Psychological Explanation, Ontological Commitment, and the Semantic view of Theories

Colin Klein

1 Introduction

Naturalistic philosophers of mind must assume some philosophy of science. For naturalism demands that we look to psychology—but to be guided by psychological theories, one must have some story about what theories are and how they work. In this way, philosophy of mind was subtly guided by philosophy of science. For the past forty years, mainstream philosophy of mind has implicitly endorsed the so-called 'received' or 'axiomatic' view of theories. On such a view, theories are sets of sentences formulated in first-order predicate logic, explanations are deductions from the theories, and the ontology of a theory can be read off from the predicates used in explanations.

The persistence of the received view in in philosophy of mind is surprising, given that few philosophers of science these days would endorse the it. An alternative, the so-called *semantic* view of theories, has become far more popular. With it comes a new view about

explanation, and about ontological commitment more generally. One might therefore worry—as I do—that many problems in philosophy of mind are actually psuedoproblems introduced by an outdated notion of theories.

Philosophy of mind has seen some important moves beyond the axiomatic view and the corresponding view of explanation in recent years (Craver (2007) is a notable example). I think, however, that philosophy of mind — and especially the metaphysics of mind — has not fully appreciated how different the landscape looks when one moves away from the old view of theories. The new wave in philosophy of mind will involve re-importing some of these lessons from philosophy of science, and re-thinking some of the old puzzles that arose in the context of the axiomatic theory. What follows is a first step in that process, focusing on the key issue of explanation and ontological commitment.

2 Two Views about Explanation

2.1 Explanatory Literalism

Consider the following pairs of explanations:

(1) (a) The square peg failed to pass through the hole because its cross-section was longer than the diameter of the hole

(b) The peg failed to pass through the hole because [*extremely long description of atomic movements*]

(2) (a) Klein got a ticket because he was driving over 60mph

(b) Klein got a ticket because he was driving exactly 73mph

(3) (a) Socrates died because he drank hemlock

(b) Socrates died because he guzzled hemlock

(4) (a) Esther ran because she was scared of the bee

(b) Esther ran because [*complicated neural description*]

Many have the strong intuition that the first sentence in each pair is a *better explanation* that the second. This is true, note, even though the truth of the second sentence guarantees the truth of the first. I want to take that intuition for granted and explore two different stories about why that might be the case.

There is a well-loved account, tracing at least back to Hilary Putnam, for the superiority of some explanations. Explanation 1a, Putnam claimed, is clearly better because

> In this explanation certain *relevant structural features of the situation* are brought out. The geometrical features are brought out. It is *relevant* that a square one inch high is bigger than a circle one inch around. And the relationship between the size and shape of the peg and the size and the shape of the holes is *relevant*. It is *relevant* that both the board and the peg are *rigid* under transportation. And nothing else is relevant. The same explanation will go in any world (whatever the microstructure) in which these higher-level structural features are present. In that sense this *explanation is autonomous* (Putnam, 1975, p296)

On Putnam's story, (1a) refers to a higher-level *property* of the peg, the shape, that is most *commensurate* with the explanandum. Following Yablo, Bontly has called this the 'Goldilocks Principle': the peg's shape is just enough (and no more) to cause its failure to pass; so too with all truly explanatory properties. (2a) is a better explanation because the property of my speed—being above the limit—was sufficient for a ticket; my exact speed was not required. (3a) is a better explanation than (3b) because it was *drinking* hemlock that was fatal, guzzled or not. (4a) is a better explanation than (4b) because the extra

detail is irrelevant: Esther would have run no matter how her fear was instantiated.

Generality does not always make for better explanation. Consider:

(4) (c) Esther ran because she was scared of the small flying thing

This is both true and more general than (4a); nevertheless, it is an inferior explanation if Esther is scared only of bees but indifferent to flies. Rather, it is *proportionality* between higher-level cause and effect that picks out the most explanatory of the causally relevant properties.

