cannot address project 4. We might want "cause" or "knowledge" to be natural kind terms that pick out something real and important; we might instead want them to be practical tools for use in informal discussion, terms that need not bear any theoretical weight.

Recent work on project 1 in the case of both causation and knowledge has considered recognizing a *context-sensitivity* in attributions, where the "context" here is that of the speaker (see De Rose (1992) and Lewis (1996) for knowledge, Menzies (2007) and Swanson (2010) for causation). Assume that two people in different contexts are talking about another person's belief. According to a context-sensitive analysis of knowledge, one speaker might say that the third person *knows* such-and-such, while the other will deny it, and neither need be mistaken by ordinary standards. Perhaps knowledge is true belief based on evidence that rules out all *relevant alternatives* to what is believed. In one context of discussion, X might count as a relevant alternative that has not been ruled out, while in another context of discussion X might not count as relevant.[4] In the case of causation, the phrase "the cause of" has often been seen as somewhat context-sensitive, because the promotion of something from "a" cause to "the" cause can be a matter of the speaker's interests. Swanson (2010) has argued that even whether something is *a* cause of an event is context-sensitive. As he sees it, our linguistic norms require that we use "good representatives" of a causal path leading to an effect, when describing the causal role of that path. Whether something is a *good* representative can depend on the context of discussion.

Suppose both analyses are right. In both cases, context-sensitivity in attributions may help with the rapid conveying of information. This would also make "cause" and "know" less useful as tools for theoretical description and explanation. This need not be a problem, as other language can be used for those tasks. (Swanson

[4] For criticism of this kind of type 1 analysis of knowledge, see Stanley (2005).

thinks that "was causally relevant to" is not context-sensitive. In the case of knowledge, one can simply talk about whether the evidence ruled out such-and-such a possibility of error.) When serious efforts at explanation are being undertaken, long-windedness is not much of an issue. So the choices here may or may not be pressing ones. But goals related to convenience and ease of use may trade off against goals related to theorizing and explanation.

Which goals are relevant? One answer is that the goals of present-day users of the term are the relevant ones. These are covered by project 1—they are part of the psychology and sociology. Then whether present usage succeeds in meeting present goals will be part of what is covered in project 3. For example, work on project 3 will tell us whether the concept in question has picked out a natural kind, and project 1 will tell us whether that outcome is a good one, given the goals of actual users. Are people *aiming* to pick out an natural kind, or do they want a linguistic tool of another type?

Existing goals are one thing, but we might have different goals, or want to reform them. A person might say: from now on I want to use "cause" for *this* purpose (more theoretical, more pragmatic…). Then given that new goal and facts of types 1, 2, and 3, a direction of reform will follow.

Something as basic as *consistency* in usage may or may not be very important. *Simplicity* can have either a minor or major role. Maybe consistency and simplicity are both always better *ceteris paribus*, but other things will often not be equal, and these two goals themselves might trade off against each other. All sorts of *possible* usages, ways of talking and categorizing, will arise as we describe facts of type 2. As we describe what there is, we will see new options for dividing it up. Whether any are worth keeping is a goal-dependent matter.[5]

A person might wonder at this stage about how we pick out "the relevant part of the world" for project 2. Where does this begin and end? Is there a risk of circularity? I reply that if there is uncertainty, one should simply expand the boundaries. Returning to the example of causation, a number of writers have argued that causal talk depends on contrasts between what actually occurs and the "normal" course of events (Menzies 2007; Hitchcock and Knobe 2009). Suppose this is right, and causal talk depends on judgments of normality. This pushes out the boundaries of project 1 for "cause", and also pushes out the boundaries of project 2. A full account of causal talk will include description of how people make judgments about what is normal, and the features of the world that talk of normality is aimed at dealing with. Further pushings-outward may be forced on us as well.

In general this possibility of expansion is not a problem, except on the practical side. In the ideal, all of our analyses would eventually knit together, yielding *total*

[5] With the possibility of reform on the table, I will briefly compare my framework with one outlined by Doug Kutach, applied especially to the case of causation. Kutach contrasts "orthodox analysis", which he opposes, with "empirical analysis": "An empirical analysis of X is a conceptual structure designed to optimize explanations of whatever empirical phenomena make X a concept worth having" (2010, p. 7 ms). Kutach's view includes a role for a purely psychological analysis of how we handle the idea of causation, and an analysis of what in the world makes this way of thinking and talking worthwhile. Our views are in the same spirit, with one difference being that Kutach's empirical analysis assumes that the concept in question *is* worth having.

type 1, 2, and 3 pictures. One problem that may arise from expansion which is not purely practical, however, concerns reform, project 4. It is easy to consider reform of our causal talk when we hold fixed other parts of our thought and talk, along with their goals. But if we are forced successively outwards, in the way illustrated above, it might become harder and harder to do this, because we are "taking a step back" from larger and larger proportions of our conceptual scheme. Too many planks of Neurath's boat are being loosened at once.

