

# Going meta on mental causation

Anthony Dardis

August 24, 2010

## Abstract

Explaining how mental causation could be possible in a physical world is a matter of explaining how creatures who have minds could cause things—perform actions—in virtue of having their mental characteristics. The mental causation problem thus is up to its ears in problematic metaphysics: properties, causes, laws of nature, physicalism . . . How could we make progress, though, in the face of so much current doubt and anxiety about whether or how metaphysics is substantive? This paper argues for a neo-Quinean approach to at least some metaphysical problems: doing metaphysics is inferring the best explanation of everything. I'll use a sample proposal about how to solve the mental causation problem as the target, my “superset” strategy, according to which mental properties are proper supersets of physical properties.

## 1 Introduction

In (Dardis, 2008) I offer a solution to the mental causation problem. In (Dardis, 2009) I discuss a general metaphysical worry that might be raised about my solution: it appears to depend on neoHumean accounts (essentially those of David Lewis) of laws and of properties and their relation to causation. I claim that the structure of my solution does not actually depend on neoHumean metaphysics, and can be used in non-Humean frameworks. I support this claim by showing that the structure of my solution could be used just about as easily in a theory that uses dispositions and a process account of causation.

Not all is smooth sailing, however. The structure itself may seem outlandish to someone with robust metaphysical sensibilities; indeed, the idea that we could be ecumenical with respect to a choice between neoHumeanism and non-Humean accounts of properties, powers and causation may also seem outlandish.

In this paper I want to try to explain my ecumenical attitude by stepping back and asking what a metaphysical theory of properties is supposed to be doing. I'm stepping squarely into the contemporary “metametaphysics” debate (Chalmers et al., 2009). My own preferred metametaphysics is once again neo-Humean: it's a kind of update of Quine's metametaphysics, and not unlike some proposals of Stephen Yablo's (Yablo, 2006, 2009).

In the first couple of sections of this paper, I'll describe the mental causation problem that troubles me the most, and a sketch of the solution I favor, and the way it might work for non-Humean ontologies. Then I'll go on to describe the metametaphysical framework I favor and show how it works for the line of thinking I've described about the mental causation problem.

## 2 My mental causation problem

The mental causation problem that concerns me is the one that Jaegwon Kim has been pressing since the late 1980's (Kim, 1989, 1998, 2005). We start with a few assumptions about the basic character of the world, human beings, and minds:

- **Physicalism:** every actual concrete individual is made up of nothing but physical matter. “Physical” means: referred to by an ideal completed theory of what makes things happen in the natural world.
- **Closure:** physics is closed in the sense that for anything that happens, it has a purely physical cause, if it has a cause at all.
- **Non-identity:** mental properties are not identical with physical properties.<sup>1</sup>
- **Supervenience:** for each mental property, there is a set of physical properties such that (a) any instance of the mental property has one of the physical properties, and (b) anything that has any of those physical properties has the mental property. This dependency holds with at least nomological necessity but not with logical necessity (otherwise the mental property would be identical with the disjunction of the physical properties).
- **Causal relevance:** when one event causes another, the cause has properties in virtue of which it causes the effect. Events are “coarse-grained”, that is, they are individuals that have many properties. Given an effect, some of the cause's properties matter to the existence and character of the effect, others do not. Call this relation on properties **causal relevance**.<sup>2</sup>
- **Exclusion:** “No single event can have more than one sufficient cause occurring at any given time—unless it is a genuine case of causal overdetermination” (Kim, 2005, 42)

---

<sup>1</sup>Although much of the current conversation is about whether mental properties reduce to physical properties, what matters to Kim's arguments is non-identity. I favor a modal argument to their non-identity: since non-physical minds are logically possible, the property of being a mind cannot be a physical property.

<sup>2</sup>Kim's formulations of the problem use a “fine-grained” conception of events, on which two events are identical iff they share the same constitutive object, time and property. There is some discussion in the literature about whether the mental causation problem depends in some way on which conception of events we adopt. I will assume that the problem can be expressed using either conception of events, and use the “coarse-grained” conception.

We then argue as follows. Assume (for *reductio*) that a mental event causes some effect in virtue of having some mental property: the mental property is causally relevant to some properties of an effect. By Supervenience, the cause and the effect have physical properties that necessitate their mental properties. By Closure, the cause has causally relevant physical properties. By Non-identity, these are not identical to any mental properties of the cause. By Exclusion, if the physical properties are causally relevant, the mental ones are not. So the mental properties of the cause are not, after all, causally relevant.

### 3 My solution to this problem

**Negative response** Kim’s various formulations of the Exclusion principle are set up for a case in which there are two putative causes of one effect. I want to take “two” here very seriously. Suppose that the mental property and the physical property are not entirely distinct. Then they aren’t two things. By Non-identity, they aren’t identical either, so they aren’t one thing or two things. (See (Sanford, 2005) for a nice discussion of the distinction between non-identity and distinctness.) There is a number of ways this might be possible.<sup>3</sup> Probably the best known are Yablo’s “coincidence” view (Yablo, 1992) and Shoemaker’s “subset” strategy (Shoemaker, 2007); see also Watkins (2002), Clapp (2001), and Shoemaker (1998).