Call someone who adopts this view a *literalist* about explanatory goodness. Literalism says that good explanations are superior to rivals because they pick out a property that their rivals don't, and that this property bears the right sort of relationship to the *explanans*. Our best explanations are thus ontologically committing. If a term φ appears in the best explanation of some phenomena, then we are, all things being equal, justified in believing that φ refers to some unique property. Hence the term 'literalism': one can read off the ontological commitments of a good explanation largely by taking it literally, and supposing that each term φ really is meant to refer to a corresponding property or entity.[1]

The literalist view is widely accepted in philosophy of mind. It has been a particular

[1] Note that one could hold a much stricter version of literalism, on which a predicate is ontologically committing only if it is *ineliminable*, or just in case it appears in the best overall axiomatization of the phenomena. I do not focus on these formulations for two reasons. First, in practice nobody actually adheres to this standard, because figuring out whether a predicate is ineliminable or part of the best axiomation *tout court* is too difficult a task. If we held ourselves to such a high standard, the game would be up from the beginning: no one should have confidence that their predicates refer, and so literalism would be a straw man. Second, formulations of the requirement in terms of axiomitization or the eliminability of predicates is so obviously derived from the axiomatic view of theories that the considerations presented in section 4 will apply directly. Thanks to Mark Sprevak for pressing me on this point.

comfort to nonreductive physicalists. The fact that (4b) is inferior to (4a) suggests that even were psychology to be reduced to neuroscience, the resulting neural explanations would be inferior to the psychological ones because they would no longer refer to the most commensurate high-level properties. Further, the explanatory superiority of proportionate properties might lead us to suppose that we have a solution to the hoary *causal exclusion argument*. The causal exclusion argument says, in simplified terms, that mental and physical properties must (if distinct) compete for causal influence, and that a plausible physicalism should force us to assign causal priority to the physical one. Not so, literalism responds: both properties are causally relevant, but only the higher-level one counts as *the* cause. It does so because it is more proportionate, or commensurate with, or otherwise better fitted to the effect. Not only is the exclusion argument avoided, but the mental is given a certain causal priority over the physical. Nonreductive physicalism is saved. For this reason, various forms of literalism are increasingly popular in philosophy of mind and philosophy of neuroscience.

Finally, literalism is simply assumed as uncontroversial by many philosophers of mind. The alternatives to literalism seem to be some sort of anti-naturalism or scientific anti-realism, neither of which are particularly attractive. That alone seems to be reason to accept it.

2.2 Explanatory Agnosticism

Literalism is not the only account of explanatory goodness available to the naturalist, however. For each pair above, it is possible to account for the superiority of one of the explanations by appealing to facts about the *language* in which the explanations are couched while remaining provisionally neutral about the *ontology* one is thereby committed to. Call this *agnosticism* about explanatory goodness. The agnostic denies that we can move easily from

language to ontology. Crudely put, the fact that a certain predicate appears in a superior explanation is no reason to believe that there is a property corresponding to that predicate.

I want to defend agnosticism about higher-level properties. Note that the position I favor is properly agnostic, rather than skeptical. I don't want to take a stand on whether there *are* higher-level properties (or determinables, or whatever). Maybe there are. Maybe there aren't. Rather, my claim is that in ordinary and scientific explanation, apparent reference to higher-level properties carries demands no ontological commitment to the existence of such properties. There may well be higher-level causes; I just don't think that our best explanations are a good guide to what they are.

Agnosticism also has a certain prima facie plausibility. First, many have noted that shifts in the presumed interests of a listener can make a difference in the explanations that it is appropriate to give (van Fraassen, 1980). Consider the explanation:

(5) Socrates died because he angered the Athenians

In certain contexts (historical/political ones say), explanation (5) is superior to either (3a) or (3b); in other contexts (physiological/medical ones), the reverse is true. Yet presumably the facts about what properties are involved and their commensurability remain unchanged.

A defense of agnosticism is strengthened by reflection on conversational pragmatics and their role in shaping our intuitions about explanations. We find the more general explanations of the pair more acceptable, says the agnostic, because of pragmatic constraints on the descriptive form of explanations (and not because they refer to more commensurate properties). These pragmatic constraints—and in particular, the Gricean maxims that underly cooperative conversation—may favor a more general description of the same circumstance, but that description is not superior because it picks out a more general *property*.

A few quick examples for how this might look. (2a) is superior to (2b) because the Gricean maxim of Relevance tells me to give only such information as is relevant to my hearer's interests (Grice, 1989). My wife wants to know why I got another ticket; the fact that I broke the speed limit is sufficient to satisfy her interests, and the specific speed is (we assume) irrelevant to her interests in the conversation. Similarly, the Gricean maxims of Quantity and Quality should forbid me from giving (4b) as an explanation when the equally effective and much shorter description (4a) is available. Indeed, to give (4b) would (on the assumption that I'm being cooperative) produce several false implicatures: that the extra detail is relevant, in the sense that counterfactuals involving small changes to Esther's neural state would result in her calm, or that I have good evidence in a particular case for the complicated neural process at which I have hinted. Neither of these is likely to be true. So to give (4b) would be misleading, in the sense that I would implicate something false to my listener. Nevertheless, it is true that the complicated neural process was *the* cause of her flight, not some additional distinct higher-level property.