I will look at one other example to conclude this section, a helpfully difficult one. This is the case of *truth*. Starting at the top, the first thing we can do is investigate the use of the term "true" and its relatives, and how the concept of truth functions in our thinking. Next we describe the part of the world that the concept seems to be used to deal with. In the case of truth, this "part" of the world is apparently a set of relations between sentences and other objects. In the case of causation, word-world relations come into the picture only in project 3, but in the case of truth, they are the subject matter of project 2. At least, that is how things initially seem; most views of truth have supposed that our talk of truth is directed on a relation between representations and the world that is of great value, often hard to attain, and can be used to explain success of various kinds. Perhaps that relation is some sort of correspondence, and perhaps it is something else. Those views, and the debates around them, easily fit the picture sketched so far. But there are also "minimalist" or "deflationist" views of truth (Horwich 1990; Field 1994). These views hold that truth is a linguistic tool with a much thinner role than philosophers usually suppose. Our ordinary talk of truth is not directed on a special but elusive relationship between words and the world. We use the word "true" just to make certain linguistic moves convenient. These include generalization—we can say "everything Sam says about Toyotas is true"—and giving credit. There is no more to truth than that.

Minimalism of this kind is a view within project 1 that has unusual consequences for projects 2 and 3. It implies a "null 2". There is nothing in the world that our idea of truth is aimed at *dealing* with, except economical expression of whatever we might be saying. There is no more of a project 2 for "true" than there is for "and" or "unless".

There is also the possibility of an "error theory" about truth. Perhaps people do treat truth as a mirroring-like relation, but are wrong to do so: the causes of practical success by means of belief and representation are disparate, and while people try to refer all cases to a kind of picturing, this is a mistake. An unorthodox view with some elements of this option, but also with features that relate to minimalism, has been defended by Price (2003). Price thinks, within project 1, that standard minimalism is false because truth has a more substantial role in discourse than minimalism supposes. Truth is associated with a special norm of discussion. Within a certain class of topics, it I say *p* and you say not-*p* then one of us has erred, and we should find out who it is. The norms of ordinary discussion seem to involve commitment to a correspondence-like notion of truth, at least of a weak kind. When we turn to projects 2 and 3, Price opposes correspondence views. *Prima facie*, an analogy might then be drawn with theism. Most people believe in a special supernatural agent (a type 1 fact). There is no such agent (type 2). So most of their

talk in this area is false (type 3). But do we want reform of this side of our thought and talk? Should we keep the type-3 surprise quiet, because of the social and moral role of theistic belief? Project 4 depends on careful consideration of purposes. This seems to apply equally to God and truth. But Price thinks the analogy between God and truth breaks down. In the case of truth, reform is probably impossible. The idea of truth is integrated into how we handle *all* discussion, and we cannot work outside it. As a result, "it may be impossible to formulate a meaningful antirealism or fictionalism about the semantic terms themselves" (2003, p. 188). But if there is no coherent anti-realist option, there is no coherent realist option either. Given this, Price thinks that philosophical analysis stops at the point where the social function of truth has been isolated: "there is no further question of interest to philosophy, once the question about function has been answered" (p. 171).

The case of truth does pose special problems. I do not think these problems prevent the kind of analysis outlined here being carried out, however. This is so even if we work within a number of Price's own assumptions. For Price himself, project 2 is coherent enough for correspondence theories to be rejected—not just as unnecessary, as I read him, but as misguided. So there are some things we can say about 2, and hence 3, despite the difficulty of operating both *on* and *with* existing norms of assertion. If we can chip away at projects 2 and 3 with negative points, we can hope to end up saying something positive. And once we have some information about 3, it is possible to wonder about doing things differently. The point of this discussion of truth is not to resolve the problem, however, but to illustrate how this case fits into the framework, given its special features.

Summarizing the projects distinguished in this section, an ideal treatment of a problematic term or concept would comprise descriptions of four things:

1. Use of the term and its relatives, and surrounding practices.
2. The part of the world that this part of our thought and talk is used to deal with.
3. Relations between 1 and 2.
4. Any desirable reforms of 1.