My own solution is the “superset” strategy. My Lewisian theory of properties—a version of what Armstrong calls “class nominalism”—says that each set of actual and/or possible individuals represents a property. By Supervenience, the actual extension of a mental property is typically larger than the extension of any one of the physical realizers; by my claim that non-physical thinkers are possible, mental properties are modally larger than physical properties as well. Hence mental properties are proper supersets of physical properties.<sup>4</sup>

So the negative response to the problem is that since mental properties are supersets of physical properties, they are not independent of them; moreover, an event’s having both a mental property and a physical property is not a matter of its effect having “more than one sufficient cause”. Hence the Exclusion principle doesn’t apply, and the argument that the mental is causally irrelevant doesn’t go through.

**Positive response** This negative response is unlikely fully to remove the sense that Kim’s argument taps into something very problematic about the possibility

---

<sup>3</sup>(Harbecke, 2008, 165) defines “new compatibilism” about the mental causation problem as a family of solutions to the problem that deny the identity of mental things and physical things, and at the same time deny their distinctness. Harbecke’s own view is a refinement of Yablo’s. Bennett (2003) also uses the expression “compatibilism” to pick out this kind of view. I believe Terry Horgan coined the expression in (Horgan, 1997).

<sup>4</sup>Put more carefully: sets that represent mental properties are proper supersets of sets that represent physical properties. I say that sets *represent* properties because of familiar difficulties about saying that the properties *just are* the sets; but for ease of exposition I’ll write as if the properties were the sets themselves.

of mental causation. One way to express it is that we have a “compactness” intuition: in seeking out the real truth about what makes things happen, wherever we can find a theory that is broader and deeper (explains more things in more fundamental terms), we are inclined to prefer that theory over whatever it explains. But this is a remark about explanation and epistemology and us, even if it does appear to have strong metaphysical echoes. It’s unclear, for example, whether we are still distracted by the “competition” picture of selecting the candidate cause that best meets the tests set by Mill’s Methods and their sophisticated progeny. Since where we have supervenience the supervening property isn’t independent of the subvening property, there is no question of competition. (Shapiro and Sober (2007) exploit this point and Baltimore (2010) criticizes their argument.)

So perhaps we could have a metaphysical picture on which there is causal relevance at many levels. Since (assuming that higher levels supervene on lower levels) the higher levels aren’t independent of the lower levels, they don’t make up “more than one” cause. Equally, there is no question that higher levels “add more oomph” to what is happening at the lower levels.<sup>5</sup>

Suppose, then, that one property stands in the causal relevance relation to another only if there is a law of nature that relates them.<sup>6</sup> A law of nature is a simplest strongest theory of everything. We think—most likely Lewis thought—of these laws as fundamental. But as Lewis points out (Lewis, 1983) we can build up a theory of everything around other than predicates for fundamental properties. If we take the predicates of a well-confirmed theory at a higher level, say, psychology, we can build up a total theory of the world around those predicates. We can draw on more fundamental theory in two places (see (Dardis, 2008, 157-162) for details): (1) higher level predicates may be multiply realized, and so will have to be supplemented with predicates that select the realizations; (2) theories at higher levels only talk about parts of the world, so the total theory just imports the bits of a more fundamental theory that talk about the rest of the world. Then we say that two properties stand in the causal relevance level if they are related by a law of nature, either at the fundamental level or from one of these higher levels.

A non-Humean can make use of the same “overlap” structure (Dardis, 2009). Let a mental disposition overlap a physical realizer disposition: they are not identical, but they are not fully distinct either. Dispositions cause by some sort of necessitation. Both the mental and the physical disposition have this kind of necessitation relation to the effect. The key thing is to remember that since the dispositions overlap, the mental disposition doesn’t *add* any “causal oomph”.

Well, if, on both pictures, the mental doesn’t *add* anything, why count it as causally relevant? Because the “making happen” relation (lawlikeness, for the

---

<sup>5</sup>Thus my solution is not a form of emergentism; I follow McLaughlin (1992) in holding that if a causal power is emergent then it makes a contribution that is not already made by the lower levels.

<sup>6</sup>These are supposed to be laws of succession, which take structured arrays of properties and yield a property: given a distribution of property instances  $ABP\dots$ , there will be an instance of property  $Q$ .

Lewis picture, and disposition-necessitation, for the dispositionalist) does hold, both at the lower levels and at the higher levels. And, while all the nomologically possible mental things are physical things as well, the mental properties are not physical properties. So there is distinctive causal relevance there, even though it isn't anything "extra".<sup>7</sup>

## 4 What's going on?

There are at least three metaphysical issues here that will seem objectionable.

1. if there are no properties, then there is no problem;
2. if there are properties, then it is not just some sort of "choice" whether they are, say, dispositions as opposed to classes;
3. is it ok to say that properties "overlap," that two properties can be neither identical nor distinct? why is it ok?