Changing conversational demands can produce shifts in explanatory goodness without shifts in interest. The patrolman is testifying. The judge, like my wife, wants to know why I got a ticket. The patrolman would have ticketed me for any speed above 60; if my speed instantiated a higher-level property of *having a speed above 60mph* that was most proportionate to my ticketing before, it continues to do so now. Yet it would now be more appropriate for the patrolman to utter (2b) than (2a). Why? The patrolman's testimony must justify his ticket-giving. Uttering (2b) implies that he has precise information about my speed—which is to say that he determined my speed by some suitably precise measurement. To utter (2a) would give the false implication that he doesn't have such information (since by

the maxim of Relevance he should be as specific as necessary for the demands of the conversation). This implication is cancelable ("He was going over 60mph—in fact, I clocked him at exactly 73mph"), but in ordinary circumstances the patrolman can achieve his ends through the parsimonious (2b).

So here is a general strategy for the agnostic: concede that the second explanations in each pair above are superior, but explain that superiority by appeal to language and conversational context, not the world. Thomas Bontly has argued (convincingly in my opinion) that the implicatures of many causal claims are nondetachable and cancelable, the standard marks of conversational implicatures.[2] Antecedents of the strategy might be found in Kim's insistence that there are only higher-level predicates, not higher-level properties (1998), and in Lewis' remarks on the pragmatics of causal explanation (1986).

2.3 The Plan

These above cases are not, to be sure, knockdown. The literalist has a ready response to them: he can say that the explanations cite facts that are causally relevant, and that shifts in context alter which causally relevant factors are appropriate to cite. But we should be suspicious of this: the evidence for literalism above was supposed to be our judgments about the appropriateness or inappropriateness of single explanations. That confidence should be undermined if we find serious context-sensitive effects on our judgments of appropriateness.

I think that agnosticism can be given a further defense. So the next section will give an

[2] See especially (Bontly, 2005, p.343). I am indebted to Bontly's article for prompting many of the reflections in this section.

extended argument in favor of agnosticism over literalism in the particular case of higher-level causal properties. The overall form of the argument is as follows: There is a set *S* of intuitions that favor the proportionality argument for higher-level causes. *S* is primarily constituted by our judgments about the examples at the beginning of section 2.1 and those like them. I claim that the pragmatics of explanation are such that we would have *S regardless* of whether there are higher-level causes or not. So the fact that we have *S* can't be part of an argument for higher-level causes.

Further, there are some more specific reasons to think that literalism itself is problematic. In particular, it is clear that there are certain predicates that are simply shorthand placeholders for functions defined in terms of other quantities. There are, I claim, good reason to treat such predicates as non-referring; even more strongly, there is no positive benefit to treating them as referring to properties. Yet literalism demands that we do so, which is a mark against literalism.

After the defense, I'll turn to diagnosis. Literalism, I'll argue, is plausible mainly because philosophers of mind are mostly wedded to a bad old sort of philosophy of science, one left over from the late positivists. I'll argue that the plausibility of literalism vanishes if we move to an updated philosophy of science. With that move, we in turn have new resources to deal with, and dissolve, philosophical puzzles that presuppose literalism.

3 Agnosticism and Derived Quantities

3.1 Derived Quantities

Working psychologists, when faced with a good explanation, can still wonder whether it is ontologically committing. When we look at the sciences relevant to philosophy of mind—psychology, cognitive science, and neuroscience, at least—we find that there is often considerable debate about whether a term used in this or that explanation actually refers to a causal property. In a classic textbook on psychometrics, for example, Nunnally warns that

> It is not necessarily the case that all the terms used to describe people are matched by measurable attributes—e.g., ego strength, extrasensory perception, and dogmatism. Another possibility is that a measure may concern a mixture of attributes rather than only one attribute. This frequently occurs in questionnaire measures of "adjustment," which tend to contain items relating to a number of separable attributes. Although such conglomerate measures sometimes are partly justifiable on practical grounds, the use of such conglomerate measures offers a poor foundation for psychological science. (Nunnally, 1967, p.3)

Consider the second possibility mentioned, that of 'conglomerate measures.' Some predicate P might have good predictive power, be measurable in straightforward ways, and appear in good explanations. Yet P may not correspond to a real property because it is simply a label that aggregates over several different psychological attributes. In short, P may be a *derived quantity*—a label for a function of other, more basic properties.