3 Working toward the ideal, and Lewis's program

The ideal described in the previous section ignored practical issues. All sorts of subtle empirical matters were assumed to be knowable. How does the ideal relate to something that philosophers can actually do?.

The first answer that might be offered is that though the ideal cannot be attained, it can be approached, and to be closer is always better. That is a reasonable first thought, though I will try to improve on it below.

When addressing something like project 1, philosophers often consult their intuitions, and those of their colleagues. Can a bullet be a *cause* of someone's death if there was another bullet right behind it? Does Jones *know* there is a barn in that field? For some critics, this shows a reprehensible narrowness. For some defenders, it reflects an appropriate place for the a priori in philosophy. Whether or not this counts as a priori, the philosopher here is exploiting the fact that he or she is a

genuine user of the language being discussed. The philosopher's intuitions are some guide to a larger set of empirical type 1 facts. Though they are *some* guide, clearly a philosopher's intuitions and usage may be different from those of other people. When type 1 details become important, the philosopher will have to either confine their claims to the potentially idiosyncratic type 1 facts that hold in their immediate community, or look for data about a larger group. This could involve "experimental philosophy", or drawing directly on linguistics and psychology.[6]

With respect to project 2, well-established science is the natural starting place, though as emphasized above, the presentation of that material by scientists will not always be in the right form for its project-2 relevance to be read off.

So the ideal is not idle; it can guide work. Further, in some ways the distinctively philosophical part of this combination is project 3. Many debates in that area are concerned with coarse-grained alternatives. Are we forced to an error theory about moral value? Are any human actions free? An answer to a project 3 question of this kind often allows for a fair amount of flexibility regarding the details of facts of type 1 and 2. This is a correction to the view expressed at the start of this section, where I said that what we are always looking for is closer and closer approach to the total body of facts. When what is sought is an answer to a classic type 3 question, such as whether there are any causal relations in the world at all, it might be possible to establish that the type 1 and 2 facts fall into a *range* that determines a definite answer to that question, and further details of types 1 and 2 do not make a difference. The type 3 answer that results will be coarse-grained, but not merely approximate.

It is also possible to look for much more detail within project 3. If we decide that there are *some* causal connections of the kind required by ordinary standards, it is natural to then ask how many there are. Then the difficulty of learning the details of facts of type 1 and 2 becomes more of a problem. Still, even when fine-grained type 3 facts are the focus, there will be a vast amount of detail about our causal talk that does not matter much to project 3 questions.

I will next compare the picture developed here with a very influential and well-developed body of work that might appear initially to fit well with the ideal, but which does not. This example is David Lewis's program in metaphysics.[7]

Lewis aimed at achieving equilibrium between a common-sense view of the world, science, and demands of consistency and economy. The resulting analyses tend to have two pieces. One is a picture of how we talk in some area (about causes, chances, or the mind). The other is a picture of the world being talked about. From the 1980s Lewis summarized much of his project as a defense of "Humean supervenience". The picture of the world he used was Humean in a particular sense; the world is seen as a "mosaic" of local matters of fact. The mosaic contains

[6] For experimental philosophy, see Knobe and Nichols (2008). Experimental philosophy has two roles that are relevant here. One is investigating actual patterns of use and their variability, and the other is formulating hypotheses about the psychological basis for human intuitive judgments, which will apply even to the philosophers.

[7] In this discussion I will not be looking at the most controversial part of Lewis's metaphysical view, his realism about possible worlds. For the work discussed here, see especially Lewis (1983, 1986, Lewis 1999).

instantiations of natural properties by particulars, with no "necessary connections" between distinct matters of fact. This view can also be described as "atomist", though in a broad sense, as Lewis does not assume in advance that everything in a mosaic world will be physical. Lewis's aim was to find good candidates in the mosaic for causes, laws, chances, and so on; he aimed to find things we can be talking about when we use the language of "cause" and "law". Facts about causation, law (and so on) will then "supervene" on the mosaic.

For example, it is initially difficult to find causation in a mosaic world. The mosaic is a world of "loose and separate" things, to use Hume's phrase, and causation is supposed to be a "cement", a connector. Lewis aims to find features of the mosaic that suffice as causal facts despite the absence of metaphysical cement. His general strategy is the reduction of facts about connection to facts about pattern. Special kinds of pattern in a world, which can be described with counterfactuals, suffice for there to be connectedness of the kind that causation as normally understood involves.

The part of this view I will focus on is the mosaic view of the world itself. Here is Lewis's best-known statement of the view.

> [A]ll there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another…. We have geometry: a system of external relations of spatiotemporal distances between points…. And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. For short: we have an arrangement of qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that. (1986, pp. ix-x)

What is the status of this picture? Why work within it, as opposed to some other picture?