My responses to these questions go like this:

1. There is no stable philosophical consensus on the existence of properties. There are empiricist and nominalist intuitions that incline philosophers to say they don't exist. There are indispensability arguments to show that they must (van Inwagen, 2004). The situation is very much like that regarding numbers. I don't see a way to stop talking about properties, even though equally I don't think they are any kind of thing in their own right. The entire situation inclines me to think that the issue *isn't factual*. Rather, it's a matter of something like a decision. We talk about properties (numbers) because inquiry is impossibly cumbersome if we avoid talking about them. Whatever the outcome of the metaphysical and metametaphysical conversations, we're going to go on talking about properties. And (see below, §8) there is good reason to think that talking about the mental properties of mental events *is* in a certain sense factual, even if there are no properties.
2. To the extent that a proponent of dispositions or causal powers etc. can show that her theory has degrees of freedom where the properties theory lacks them, and we need those degrees of freedom, then to that extent we have a good reason to adopt the dispositional or causal powers etc.

---

<sup>7</sup>Another way of expressing this point goes like this: the higher levels (since they are connected with robust, well-confirmed sciences) pick out a distinctive pattern of "making happen" in the natural world. We should retain that much of the "special sciences" idea: geology, for example, picks out something that really happens, and that real fact is not captured by the vast miscellaneous collection of physical ways things can be geological objects. The problem with expressing the point this way is that it comes close to being objectionably epistemological: this real fact is interesting to us (we care about mountains and rocks) but since it isn't really anything other than just that vast miscellaneous collection, why should it be thought of as a real part of the world?

ontology. In that sense there is clearly nothing like a “choice”. But if either theory can assert the same things and deny the same things, then there’s no point in arguing about which is right. It’s a matter of taste, ease of exposition, economy and elegance of expression. Since I doubt that any of these competing accounts of properties, powers, etc., have a decisive advantage, then I think we are left with a choice.

3. The point of a metaphysical theory is to set out or set up a framework within which inquiry may proceed. It’s up to us how to set it up. The test of a good framework is how well it conduces to good results in inquiry. Since inquiry is constrained by how the world actually is, a good framework is one that permits the generation of good theories about what the world is like. If nothing stands in the way of saying that properties overlap, and saying that they do overlap helps generate good theories about what there is, then to that extent it’s ok to say that properties overlap.

These responses are no doubt familiar. I want to spend some time articulating them and contrasting them with alternatives.

## 5 Quine’s “On what there is”

What are we doing when we do metaphysics? Clearly not ordinary empirical enquiry (“philosophers organize expedition to Paris to discover length of meter bar”).

We might be doing conceptual analysis (Thomasson, 2009) or meaning analysis of the words we use (‘property,’ ‘object,’ ‘cause,’ etc.) but on the face of it this doesn’t look plausible. A nominalist about numbers isn’t analyzing “the” meaning of the word ‘number;’ rather she is constructing something that would do the job that numbers do, acceptable to her ontological scruples. Do tables exist in addition to persons? Conceptual analysis doesn’t look like the right tool to decide this question. Nor does it seem satisfactory to say that if the meanings of the words ‘table’ and ‘person’ don’t settle an answer to that question, then that’s all there is to know or say.

We might be just making up languages (cf. Hirsch (2009) on the endurantism/perdurantism debate). This seems quite seriously to underplay the difficulty of doing first-order metaphysics and its point. Philosophers are trying to get clear on what the world is like. It would be a fine thing if it was essentially provable that different metaphysical setups all really “say the same thing” (Church’s and Turing’s theses about computability, for example). But it doesn’t seem to be provable, and Hirsch’s claim that they do say the same thing depends quite heavily on a very substantive set of ontological claims (Hawthorne, 2009). Hirsch is articulating one version of Carnap’s critique of metaphysics (Carnap, 1956): Carnap’s view also seems to me to give a picture of difference between metaphysical theories that really trivializes it too much, and leaves no room for any account of the difference other than “a practical decision”. Typically it takes considerable effort to get clear on how exactly two metaphysical theories

are related, so it can't be a matter of a "mere" decision to pick one rather than the other. Also, metaphysical theories usually literally contradict one another ("there are tables" v. "there are no tables") and so it seems unlikely that they all "say the same thing" at some appropriate level.

Perhaps we are investigating the facts about the genuine constituents of the world, or the basic constituents of reality (Fine, 2009). Or perhaps we are investigating what is fundamental and the ways it grounds what is less fundamental (Schaffer, 2009). This certainly seems closer to what metaphysicians have thought they were doing, and what they have wanted from their theories.

What do "basic constituents of reality" or "fundamental" mean? Fine and Schaffer take these as *primitive*, not to be explained in other terms (cf. Lewis and Sider on "natural" (Lewis, 1983; Sider, 2009)). It's not fair to complain we have *no idea* what the expressions mean, or that taking them as primitive makes the theory "esoteric," available and understandable only to initiates (Hofweber, 2009). We do know roughly what those expressions mean. But *simply* taking them as primitive certainly invites the charge. Something needs to be said about what "taking as primitive" is supposed to mean.

Quine's metametaphysics (Quine, 1948) is one way to explain this. Quine takes the problem of ontology to be to say what there is. The method of ontology goes like this: by empirical inquiry, find the best theory of everything; by reflection on language and argument, find the best logic; translate the best theory into the best logic, paraphrasing for economy and in order to eliminate unnecessary entities. Assume your best theory of everything is true, and identify what things there are in order for it to be true. That's what there is.