The problem of derived quantities has been overlooked by philosophers of mind because most of the explanations we tend to consider are toy examples that connect two simple events under ideal circumstances. The primary criterion for acceptability in simple explanations is simply that the *explanans* be described in the simplest, most informative way. These simple re-descriptions look a lot like the attribution of higher-level properties, and that in turn goes a long way to explaining the plausibility of literalism. That plausibility vanishes when we move to more realistic scientific explanations. So I'd like to look in depth at a case

from neuroscience to explain just why derived quantities are problematic for literalism.

3.2 The Problem for Literalism

The Hodgkin-Huxley model of the action potential has received renewed attention from philosophers of neuroscience. Hodgkin and Huxley showed that the changes in membrane potential of the neuron are determined by G_{Na} and G_k, functions that determine the sodium and potassium conductance, respectively, as a function of membrane potential. Briefly: the membrane potential is a function of the differential concentrations of Na^+ and K^+ ions on either side of the neural membrane. The membrane is studded with channels that open at an overall rate dependent on the membrane potential; the opening and closing of these channels in turn changes the membrane potential by changing the relative concentration of those ions. Hodgkin and Huxley's experimental determination of G_{Na} and G_k allows accurate derivation of the shape and amplitude of the action potential; it is a great triumph in that regard.

One thing that Hodgkin and Huxley's work explained was the fact that action potentials are threshold phenomena: membrane potential is stable below a certain threshold but rapidly depolarizes above it. We can explain this by noting that:

(6) Below the threshold membrane potential $G_{Na}/G_K = 1$, and so small depolarizations result in offsetting Na and K currents. Above the threshold, $G_{Na}/G_K > 1$, which results in a net Na current with positive feedback.

(6) is a testament to the explanatory fertility of the Hodgkin-Huxley model. Not only does it explain the threshold phenomena in action potentials, but implies a number of useful, testable, true counterfactuals (for example, that action potentials would not be generated if G_{Na}/G_K could be artificially pegged to ≤ 1, as it is by certain toxins.) Further, by parallel with explanations (1a) and (1b), it is arguably a better explanation than one that goes into the

details of the opening of sodium channels, and for the same reason: it gives us precisely the information needed to explain the threshold and no more. Further, the details of the mechanism wouldn't add anything to (6)'s goodness. This is not because the details aren't causally important—they are—but rather because (6) has already told us all we need to know about those mechanisms. Like the other good explanations above, (6) implies precisely the right sorts of counterfactuals, in the right way, and so on.[3]

Suppose we do think that (6) explains why neurons fire in an all-or-nothing way. The literalist faces a dilemma. He could say that the expression 'G_{Na}/G_k' does not itself designate a property—that it only stands for a mathematical operation defined over the determinate value of two distinct properties. But we could as a matter of convention introduce a singular term (say φ) to stand in for G_{Na}/G_k. φ would be a derived quantity. Since G_{Na}/G_k did not designate a property, φ should not either. But that is to concede the main claim of the agnostic view: that one cannot unproblematically move from the

[3] Here, some care is needed. It has become recently fashionable to claim that the Hodgkin-Huxley equation does not explain anything, but merely describes the shape of the action potential (Craver, 2007, Ch3). It is true that insofar as the above is explanatory, it is not because it constitutes a deduction from the more general laws postulated by Hodgkin and Huxley. Rather, (6) is explanatory because it details some facts about the mechanism that underlies the action potential, and then uses facts about that mechanism to explain the threshold. It does not detail the mechanisms by which the voltage-gated ion channels work; to the extent that the detailing of those mechanisms was part of neuroscientists' shared explanatory interests, Hodgkin and Huxley fell short of explaining everything there was to explain about the action potential. But that does not mean that the equations they experimentally derived were not themselves explanatory of *some* phenomena. Thanks to Carl Craver for helpful discussion on this point.

presence of a singular term to a causal property, even in our best explanations.

On the other hand, the literalist could say that $G_{N\,a}/G_k$ designates a distinct, higher-order causal property. This is implausible for at least three reasons. First, it is an unnatural reading of (6): the most natural reading of is as expressing a relationship between $G_{N\,a}$ and G_k. This reading connects the explanation to other explanations in terms of $G_{N\,a}$ and G_k. (For example, we can explain the refractory period of the neuron by talking about the time-sensitive decay of $G_{N\,a}$.) The connection between this explanation and (6) is lost, or at least obscured, if we think that there are two distinct properties involved in the threshold and the refractory period.

Second, the mathematical form of the *explanans* is important: the mathematical properties of ratios can be used, along with the mathematical properties of facts about $G_{N\,a}$ and G_k, to explain further facts about the shape of the action potential. Treating $G_{N\,a}/G_k$ as a single property again obscures this explanatorily useful relationship.