Sometimes Lewis says that his work is a "campaign on behalf of" this view (1986, p. ix). But Lewis accepts that Humean supervenience does not fit comfortably with current physics. The problem is not that physicists do not talk this way; we might have reason to describe their findings differently from the way they do. But modern physics seems to be positively at odds with it.[8] Humeanism of this kind might be seen as a philosophical interpretation of older and less exotic physics. Even that is doubtful; Bricker (1993), Maudlin (2007), and Hall (2010) argue that Lewis's treatment of space, for example, has problems even assuming classical physics. The "triangle inequality" looks to be a necessary truth about spatial relations, but Lewis would then have to treat it as a necessary connection between distinct existences.[9] This issue about space is of secondary importance in the present context, however, as it might be an unseen internal difficulty rather than something compromising the intended role of Lewis's Humeanism.

[8] For a detailed discussion of Humean supervenience and physics see Karakostas (2009).

[9] "If the distance between points A and B is x, and the distance between B and C is y, then the distance between A and C cannot be more than $x + y$. But why should this constraint hold, if the spatial relations between A and B, and B and C, on one hand, place no constraints on the spatial relations between A and C, on the other?" (Hall 2010).

Another puzzling feature of Lewis's position is the apparent weight given to aesthetic criteria. It is hard to track Lewis's tone on this issue: "Perhaps there might be extra, irreducible external relations, besides spatiotemporal ones … It is not, alas, unintelligible that there might be suchlike rubbish" (1986, p. x). Aesthetic criteria have no role in project 2 as I understand it, unless we have some reason to think that they point towards truth. Why should not the world contain rubbish? It won't contain incoherent or impossible rubbish, but why not coherent rubbish? I don't know of a passage where Lewis argues that aesthetic criteria of this kind do point towards truth.

Lewis says some things intended to clarify his campaign.

> What I want to fight are *philosophical* arguments against Humean supervenience. When philosophers claim that one or another common-place feature of the world cannot supervene on the arrangement of qualities, I make it my business to resist. (1986, p. xi)

We might see Lewis as concerned only with a conditional: *if* Humeanism is true *then* (given the right patterning) there is causation. But why is this conditional important if Humeanism is not likely to be true? I see Lewis's most promising defense as the one he reaches at the end of this next passage:

> The point of defending Humean Supervenience is not to support reactionary physics, but rather to resist philosophical arguments that there are more things in heaven and earth than physics has dreamt of. Therefore if I defend the *philosophical* tenability of Humean Supervenience, that defense can doubtless be adapted to whatever better supervenience thesis may emerge from better physics. (1994, p. 226)

This will be a good defense, however, only if a further assumption is true, the assumption that there will a certain kind of continuity between classical physics and a better future physics. Why should we believe that assumption? I do not know of an argument that Lewis gives on this point. But Brian Weatherson gives an argument that I see as developing the defense above while asserting a different relationship between Humeanism and future physics.

> What Lewis's defense of Humean supervenience gives us is a recipe for locating the nomic, intentional and normative properties in a physical world. And it is a recipe that uses remarkably few ingredients; just intrinsic properties of point-sized objects, and spatio-temporal relations. It is likely that ideal physics will have more in it than that. For instance, it might have entanglement relations, as are needed to explain Bell's inequality. But it is unlikely to have less. And the more there is in fundamental physics, the *easier* it is to solve the location problem, because the would-be locator has more resources to work with…. So Lewis's defense of Humean supervenience then generalises into a defense of the compatibility of large swathes of folk theory with ideal physics. (2010)

Weatherson claims not that he can predict that future physics will be *similar* to Humean physics, but that another feature of the Humean picture makes it likely that

Lewis's larger project can succeed. This is the *weakness* of the physical assumptions used by Lewis to show the reality of causes (and other problematic relations); if we can find causes in a Humean world, we can find them anywhere.

The details of Weatherson's bridge between Humeanism and future physics can be questioned. Physical connectedness might be a good thing for causation, but this is a case where you can have too much of a good thing, at least for causation in the ordinary sense. A very holistic picture of the physical universe, in which everything is "responsible" for everything else, is a picture in which our ordinary concept of cause faces as many problems as it does in a disconnected world. So Weatherson's argument seems too quick. We need to believe that future physics will recognize no less structure than Humeanism recognizes, and will not add the wrong things either. This, however, is a secondary point. What I would emphasize is that the best reason we have ended up with for working within Humeanism does not involve a "campaign" for the truth or even "tenability" of the view. This defense gives the Humean mosaic a quite different role. Turning to a different language, it is treating the Humean mosaic as a *minimal model* of the physical world, which can be used to explore how easy or hard it is to solve various location problems. A minimal model can be useful without being accurate. Once we say that, it shows a way to start afresh on this family of issues.