These "stages" (Schaffer, 2009) aren't really stages, of course. Each operates with the others in view. Logic and mathematics develop together: as we find more sophisticated and more subtle theories in either, typically we find new ways to do things in the other. We need both to do empirical inquiry, and empirical inquiry leads certain kinds of development in math and logic. The strategy of paraphrase (Russell, 1905) developed out of logic and has flourished in the hands of linguists and friends of "semantic methods". So the ontological heavy lifting of Quine's method is happening all the way along. Working out the best theory of everything involves trying to figure out what kinds of things there are, that they might combine together to make up the world. Some of that figuring out is structural, particularly in physics: what kinds of structures do we need in order to make sense of what we're observing? That demands mathematics and sometimes new mathematics. Paraphrasing is more or less continuous with structural "figuring out."

I suggest that Quine's picture provides an account of what it means to "take something as primitive". The point of ontology is to say what there is, not what there could be. That means that it is neither analytic nor known *a priori*. The standard form of argument for a synthetic *a posteriori* theory is inference to the best explanation. An explanation of everything is an explanation that has substantial structure, and using it involves non-trivial mathematical inference. A theory sets up a network of categories for understanding the world, a "picture" or a model of what the world is like (see below, §7). Some of these categories

are basic, others are derived, and the theory specifies how to manipulate propositions expressed using these categories. “Taking something as primitive” then means taking it as a basic theoretical category.

That’s where we should locate the primitive concepts. They are the ones that base all our explanations of other things. They aren’t “first principles” that are known by pure rational insight. They are primitive in the sense that there isn’t anything more basic or fundamental in terms of which we can explicate them. That doesn’t mean that there’s no account whatsoever of what they mean. On the contrary: they get their meaning from their job, and their job is to contribute to a special sort of explanation. Since this kind of explanation is in principle responsive to absolutely everything, the content of the primitive vocabulary depends on the way that the whole theory responds to everything.

## 6 Quine? Really?

Quine is not actually such a popular metaphysician these days, and so much has been said against his metametaphysics. (Although (Yablo, 1998, 259) reveres “it as ontology’s last, best hope” and (van Inwagen, 2009, 506) says “if these [Quinean meta-ontological] ‘rules’ are not followed . . . it is almost certain that many untoward consequences of the disputed positions will be obscured by imprecision and wishful thinking”.)

Fine (2009) takes Quine’s picture to be restricted to ontology as it is done in the service of serious scientific theorizing, whereas in fact most ordinary entities have very little to do with explaining anything (chairs, tables and couples, for example, and numbers).

Schaffer (2009) argues that Quine’s ontological question (“what is there?”) has overwhelmingly trivial answers: there are numbers, properties, propositions, species, fictional characters, proper parts—the atheist isn’t, for example, denying that God *exists* but saying rather that God is a fictional character. The better (and older, and Aristotelian) ontological question is “what grounds what”: what is fundamental, and how are other things dependent on that? Schaffer hammers Quine for presupposing answers to this question at every “stage” of the ontological enterprise.

I think Fine underestimates Quine’s idea of the scope of the relation between inference to the best explanation and existence questions. Quine thinks of *everything* as a “posit”: “the myth of physical objects [is] . . . a device for working a manageable structure into the flux of experience” (Quine, 1951, 41). I think there is considerable truth to this—inference to the best explanation seems like a powerful and pervasive psychological mechanism for minds like ours.

I think Schaffer mischaracterizes the way the question of grounding pervades Quine’s metametaphysics. *Of course* Quine presupposes questions about grounding at every stage. The structure of reality does really involve more and less fundamental things. If ever we figure out how consciousness works, we will (won’t we?) see how it is grounded in some kind of structures which in turn are grounded in neural networks in the brain, which in turn are grounded in cellular

mechanisms, and so on down. The only stage at which the ontology turns out to be entirely “flat” is at the bottom, which by definition is ungrounded and grounds all else. Certainly if we construe Quine’s interest as exclusively in that level, then the grounding questions fade into the background. Indeed, pursued rigorously, Quine’s metametaphysics leads to an implausible sort of Pythagoreanism or structuralism: the important thing about the theory that comes out of ontological investigation is its cardinality and which structures it defines; that’s really all that’s important to the kinds of things there are. We don’t have to follow Quine that far. But it’s not going to help to say, some moves are just not permitted, or that there is a privileged domain, or that some things are primitive. Any of these moves is a move made within a certain kind of inquiry: metaphysics. We need a characterization of what kind of inquiry that is before we can understand what these moves are for. And my suggestion is that at the end of the day, Quine is basically right about the kind of inquiry it is. We’re trying to figure out what there is and how it hangs together. The way we do this is the way we figure anything out: we try to find structure and order in our experience of things. We build a theory.

## 7 Models, maps and measurement

Ok, so the point of metaphysical inquiry is theory building and the method is inference to the best explanation. The mental causation problem leads to questions about the existence of properties and about what we can and cannot say about them.

As I hinted above, I don’t think properties exist, yet I think they are indispensable for theorizing about mental causation. That looks about as close to a contradiction as one might like. So I need to show how a theory, in the sense sketched above, can say this about properties.