Third, treating $G_{N\,a}/G_k$ as designating a property leads to an unreasonable proliferation of causal properties between which the literalist can offer no ground for decision. For if $G_{N\,a}/G_k$ designates a property, then so should $2(G_{N\,a})/2(G_k)$, $3(G_{N\,a})/3(G_k)$,... Each of these properties are causally commensurate with the threshold effect, since in each case the action potential fires iff the property had a value ≥ 1. There is nothing, from the point of view of causal facts, to distinguish them. This proliferation is a *reductio* against literalism.

Of course, scientists might prefer to use the unadorned $G_{N\,a}/G_k$ to its multiples. This is no argument, however; indeed, it's rather uncomfortable for the literalist. For surely what exists doesn't depend on what people prefer to talk about. So the expression $2(G_{N\,a})/2(G_k)$ either refers or not. If it does refer, we have an explosion; if it doesn't, I fail to

see an argument for why it doesn't refer that doesn't also impugn $G_{N\,a}/G_k$ itself.

Insofar as (6) is preferable to explanations in terms of, say, $2(G_{N\,a})/2(G_k)$, it is for pragmatic rather than causal reasons. The more complex formulation would be inappropriate because it implies that the extra complexity is somehow relevant. By the maxims of quality and relevance, we should prefer to give a simpler, shorter, less complex explanation if it would suffice. That's what's makes (6) better than other, mathematically equivalent counterparts.

So either way the literalist treats $G_{N\,a}/G_k$, he must say that the quality of some explanations lies in how they describe a set of causal properties, not just that they describe causal properties. But that is to concede the agnostic's main point.

4 A Diagnosis

What's the lesson from all of this? One could, I suppose, use it to defend a crude sort of old-fashioned reductionism. That is, one could argue that all mental predicates are simply derived quantities, and that the only real causal properties are the physical ones and the properties that are identical to them. (Indeed, much of the above was inspired by Kim's remarks about second-order descriptions in science in chapter four of his (1998), and could be thought of as one way of unpacking them.) I think, though, that we can draw another, deeper conclusion. The real question is why literalism seems so *plausible* even if it's problematic, especially to naturalistically-minded philosophers of mind. Here, I think I can offer a diagnosis.

4.1 Literalism and the Axiomatic View

Literalism's plausibility has a historical origin. Many classic papers in metaphysics of mind developed against the background of the late positivist conception of theories as

developed in the writings of Carl Hempel and Ernest Nagel.[4] This is sometimes called the 'received view' or 'standard view' of theories (though that is now an anachronism). I'll call it the *axiomatic* view of theories, because on it theories are conceptualized as the best axiomatizations of a domain of phenomena.

On the axiomatic view, a theory consists of two parts. The first part consists of a set of theoretical postulates : a finite set of sentences, constructed from a basic vocabulary containing a fixed set of names and predicates, and augmented with the resources of the first-order predicate calculus. Speaking loosely, the predicates in the standard vocabulary are the properties and relations that the theory attributes to the world. The universally quantified statements among the theoretical postulates are the laws of a theory. The laws, together with statements of particular fact, allow us to derive particular consequences that predict and explain phenomena.

The second aspect of theories is a set of coordinating definitions, which supply a semantics for the theory by connecting at least some of the terms in the basic vocabulary to the world. By the time of Hempel, it was widely agreed that this connection would not involve an exhaustive characterization of theoretical terms via observational terms. Instead, in Hempel's formulation—later imported into philosophy of mind by Lewis (1970)—coordinating definitions link theoretical terms to other terms we already have a handle on, often because they occur in natural language. The coordinating definitions, together with the interrelations between terms described by the theoretical postulates, provide a partial interpretation of the

[4] See Nagel (1961) for a classic statement, and Suppe (1989) for a contemporary reconstruction and discussion.

theoretical terms. This partial interpretation allows us to connect the predictions of the theory to the world, and so to give our theories empirical content.

If you endorse this view of theories, then literalism and its conclusions are nearly inevitable. Theories are individuated by the languages they use. A different language just gives you a different, and therefore competing, ontology. Assuming that this different language is not simply reducible to the original, then there really are two sets of properties in the world that compete for the title of the most explanatory.