4 Modeling

Suppose you are a social scientist or biologist, and you want to understand how cooperation can survive in environment in which agents are generally self-interested and opportunities for exploitation reliably arise. The systems you are interested in are extremely complex and subject to diverse influences. One thing you might do is describe an imaginary set-up which has some core features of what you care about. This system is imaginary or hypothetical, but you want to see how it *would* behave. To constrain this work, you connect the imaginary system to some mathematical structure if you can, at least imagine it in a mathematically guided way. Once you know how the imaginary system behaves, you try to use it a guide to what goes on in the more complicated real system with which you began.

A famous example of this sort of work is Robert Axelrod's *The Evolution of Cooperation* (1984). Axelrod used game theory, especially the "iterated prisoner's dilemma". He showed the power of simple forms of reciprocity in maintaining cooperation, even when immediate benefits can be gained by exploitation, especially the power of a behavioral rule invented by Anatol Rapaport called "tit-for-tat". In an iterated interaction involving a pair, where choices are simultaneous and the only options are "cooperate" and "defect", tit-for-tat is a strategy that begins by cooperating and then copies whatever its partner did on the previous trial. So it starts friendly, retaliates quickly, but forgives quickly too. Axelrod's method involved a mixture of computer modeling and algebra, going back and forth between the two. In a typical scenario, we assume a population with different types of individuals, playing different strategies. There is some rule of interaction, and some rule of updating of the population composition according to

the payoffs that individuals receive. Axelrod showed that although it can be hard for a strategy like tit-for-tat to initially get established in an unfriendly population, it can be quite successful once it has a foothold.

Axelrod's work on this topic involves constrained acts of imagining of hypothetical or fictional situations. Most of what is described does not occur in any real system. The "rules of updating" mentioned above, in particular, are usually extremely unrealistic—often a form of asexual reproduction. But what happens in such systems can be relevantly similar to what occurs in real situations. So work proceeds by working out how the imagined systems behave, and then looking at relations between the imaginary and the empirical. Complexity can be added to the model, narrowing the gap between imagined and real, but this occurs in the context of a trade-off between complexity and tractability: the more you add to the model, the less transparent is its behavior. In successful work of this kind, people often move back and forth between slightly different packages of assumptions, with each package containing a mix of realistic and unrealistic elements.

Assertions within this work are made in several modes. Some are made "inside" the model. A person might say that such-and-such a system will not reach an equilibrium, even though everyone has assumed that it would. Other claims are overt model-world comparisons: here are the realistic features of the model, and here are the simplifications. Axelrod wanted to understand the rise of un-sanctioned cooperation across No Man's Land in the Western Front during World War I. So he compared what happened there with what happens in his models. One can also try to make further claims about the actual world, including empirical predictions, on the basis of the model. The upshot of a model like this can be expressed with conditionals: "If X holds, Y will happen". This is said while knowing that X does not hold in the actual system. We might know that *approximately* X holds, however. If we know that "If X then Y" and "Approximately X", this is no guarantee that Y, or approximately Y, will occur. But a conditional of this kind can often be a good guide to actual events. Working out when is part of the craft of modeling.[10]

This kind of model-building is ubiquitous in biology, social science, and areas like climatology. It is not characteristic of all scientific theorizing, and the word "model" itself is sometimes used for work that is different, such as work that is cautiously presented but not intended to take a deliberate detour through the imaginary. Taking such detours is one strategy, a strategy of imposing simplicity and order in the service of getting results. These "results" apply in the first instance to an imaginary case, but often also tell us about real cases.

What role might this have in philosophy? Philosophy obviously contains much use of fiction and the imagination; it is full of thought-experiments and imaginary cases designed to illustrate principles and probe intuitions. Some of these are quickly devised and thin: "Well, what if we had someone whose brain was attached to a …" Others are developed more carefully, such as the "original position" of John Rawls' political philosophy (1971). I think there are also larger-scale and less

[10] For relevant discussions of modeling in science see Godfrey-Smith (2006a) and Weisberg (2007). For a relevant discussion of inferences in science involving conditionals and "approximately" operators, see Marshall (forthcoming).

obvious examples of model-like structures.[11] In particular, the kind of system Lewis was working with, his mosaic, was a model-like construction. He did not see it that way, as far as I know. He saw it as a good candidate for a full account of the fundamental structure of the world. So I call it model-*like*; it is something that would better have been treated as a model, and can now be treated that way, even though it was not treated that way by Lewis. The Humean mosaic is a minimal model of the universe, which can be used to explore how various project 3 questions *could* be answered, especially in cases where it seems hard to do so.