The standard way a Quinean argues that some class of things doesn’t exist is by explaining how that class of things is really something else (and, even better, by exposing some more-or-less concealed fault with the idea of the class (Churchland, 1981)). That’s how Field and Chihara argue against numbers: they offend against a certain sense of reality, and we can show how to do without them. The standard Quinean argument *for* a class of things is the indispensability argument: there just is no other way for some stretch of important discourse to come out true unless those things exist. I don’t want to engage in either kind of argument, because the position I want says neither that properties exist, full stop, nor that they don’t, full stop.

Let me say a little more about how I am understanding theories. I’m following the “semantic” view of theories (van Fraassen, 1980). The “syntactic” or logical positivist view took theories to be a set of sentences. The semantic view takes a theory to be a class of models. Sometimes models are real, concrete objects, as for instance Watson and Crick’s model of the DNA molecule. Sometimes they are fictions (Kuhn’s example is the use of the “two spheres” model of the world in practical celestial navigation). Sometimes they are ab-

stract mathematical structures, for instance set-theoretic structures, or perhaps a set of equations, for example for the time-development of a system in some state space. Sometimes models are understood to be the objects theorized about themselves, understood in some stylized or abstracted way.

There is a fair amount of controversy about how models and scientific theories and scientific practice are related (Frigg and Hartmann, 2009). My question concerns theories in metaphysics. What a metametaphysical theory may say about these theories is less constrained by “first-order practice” than what philosophers of science may say about scientific theories. So I want to try to work out an account of how metaphysics, inference to the best explanation, and models and theories could work together.

A theory, as I’m understanding it, specifies a set of models, a set of entry conditions and a set of exit conditions. The entry and exit conditions relate the theory to something else: either things in the world, or perhaps entities described by some other theory. Copernicus’s theory explained the retrograde motion of Mars. The entry conditions are observations of the sky, and yield a state for the heliocentric model. The exit conditions take the state of the model and yield claims about the actual state of things in the sky, and predictions about observations.

The theory’s specification of the model describes, in more or less detail, a set of objects and their relations. One key kind of relation involves the time-development of configurations of objects. At least in the sciences that’s what we’re most interested in: the retrograde motion of Mars is something that happens in time, and so we want the model to display how the objects in the heliocentric model progress through time so as to generate the appearance of retrograde motion.

For any theory and model, there is a question about how the model is supposed to relate to “the world”. Suppose, first, that we have a strongly realist picture of what the world is like. There are things and their properties and relations, and the point of the theory is to tell the truth about those things and their properties and relations. That is, we’re assuming that we know the basic categorical structure of the world, and that the truth about the world will be a specification in terms of that categorical structure. The model of the world offered by the theory may, or may not, be intended to be isomorphic with the world. It will be isomorphic if the objects in the model correspond 1-1 with the objects in the world: for any particular application of the theory, to each real-world object there corresponds an object in the model, and *vice versa*. For each modelled relation on the objects in the model, the real-world objects should stand in the corresponding relation. (Since any real-world object has indefinitely many relations to everything else in the world, it would be too much to ask of a theory that it model *all* the relations of the real-world objects.) A model can fail to be isomorphic, in this sense, in at least 3 ways. First, there may be relations in the model that aren’t intended to match actual relations of the objects. Second, there may be objects in the model that aren’t intended to match real-world objects. Third, there may be real-world objects in the application situation that are ignored. Kuhn’s example of celestial navigation has

objects that aren't intended to match real-world objects (celestial spheres), relations that aren't intended to match (the stars are mounted on those spheres), and many objects in the actual situation that aren't represented in the model.

What if we relax the assumption that we know the categorical structure of the world? We could take the theory in a strongly anti-realist spirit, and say that the theory and the model are a calculation device: the exit conditions only specify predictions about observations, and say nothing about the objects and relations present in the application situation. We could do some careful work picking out the realist from the non-realist components of the model. Indeed we could work out a theory of which bits of theories should be taken in a realist spirit. Ian Hacking's famous remark that electrons are real because you can spray them would be one such theory. Since, in that case, the theory of electrons specifies the sort of rich causal interconnectedness that we find in paradigmatic real-world objects, the theory says that electrons are real. Or we might remain cautiously neutral, in something more than an anti-realist spirit but less than a strongly realist spirit. The theory picks out something real in the world, we might say, but we don't know what it is. Assuming that the theory is strongly confirmed, we can be confident that something is going on in the world that takes the entry conditions to the exit conditions. There may be bits of the model that correspond to bits of the world.

The question of the match between the model and the real world is very complex for theories in the natural science, and a large number of possible kinds of match has been developed and explored. I take it that this work provides resources for articulating subtle pictures of how metaphysical theories and their models relate to the real world.

Let me close this section with two other well-known and well-understood theory/world relations that exemplify the kind of pattern I'm interested in: maps and measurement.

**Maps** The Google Maps map of my address shows a white road, beige lots with grey boundaries, and a blue creek. It shows the road I'm on reaching a corner about 200 feet from my house, where it becomes a different road which leads off in a northwesterly direction.

Of course, the road isn't white, the lots aren't beige, the lot boundaries don't follow the actual lots, and they're legal rather than physical (there are fences of various kinds between the lots, but they aren't grey either), and the creek is, well, a kind of brownish green. (It was bright blue about a month ago after a chemical plant burned upstream.) The road does indeed do what the map depicts it as doing.