4.2 The Semantic View of Theories

The axiomatic view is no longer popular among philosophers of science. It fell out of favor for a number of reasons. [5] Two in particular are worth noting. First, as Suppes notes, first--order formulations of theories are inadequate for many scientific purposes. Any theory that requires, say, the real numbers will be difficult to capture in first-order language. Further, axiomatizing both the theory and the accompanying math would be, in Suppes' words, "awkward and unduly laborious" (Suppes, 1967, p.58). By this, I take it that Suppes means that even if we can axiomatize the relevant math, it would be inappropriate to include mathematical apparatuses in the theory itself—certainly it is more natural to describe set theory as something that we use to talk *about* various theories, not something that happens to be part of

[5] See chapter 2 of Suppe (1989) for an extended discussion of problems with the axiomatic account. The essays in Salmon (1998a), especially Salmon (1998b), also contains a number of useful critiques of the deductive-nomological view of explanation associated with the axiomatic view.

many distinct theories.

Second, the axiomatic view requires theories to be axiomatizable. Theories that can be axiomatized turn out to be rare, and theories that are actually treated as a set of axioms rarer still. This was bad enough in disciplines like biology and psychology, where it was hard to find things that counted as laws. But it seemed to be true even of physics: as van Fraassen notes, many useful treatments of quantum mechanics are non-axiomatic in form (1970). Even if we are confident that theories could be identified with sets of axioms, then, it seems like a stretch to claim that the axiomatic view has captured how *scientists* treat theories.

From these criticisms, an alternative naturally follows. The *semantic view* of theories claims that theories are to be identified with sets of models, rather than sets of sentences. These models are real structures—abstract entities like sets or state-spaces in Suppes and van Fraassen, concrete objects in more recent treatments.[6] These structures are meant to be isomorphic to the world in some respect. Theoretical models are often *described* using language, but the important linkages hold between models and the world, not between any canonical description and the world. So on the semantic view, a theory consists of two parts: a set of models, and a postulation of isomorphism between certain respects of models and parts of the world.

The semantic view seems to fit better with scientific practice; many disciplines present models of some target phenomenon and then reason about them. This is most obvious in fields

[6] For the latter see (Giere, 1988; Godfrey-Smith, 2006). I prefer concretist accounts, both because I find them more natural for sciences like psychology and also for the problems recently raised by Hans Halvorson against more mathematically-oriented approaches in his (2012).

like cognitive psychology. Models of facial recognition, say, are never presented as sets of laws. Instead, one is presented with a model mechanism and an assertion that this is what the brain does—that is, an assertion that the brain is isomorphic to the model in some respect. Similarly, as Lloyd has shown, many of the central claims of evolutionary theory can be interpreted as models of systems under selection (1994). Newtonian mechanics can be interpreted as the postulation of certain models, the permissible Newtonian spaces (van Fraassen, 1970). And so on.

The semantic view is problematic for literalism. On the semantic view, there is no presumption that the *language* in which theories are designated is at all important. The same set of models can be described using a variety of different terms, none of which need pick out the driving causal properties in a model (van Fraassen, 1989, Ch9). As a simple example, Hodgkin and Huxley could be thought of as specifying a state-space for neural processes. Later work on the molecular configuration of sodium and potassium channels described the same state-space using the language of molecular biology. Same models, same theory, completely different language. Again, literalism is unwarranted. Similarly, the relationship between model and world need not be exact: model-world mappings can be inexact, fuzzy, or otherwise complex (Godfrey-Smith, 2006). So the mere fact that there is an element in a model does not warrant concluding that there is an isomorphic property or object in the world: that depends, at the very least, on the intended model-world mapping.[7]

In addition to fitting the apparent practice of science, the semantic view also provides a neat solution to the role of mathematics in science. Mathematics is something we use to reason *about* the models. Mathematics is not a part of any theory, but is available to all. Thus,

[7] Thanks to Dan Weiskopf for drawing my attention to this point.

$_{max}$ III II

as van Fraassen puts it, physics first sets up a framework of models and then, having done so, "The theoretical reasoning of the physicist is viewed as ordinary mathematical reasoning concerning this framework" (van Fraassen, 1970, p.338).

With that in mind, consider, mathematically complex claims like the Hodgkin-Huxley equation, or mathematically complex expressions like the one describing $G_{N\,a}$,

$$G_{N\,a} = g_{N\,a} \quad ^{3}$$

where *m* and *h* themselves stand for complex exponential functions governing activation and inactivation of the sodium channel. It would be a mug's game to try to recast any of these in a first-order language. If you can't, then the received view forces you to treat things like ratios, products, and multi-variable embedded functions as causal properties. As we saw in section 3.2, this isn't a very plausible reading of explanations like (6). Further, recasting (6) this way would be a futile exercise: you can keep your ontology trim by including only the individual properties in (6) along with math.