As in science, modeling of this kind is an optional move, one strategy among several. A good metaphysical model will be one that has relevant similarities to a puzzling target system, while being sufficiently simple for its exploration and manipulation to yield definite results. To see the contrasts between a modeling strategy and others, we can compare Lewis with another body of work on causation, developed around the same time, but largely supplanted by Lewis in ongoing discussion. I have in mind John Mackie's *Cement of the Universe* (1980). Mackie's approach was close to a simple application, with no use of models, of the approach sketched earlier in this paper. He studied what he took to be "our concept" of causation, then noted what the world seems to contain, and gave a complicated—in some ways awkward—account of the relation between the two. For Mackie as for Lewis, causal thinking about singular cases is closely tied to counterfactual thinking, but the counterfactuals that are relevant do not have determinate truth values. Singular causal claims inherit this feature, even though we can see how some real characteristics of the world—regularities and certain kinds of physical continuities—make it useful to engage in causal thought and talk of the kind that is familiar to us. They "sustain" our causal talk even if they do not make it true. Mackie did not use a total metaphysical system such as a Humean mosaic in his account of the type 2 facts. In retrospect, we can see the strengths of both approaches; one treatment was more elegant and fruitful, the other better grounded.

Once a theoretical construction is being treated as a model, some distinctive epistemological ideas become relevant. "Robustness" arguments are an example. In my discussion of game theory above, I noted that successful work often considers several variants of an imagined scenario, with different simplifications. The aim is to see whether certain results are "robust" across these different versions of the model. Robustness-based methods could have a role in the projects I have been discussing here. We would first imagine simplified versions of the type 1 and type 2 facts, in order to answer a question in project 3, and then look at how sensitive our conclusion was to the specific simplifications that were made. There are also relatives of this procedure that would not be so good: re-shaping facts of type 1 and 2 to make them amenable to a desired project 3 outcome, and then arguing that the type 1 facts can be "reconstructed as" fitting this imagined version, or that we have "no choice" but to accept the project 2 description that fits with the desired outcome in project 3. Coming up with a modification of the type 1 and 2 facts that brings them into line with attractive conclusions about project 3 is something we could

[11] Other aspects of this suggestion are discussed in Godfrey-Smith (2006b). See also Paul (forthcoming) and Balcerak Jackson (forthcoming) for different ways of treating philosophical work as modeling.

certainly do as an exercise, to see how it would look, but that is very different from trying, even with the aid of models, to bring the real 1 and 2 together.

I will discuss one more issue concerning modeling. The emphasis above was often on simplification. But not all modeling maneuvers are solely a matter of simplification. There are also considerations of *tractability*, which is related to simplicity but not the same thing. For example, in science choices are often made to treat a system as discrete when it is probably continuous, or vice versa. Neither of these is obviously a move in the direction of simplicity, but it often makes for tractability, as a result of the methods at hand.

A good proverb has it that "to someone who only has a hammer, everything looks like a nail". Not all things are nails, but there is a legitimate modeling maneuver here: "Given how good my hammer is, I will pretend these things are nails and see if it gets me somewhere". Part of the craft of modeling is realizing when this approach is helpful and when it is not. Without working within a framework of the kind I use here, Hall (2010) has suggested that a phenomenon like this can be seen in further features of Lewis's work on Humean supervenience. In Lewis's mosaic, particulars instantiate "natural properties". Particulars can instantiate any combination of natural properties, except when one logically excludes another, and this exclusion is the kind treated in predicate logic. Hall argues that this fits poorly with the ontology used in physics for centuries now, in which physical *magnitudes* are attributed to things. Suppose an object has a mass of 5 kg. This excludes it having a mass of 7 kg, but if *having mass 5 kg* is one natural property, and *having mass 7 kg* is another, there is no apparent reason why one should exclude the other. There would have to be some reduction of these magnitudes to more basic properties, so that *being G* is a matter of *not being F*, and so on. But magnitudes like mass are as basic in physics as anything gets.

As Hall sees it, Lewis is influenced here by the desire to express as much as possible in the language of first-order logic. That logic is not well-equipped to deal with magnitudes of the kind that physics now uses. This is reminiscent of the proverbial hammer. To someone whose favorite format is predicate logic, everything looks like an array of particulars instantiating properties and standing in relations.