The model or representation—the map—uses some medium to represent features of the landscape. Some aspects of the representation are intended to represent aspects of the landscape, and others are not. We use blue as the conventional map color for water, but most maps in everyday use are not intended to represent the color of the water, only its presence.

A map like those served by Google Maps is intended primarily for navigation:

to place locations relative to other locations. So the map of my address (pretty accurately depicts distances and directions. When the scale is 2000ft per inch, the fact that my house is about 2 linear inches from the Philosophy Department shows that my house is about 4000 feet (as the crow flies) from the Philosophy Department. If we want to know how far it is along the roads, Google Maps traces a route suitable for a car and offers up that it is about 1.1 miles. Its representation of which turns to take is also accurate.

**Measurement** This is the heart of measurement (Suppes, Patrick and Zinnes, Joseph, 1963). The world contains various things related in various ways. A measurement system contains a set of objects and their relations: we measure temperature with numbers, and we measure spatial relations in the world with spatial relations on a map. To prove that a measurement system is doing what we want it to, we need to prove a representation theorem and a uniqueness theorem. The representation theorem tells us that the relations we want to measure are isomorphic with the relations in the measuring system. The uniqueness theorem tells us what kinds of transformations of the measuring system preserve the relations we are measuring. So for temperature, a representation theorem will show that something in the world corresponds to increments of temperature in any of the scales; the uniqueness theorem shows us that temperature is measured on an interval scale: you can transform any temperature measurement scale into another that serves just as well, by multiplying by a constant and adding a constant; also, there is an absolute origin. Given the representation and uniqueness theorems, you can say exactly what it is about the world that is being measured, exactly which features of the measurement system represent that, and exactly which features of the measurement system are simply arbitrary, like the conventional map color of water.

## 8 Properties

My proposal about properties, then, is this. We can treat discourse about the properties of things as deriving from a theory about the world. The theory involves a model, which says that there are individuals and their properties. For starters, as Aristotle (1996) could have said, properties are repeatable, in the sense that many individuals (and many properties) can have the same property, but individuals are not repeatable, in that nothing (in this sense of “have”) has an individual.

What else should the theory say? The following is tempting: there are two kinds of things, the individuals, and the properties. Individuals have properties. This looks paradigmatically like a place where we want to *deny* that the model is isomorphic with reality: if we took seriously the idea that real-world individuals have property instances, we have the “pincushion” model of property possession, on which property possession is like pins stuck into a substrate.<sup>8</sup>

---

<sup>8</sup>Or the Christmas orange pomander theory: take an orange and stick cloves in it. When

Armstrong says that universals are immanent and that they are wholly present in each of their instances. The “immanent” part at least entails that if the world contains no instances of  $P$ , then  $P$  is not a universal. One traditional additional idea is that a “transcendental” universal is one that exists (a) even if there are no instances, (b) neither in space nor in time—so it exists in reality in a different way (or something) than spatio-temporal things. Should a theory of properties say (or deny) these extra things about reality? The “wholly present” part at least entails that everything a property can contribute to its instances it does contribute (one of Plato’s example from *Phaedo*: a triple of things is three, and, being three, is numbered oddly). Taken literally, though, the expression is absurd: if being human is a universal, and the universal is wholly present in me, then if I am wholly destroyed, then the universal is wholly destroyed. Once again: does the theory of properties have to say that it’s literally true *of reality* that an individual—the universal—is wholly present in all its instances?

Armstrong notes at the end of (Armstrong, 1989) that the issue between his scientific realist position and Martin’s trope theory is merely terminological. Daly (1997) argues that a universals theory can do and say anything that a tropes theory can do or say, and better and more economically.

What should we do? The history of attempts to come up with a coherent theory of properties is not encouraging.

The semantic account of theories offers a strategy. We are at liberty to put in pretty much what we like as the model for properties. There are a few basic things that have to be preserved: whatever properties are, they are shared among individuals, they are one way to explain sameness, there’s a close tie between the existence of an instance and individuals, there are certain quantificational moves that should come out ok,<sup>9</sup> and so forth. We then have the range of possible attitudes toward realism to decide on. We can say that the theory is intended entirely anti-realistically: there are no such things as properties, the theory is a mere calculation device for keeping track of ways things are. We can take the theory quite realistically: properties enjoy real existence just like individuals, and the job of a theory of properties is to tell us what that kind of existence is like. Or we can take the theory of properties in a neutral way. To the extent that it gets right the way property possession bears on change, sameness, etc., it “picks out something real”. Where exactly the model corresponds to reality we can leave up in the air.

Back to my contradiction: there are no properties, but we have to talk about them (at least my mental causation problem requires talking about them). My own inclination about the theory of properties is anti-realist about the properties

---

it dries, you have an aromatic Christmas tree ornament. The orange is the individual and the cloves are the property instances.

<sup>9</sup> ‘Any two mature, well-formed spiders of the same species have the same anatomical features; *Hence*, An insect that has some of the same anatomical features as some mature, well-formed female spider has some of the same anatomical features as any mature, well-formed female spider of the same species.’