Once we move to a view on which scientific theories are not artificially hampered in their expressive power, something like agnosticism is forced upon us. On the received view, there is one best way to state the content of an explanation, because there are so few ways to express anything. On a semantic view, by contrast, one has the possibility of talking about models in a variety of different ways. When that happens, one will need to take pragmatic factors into account when we evaluate the goodness of explanations. Figuring out the ontological commitments of an explanation is a complicated, hermeneutic process, not a straightforward leap from terms to the world.

5 Going Further

Literalism ultimately relies on an unrealistically simple view about how scientific theories work. Attention to the pragmatic aspects of explanation shows several reasons why this simple view must be abandoned. Good explanations often involve abstract re-descriptions of specific, lower-order properties; these re-descriptions are required for pragmatic reasons, not for ontological ones. This in turn fits well with the semantic view of theories, which carefully separates the language in which models are specified from the models themselves and the model-world relationships asserted by the theory.

I want to conclude by considering ways in which the abandonment of literalism might matter for philosophy of mind. I have argued elsewhere that once we break the link between theories and the language in which they are formulated, traditional arguments for multiple realizability fail.[8] This is because traditional arguments for multiple realizability suppose that the only explanations available to physics are those that describe atoms and their motions in tedious detail. The idea that physics only describes mereological simples is almost unavoidable on the axiomatic view, for reasons outlined above. It is also patently absurd: physicists spend most of their time trying to give high-level abstract explanations of physical phenomena. Once we realize this, and the role of model re-description in science, multiple realizability becomes difficult to motivate.

Indeed, I think there's a more general point that can be made here about the individuation of scientific disciplines. There has been an assumption that scientific disciplines

[8] I develop this point further in (Klein, forthcoming). See also my (2009) for an earlier exploration of this idea in the context of Nagel's theory of reduction.

are individuated by their *domains*: that is, what's characteristic about physics or biology is primarily the set of things that fall under their laws. This view is again almost unavoidable on the axiomatic view: the domain of a science just is the domain of its quantifiers. This turn leads to the a hierarchical, striated view of reality made famous by Oppenheim and Putnam (1958). On such a view, each scientific discipline corresponds to a distinct level of reality. Again, a metaphysical point grows out of a substantive view about philosophy of science. If my argument is right, however, we should be wary of this view of the world. Sciences have more descriptive flexibility than the philosopher of mind tends to ascribe to them, and there is no reason why scientific disciplines must carve the world into non-overlapping spheres of influence.

The semantic view of theories permits an alternative view of disciplinary individuation: what I'll call (with some trepidation) a *paradigm-based* view. Every discipline or sub-discipline starts with a set of characteristic phenomena that it tries to explain: living things for biology, minds for psychology, nerves for neuroscience, lenses for optics, and so on. The investigation of characteristic phenomena often hinges on creating local levels—again, it's scientifically useful to abstract, to decompose, to divide things up by size, and to look at the behavior of aggregates and compounds.

This is makes the standard level-based view of the world problematic, for two reasons. First, there's no guarantee that some sciences, when decomposing things into their parts won't run into another science that cares about aggregates (or vice versa). Often, these distinct subdisciplines bump into each other: seeking to explain the behavior of life, biologists decompose living things into cells, and cells into organelles, and organelles into their parts. At that point, it bumps into chemistry, which has been investigating the same

phenomena as a special case of some more general abstract principles. That's not, note, due to some overarching commitment to a 'unity of science' program: this is normal science within one discipline extended until it—as a matter of contingent, empirical fact—hits normal science that started with a different set of characteristic phenomena. Sometimes when this happens there is a complete merger—as, for example, when the science of lenses came to be swallowed up to become a special branch of physics. Other times the merger is tentative or incomplete, as it currently is with biochemistry or cognitive neuroscience. These mergers should, in my opinion, be counted as forms of intertheoretic reduction. But note that the picture of reduction that emerges will not be an imperialistic one: there is not the science of one level of being co-opting a distinct one. Instead, insofar as disciplines evolve and merge, it is an outgrowth of perfectly ordinary *intra*theoretic endeavors on each side.

Second, many sciences care about making models at a relatively high level of abstraction. The same model of oscillatory motion turns out to be useful both for the investigation of springs and for the vibrations of electrons. Again, this is one of the things that physics is good at: taking the behavior of a specific set of things, and showing that at some level of abstraction it is the behavior of a diverse set of things. This sort of abstraction is, as I conceive of it, intra-theoretic: it is part of ordinary scientific practice within a discipline. But models formulated at that level of abstraction also often turn out to have unexpected uses in other domains: modeling (say) electrical circuits, or the oscillatory firing of neurons. In these cases, it's natural (and, note, actual) that models from one science get imported into another, largely unchanged. But this again makes problems for hierarchical concepts of nature.