Earlier I compared Lewis's work with Mackie's. A comparison can also be made with Rudolf Carnap (especially Carnap 1950). Carnap's aim was not to analyze causal connection, but induction. He wanted to know how observing some parts of the world can give us reason to draw conclusions about other parts. Carnap worked within an atomist framework similar to Lewis's. But he was quite explicit about the role of simplification, and imagined modification of the world, in this work. He says that the most straightforward application of the system in *Logical Foundations of Probability* is to a "simplified" universe in which qualities and relations behave quite differently from how they do in reality (1950, pp. 73–74). He also notes that because his framework cannot handle quantitative magnitudes like mass, it is "not yet applicable to the entire language of science" (p. v).

Taking a longer view, atomism is a long-standing preference in the Western theoretical tradition. People were attracted to atomist views about the physical world long before there was good evidence for them. Locke and Hume, influenced

by Newton, exported an atomist pattern of explanation into a new domain, the mind. 20th century "logical atomism", to use Bertrand Russell's term, continued along this path. Russell's logical atomism (1918/1985) reflected both dissatisfaction with the "monist" views of 19th century idealism and a sense that much could be *done* with new methods, derived from mathematics and logic, that can be employed within an atomist picture. The new methods were indeed successful. Topics that had seemed to generate nothing but rubbish became the focus of rigorous and cumulative work. Atomism is a particularly *useable* metaphysics.

John Dewey (1929/1988) gave a diagnosis of metaphysical system-building, with particular attention to atomism. Dewey saw philosophical systems as products of "selective emphasis". The philosopher seizes upon one aspect of experience and tries to make it the basis of everything. Other things we encounter in experience receive some mixture of reduction, deflation, and elimination. In different cultural and scientific circumstances, different aspects of experience will suggest themselves as fundamental. For Dewey, selection-based constructions of this kind are not problematic, and can be used to reveal useful new facts, as long as the act of selection is not obscured. Obscuring it is "the source of those astounding differences in philosophic belief that startle the beginner and that become the plaything of the expert" (p. 30).

Applying Dewey's idea to the cases discussed here, ordinary experience of the world contains a mix of the separated and the connected. Philosophically, either can be taken as basic, requiring an explaining-away of the other. This leads to the astounding differences between everything-is-connected holist philosophies and nothing-is-connected atomist ones. In 19th century philosophy the unified and connected were often given primacy. In extreme forms, of the kind Moore and Russell reacted against, hidden connection was attributed to everything. In the 20th century, the separated and discrete were given primacy instead, especially as the new logic came into philosophy.[12] This move was fruitful. Was it fruitful in a way that provides support for the truth of an atomist view? Answering that question would require a close look at the different fruits. Establishing interesting conditionals, of the form "*if* atomism was true, the world could still contain relations recognizable as causal" is fruitful, but not in a way that bears on the truth of the antecedent. Dewey would say that as far as purely philosophical modeling goes, neither the atomist nor holist picture can claim to be more faithful to how things are, as both involve acts of selection that have a similar status. Something Dewey may not have countenanced is that one side or the other, understood as a view of the world as a whole, may be vindicated by physics.

In the first part of this paper I outlined an ideal which can be used to think about philosophical work of a certain kind and what it achieves. I then gave a sketch of how various methods that people do or might use relate to the ideal. Modeling is one of these. In particular, modeling is one way of dealing with the combination of difficulty and explanatory ambition seen in metaphysics. Lewis's metaphysical

[12] There are relevant connections between some of this early 20th century work, represented by the Vienna Circle, and the development of "modernist" ideas outside philosophy, for example in architecture and design. The close interest in these currents on the part of Carnap and other members of the Vienna Circle is charted in Galison (1990).

work has a modelish character even though it was not presented or apparently understood that way. Mackie's work shows what work on similar topics conducted outside a model-like system looks like. Carnap is an example of someone who was explicit about maneuvers made in the service of tractability—in the service of making his preferred tools useable.

Acknowledgment A version of this paper was given at the Summer Causation Workshop at the Center for Time, Sydney University, in 2011. The paper has also been influenced by discussion at the Surrogate Reasoning in Philosophy and Science Workshop at the RSSS, Australian National University, 2010, and in Ned Hall's "Explanatory Structures" Seminar at Harvard University in 2010. Thanks are due to an anonymous referee for very helpful comments.