(van Inwagen, 2009, 506)

themselves. Whatever we need to say about properties is handled by the entry and exit conditions of the theory of properties, together with the relations the theory describes for properties. (For example: if I'm looking at three apples, then the entry conditions yield a configuration in the model of three individuals, and the three individuals are an instance of the property of being a triple. Since that property brings along with it the property of oddness, the exit conditions tell us that these three apples have the property of being oddly numbered.) We are going to talk about properties of things *anyway*; property talk offers a semantic mechanism that is far too intertwined in our languages and ways of talking to give up.

And back to my three objectionable metaphysical issues (§4). We can, as I've just argued, say that there are no properties, but talk about properties anyway. We are more or less at liberty to set up the theory of properties as we like—universals, tropes, classes, whatever—so long as we preserve agreed-on aspects of property-talk. That means, for example, that we are at liberty to say that properties (or tropes, or dispositions, or universals) “overlap” in the way needed for my solution to the mental causation problem. The point of a theory of properties is to set out a model of what they are that permits us to set up elegant and productive accounts of how the world works. I submit that property overlap does just that for the mental causation problem. (And it's not an *ad hoc* move: it solves the “everything not basic causation” problem as well: if mountains or tables or cats or stars can cause things in virtue of their higher level properties, then property overlap is a good way for them to do it.)

## 9 Presupposition

Yablo (2009) suggests non-catastrophic presupposition failure as a model for how we may understand apparently (but problematically) referential discourse. “The present King of France isn't wearing my pajamas” is unproblematically true, even though there is no such person. The sentence (perhaps) presupposes that there is such a person; but at least one of its entailments (that of all the people who are wearing my pajamas, the present King of France isn't one of them) is true and does not depend on the existence of said King for its truth.

Suppose something like this works for numbers. The truths of arithmetic turn out to be presupposition-free consequences of claims about numbers. So, do numbers exist? Well, *all there is to know* about whether they do or don't comes from arithmetic. And all there is to know from arithmetic is the many truths of arithmetic. And these, by hypothesis, don't say whether or not numbers exist. Since there's nowhere else to look for an answer to this question, it just has no answer.

I like this idea a lot. I'm not clear yet what its relation is to the semantic conception of theories. Certainly Yablo's motivating idea in ontology is very similar to mine: these are questions we shouldn't be worrying about, not because they have answers, but because answering them is the wrong thing to be trying to do.

## 10 Advantages

### 10.1

Sensible account of methodology of metaphysics. (Contra Thomasson.)

### 10.2

Sensible account of origin of necessity in metaphysics.

### 10.3

Sensible account of disputes about what metaphysics is supposed to be doing (Quine/van Inwagen v. Schaffer/Fine).

### 10.4

Answer to Bennett: there are no facts there.

## 11 Problems

### 11.1

What are models? If models are fictions, we need a metaphysical account of fictions, and so the theory of properties depends on the theory of fictions. I don't know which is more obscure. If models are abstract objects (sets, for example), then we need a metaphysical account of the being of abstract objects.

The cardinality of a theory of properties is probably extremely large (Lewis guesses that the number of properties is at least  $\Omega_3$ ). So a metaphysical theory of properties has a model that has a very high cardinality. So we need an account of how that works.

### 11.2

Many metaphysicians (Hawley is one) are explicit that the method of metaphysics is something like inference to the best explanation. Ladyman and other philosophers of science criticize the metaphysicians. Inference to the best explanation in the natural sciences is very constrained by data and by other theories. IBE in metaphysics is really not constrained by much at all.

Granted. I am urging that the metaphysical disputes of most interest for the mental causation problem—those about laws, properties, powers, etc.—are to a large extent bootless. I do think there are a few definite things to be said about properties, for example, but the rest is really a matter of style.

### 11.3

Inference to the best explanation is a highly suspicious form of inference. Understanding it would require understanding explanation (or: understanding IBE in metaphysics would require understanding what explanation in metaphysics is).

## 12 Conclusion

A lot of work on the mental causation problem over the last 20 years or so has advocated moving out of traditional (Humean, logical positivist and post-positivist) metaphysics, and a lot of exciting new work in philosophy of science and the metaphysics of science has depended on exploring new metaphysical accounts of the categorical landscape. I've argued that we can work out a solution to the problem by adding some degrees of freedom to theories of properties that might seem objectionable. The questions, then, are, what are we doing when we do metaphysics, and how can we tell whether we're doing it correctly or incorrectly? My claim has been that Quine was right, metaphysics is really the same kind of thing as empirical theorizing, and so the test of doing it correctly is whether we get good theories of everything. If I'm right, many questions in metaphysics aren't so much questions about reality but about styles of modelling and theory presentation.