In conclusion, the shift from an axiomatic to a semantic view of theories should result

in a shift in how naturalistically-inclined philosophers approach scientific language. The very same theory can be couched in different language, and even canonical formulations of a theory can hide considerable complexity in the real-world properties to which a model corresponds.

Fodor was once able to write confidently that "Roughly, psychological states are what the terms in psychological theories denote if the theories are true." (Fodor, 1997, p.162 fn1). Moving to the semantic view of theories should sap that confidence. With new humility, however, also comes new opportunities for close reading of scientific theories, and a more engaged approach to determining the ontology to which psychological explanations actually commit us.[9]

[9] Thanks to Carl Craver, David Hilbert, Esther Klein, Tom Polger, and several APA audiences for comments on previous drafts. Special thanks to participants in the New Waves online conference organized by Mark Sprevak and Jesper Kallestrup for many helpful comments.

References

K. Bennett (2003) ‘ Why the exclusion problem seems intractable, and how, just maybe, to tract it’ *Nous*, 37(3), 471–497.

T. Bontly (2005) ‘ Proportionality, causation, and exclusion’ *Philosophia*, 32(1), 331–348.

C. Craver (2007*) Explaining the brain*. (New York: Oxford University Press).

J. Fodor (1997) ‘ Special sciences: Still autonomous after all these years’ *Philosophical perspectives: Mind, Causation, and World*, 11, 149–163.

R.N. Giere (1988) *Explaining Science: A Cognitive Approach* (Chicago: The University of Chicago Press).

P. Godfrey-Smith (2006) ‘ The strategy of model-based science’ *Biology and Philosophy*, 21, 725–740.

H.P. Grice (1989) *Studies in the Way of Words* (Cambridge: Harvard University Press).

H. Halvorson (2012) ‘ What scientific theories could not be’ *Philosophy of Science*, 79, 183–206.

J. Kim (1998) *Mind in a Physical World* (Cambridge: MIT Press).

C. Klein (2009) ‘ Reduction without reductionism: A defence of Nagel on connectability’ *The Philosophical Quarterly*, 59(234), 39–53.

C. Klein (Forthcoming) ‘ Multiple realizability and the semantic view of theories’ *Philosophical Studies*. DOI: 10.1007/s11098-011-9839-6.

S.W. Kuffler, J.G. Nicholls, and A.R. Martin. (1984) *From Neuron to Brain: A Cellular Approach to the Function of the Nervous System*, 2nd edn (Sunderland: Sinauer Associates, Inc.)

D. Lewis (1970) ‘ How to define theoretical terms’ *The Journal of Philosophy*, 63(13), 427–446.

D. Lewis (1986) ‘ Causal explanation’ In *Philosophical Papers*, volume 2. (New York, Oxford University Press).

E. Lloyd (1994) *The Structure and Confirmation of Evolutionary Theory* (Princeton: Princeton University Press).

E. Nagel (1961) *The Structure of Science: Problems in the Logic of Scientific Explanation* (New York: Harcourt, Brace, and World, Inc.).

J. Nunnally (1967) *Psychometric theory* McGraw-Hill, New York.

P. Oppenheim and H. Putnam (1958) 'Unity of science as a working hypothesis' *Minnesota Studies in the Philosophy of Science*, 2, 3–36.

H. Putnam (1975) 'Philosophy and our mental life' In *Mind, Language, and Reality* (London, Cambridge University Press).

W. Salmon (1998a) *Causality and Explanation* (New York, Oxford University Press)

W. Salmon (1998b) 'Deductivism visited and revisited' In Salmon (1998a), pages 142–177.

F. Suppe (1989) *The semantic conception of theories and scientific realism* (Champaign: University of Illinois Press).

P. Suppes (1967) 'What is a scientific theory?' In S. Morgenbesser (ed.) *Philosophy of Science Today* (New York: Basic Books)

B. van Fraassen (1970) 'On the extension of Beth's semantics of physical theories' *Philosophy of Science*, 37(3), 325–338.

B. van Fraassen (1980) *The Scientific Image* (New York: Oxford University Press)

B. van Fraassen (1989) *Laws and Symmetry* (New York: Oxford University Press)

S. Yablo (1992) ‘Mental causation’ *The Philosophical Review*, 101(2), 245–280.