References

Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.
Balcerak Jackson, M. (forthcoming). Thought experiments and model-based philosophy.
Blackburn, S. (1993). *Essays in quasi-realism*. Oxford, NY: Oxford University Press.
Bricker, P. (1993). The fabric of space: Intrinsic vs. extrinsic distance relations. *Midwest Studies in Philosophy, 18*, 271–294.
Carnap, R. (1937). *Philosophy and logical syntax*. London: Kegan Paul, Trench, Trubner and Co.
Carnap, R. (1950). *Logical foundations of probability*. Chicago, IL: University of Chicago Press.
Collins, J., Hall, E., & Paul, L. (Eds.). (2004). *Causation and counterfactuals*. Cambridge, MA: MIT Press.
De Rose, K. (1992). Contextualism and knowledge attributions. *Philosophy and Phenomenological Research, 52*, 913–929.
Dewey, J. (1929/1988). *Experience and nature* (Revised edn). In: J. -A. Boydston (Ed.), *John Dewey*: *The later works*, *1925–1953*, vol. 1 (reprinted). Carbondale, IL: Southern Illinois University Press.
Field, H. (1994). Deflationary theories of truth and content. *Mind, 103*, 249–285.
Galison, P. (1990). Aufbau/Bauhaus: Logical positivism and architectural modernism. *Critical Inquiry, 16*, 709–752.
Godfrey-Smith, P. (2006a). The strategy of model-based science. *Biology and Philosophy, 21*, 725–740.
Godfrey-Smith, P. (2006b). Theories and models in metaphysics. *Harvard Review of Philosophy, 14*(2006), 4–19.
Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford, NY: Oxford University press.
Gopnik, A., & Schulz, L. (Eds.). (2007). *Causal learning: Psychology, philosophy, and computation*. New York, NY: Oxford University Press.
Hall, N. (2010). David Lewis's metaphysics. In: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (*Fall 2010 Edn*). http://plato.stanford.edu/archives/fall2010/entries/lewis-metaphysics/. Accessed 1 April 2011.
Hare, R. M. (1963). *The language of morals*. Oxford, NY: Clarendon Press.
Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy, 106*, 587–612.
Horwich, P. (1990). *Truth*. Oxford, NY: Blackwell.
Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford, NY: Oxford University Press.
Karakostas, V. (2009). Humean supervenience in the light of contemporary science. *Metaphysica, 10*, 1–26.
Knobe, J., & Nichols, S. (Eds.). (2008). *Experimental philosophy*. New York, NY: Oxford University Press.
Kutach, D. (2010). Empirical analyses of causation. In A. Hazlett (Ed.), *New waves in metaphysics*. New York, NY: Palgrave Macmillan.
Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50*, 249–258.
Lewis, D. K. (1983). *Philosophical papers, volume I*. Oxford, NY: Oxford University Press.
Lewis, D. K. (1986). *Philosophical papers, volume II*. Oxford, NY: Oxford University Press.

Lewis, D. K. (1994). Humean supervenience debugged, *Mind* 103 [Reprinted in Lewis (1999), pp. 224–247.
Lewis, D. K. (1996). Elusive knowledge. *Australasian Journal of Philosophy, 74*, 549–567.
Lewis, D. K. (1999). *Papers in metaphysics and epistemology* (pp. 224–247). Cambridge, MA: Cambridge University Press.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Harmondsmouth, NY: Penguin.
Mackie, J. L. (1980). *The cement of the universe* (2nd ed.). Oxford, NY: Oxford University Press.
Marshall, D. (forthcoming). Geometry and nature.
Maudlin, T. (2007). *The metaphysics within physics*. Oxford, NY: Oxford University Press.
Menzies, A. (2007). Causation in context. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's Republic Revisited* (pp. 191–223). Oxford, NY: Oxford University Press.
Paul, L. A. (forthcoming). The Handmaiden's tale: Ontological methodology.
Price, H. (2003). Truth as convenient friction. *Journal of Philosophy, 100*, 167–190.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
Russell, B. (1917). On the notion of cause. *Proceedings of the Aristotelian Society, 13*, 1–26.
Russell, B. (1918/1985). *The Philosophy of Logical Atomism*. In: D. Pears (Ed.). Chicago, IL: Open Court, 1985.
Stanley, J. (2005). *Knowledge and practical interests*. Oxford, NY: Oxford University Press.
Swanson, E. (2010). Lessons from the context sensitivity of causal talk. *Journal of Philosophy, 107*, 221–242.
Weatherson, B. (2010). David Lewis. In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (*Summer 2010 edn*). http://plato.stanford.edu/archives/sum2010/entries/david-lewis/. Accessed 1 April 2011.
Weisberg, M. (2007). Who is a modeler? *British Journal for the Philosophy of Science, 58*, 207–233.