## References

- Aristotle. 1996. "Categories." In *Introductory Readings*, 1–8. Hackett Publishing Company. Translated, with Introduction, Notes, and Glossary, by Terence Irwin and Gail Fine.
- Armstrong, D.M. 1989. *Universals: An Opinionated Introduction*. Boulder, CO: Westview Press.
- Baltimore, Joseph. 2010. "Defending the piggyback principle against Shapiro and Sober's empirical approach." *Synthese* 175:151–168.
- Bennett, Karen. 2003. "Why the Exclusion Problem Seems Intractable, and How, Just Maybe, To Tract It." *Noûs* 37:471–497.
- Carnap, Rudolf. 1950. "Empiricism, Semantics, and Ontology." *Revue internationale de philosophie* 4:20–40.
- . 1956. *Empiricism, Semantics, and Ontology*, chapter Appendix. Chicago, IL: University of Chicago Press. Originally published as Carnap (1950).
- Chalmers, David J, Manley, David, and Wasserman, Ryan (eds.). 2009. *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Clarendon Press.

- Churchland, Paul. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78:67–90.
- Clapp, Lenny. 2001. "Disjunctive Properties: Multiple Realizations." *The Journal of Philosophy* 98:111–36.
- Daly, Chris. 1997. "Tropes." In D. H. Mellor and Alex Oliver (eds.), *Properties*, 140–159. Oxford: Oxford University Press.
- Dardis, Anthony. 2008. *Mental Causation: The Mind-Body Problem*. New York: Columbia University Press.
- . 2009. "A structure for mental causation." <http://philsci-archive.pitt.edu/archive/00004739/>. Presented at Metaphysics of Science conference July 2009, Melbourne.
- Fine, Kit. 2009. "The question of ontology." In Chalmers et al. (2009), 157–177.
- Frigg, Roman and Hartmann, Stephan. 2009. "Models in Science." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2009 edition.
- Harbecke, Jens. 2008. *Mental Causation: Investigating the Mind's Powers in a Natural World*. Frankfurt: ontos verlag.
- Hawthorne, John. 2009. "Superficialism in Ontology." In Chalmers et al. (2009), 213–230.
- Hirsch, Eli. 2009. "Ontology and alternative languages." In Chalmers et al. (2009), 231–259.
- Hofweber, Thomas. 2009. "Ambitious, Yet Modest, Metaphysics." In Chalmers et al. (2009), 260–289.
- Horgan, Terence. 1997. "Kim on Mental Causation and Causal Exclusion." *Philosophical Perspectives 11: Mind, Causation, and World* 165–184.
- Kim, Jaegwon. 1989. "The Myth of Non-reductive Physicalism." *Proceedings and Addresses of The American Philosophical Association* 63:31–47.
- . 1998. *Mind in a Physical World*. The MIT Press.
- . 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- Lewis, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61:343–377.
- McLaughlin, Brian P. 1992. "The Rise and Fall of British Emergentism." In Ansgar Beckerman, Hans Flohr, and Jaegwon Kim (eds.), *Emergence or Reduction?* Berlin: De Gruyter.

- Quine, Willard van Orman. 1948. "On what there is." *Review of Metaphysics* 2:21–38.
- . 1951. "Two dogmas of empiricism." *The Journal of Philosophy* 60:20–43.
- Russell, Bertrand. 1905. "On denoting." *Mind* 14:479–493.
- Sanford, David. 2005. "Distinctness and non-identity." *Analysis* 65:269–74.
- Schaffer, Jonathan. 2009. "On What Grounds What." In Chalmers et al. (2009), 347–383.
- Shapiro, Larry and Sober, Elliot. 2007. "Epiphenomenalism — The Do's and the Don'ts." In Gereon Wolters and Peter Machamer (eds.), *Thinking about Causes: From Greek Philosophy to Modern Physics*. Pittsburgh: University of Pittsburgh Press.
- Shoemaker, Sydney. 1998. "Realization and Mental Causation." *Proceedings of the Twentieth World Congress of Philosophy* 9:23–33.
- . 2007. *Physical Realization*. Oxford: Oxford University Press.
- Sider, Theodore. 2009. "Ontological Realism." In Chalmers et al. (2009), 384–423.
- Suppes, Patrick and Zinnes, Joseph. 1963. "Basic Measurement Theory." In R. Duncan Luce, Robert R. Bush, and Eugene Galanter (eds.), *Handbook of Mathematical Psychology*, volume I, 1–76. New York: Wiley and Sons.
- Thomasson, Amie. 2009. "Answerable questions and unanswerable questions." In Chalmers et al. (2009), 444–471.
- van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Clarendon Press.
- van Inwagen, Peter. 2004. "A theory of properties." In Dean Zimmerman (ed.), *Oxford Studies in Metaphysics*, volume 1, 107–138. Oxford: Oxford University Press.
- . 2009. "Being, Existence, and Ontological Commitment." In Chalmers et al. (2009), 472–506.
- Watkins, Michael. 2002. *Rediscovering Colors: A Study in Pollyanna Realism*, volume 88 of *Philosophical Studies*. Dordrecht: Kluwer Academic Publishers.
- Yablo, Stephen. 1992. "Mental Causation." *The Philosophical Review* 101:245–280.
- . 1998. "Does ontology rest on a mistake?" *Proceedings of the Aristotelian Society - Supplementary* 72:229–261.

- . 2006. “Non-catastrophic Presupposition Failure.” In Judith Thomson and Alex Byrne (eds.), *Content and Modality: Themes from the Philosophy of Robert Stalnaker*, 164–190. Oxford: Clarendon Press.
- . 2009. “Must Existence-questions Have Answers?” In Chalmers et al. (2009), 507–525.