The Philosophical Review 116.3 (July 2007): 401–425

The Problem with Charlie:
Some Remarks on Putnam, Lewis and Williams[*]

Timothy Bays

In his new paper, "Eligibility and Inscrutability," J. R. G. Williams presents a surprising new challenge to David Lewis' theory of interpretation. Although Williams frames this challenge primarily as a response to Lewis' criticisms of Putnam's model-theoretic argument, the challenge itself goes to the heart of Lewis' own account of interpretation. Further, and leaving Lewis' project aside for a moment, Williams' argument highlights some important—and some fairly general—points concerning the relationship between model theory and semantic determinacy.

In these remarks, I plan to do three things. First, I'll provide a brief overview of the three arguments that I'll be looking at: Putnam's original model-theoretic argument, Lewis' response to this argument, and Williams' new model-theoretic argument. In giving this overview, I'm going to suppress almost all of the model-theoretic details, so that the arguments' philosophical structures come out more clearly. For similar reasons, I'll feel free to shamelessly oversimplify the arguments. Where this oversimplification becomes *too* shameless, I'll try to redeem my scholarly credentials in the footnotes.

Second, I'll lay out what I take to be the most obvious objection to Williams' new argument, and I'll then look at several ways of modifying Williams' model theory so as to overcome this objection. Finally, I'll step back and make a few remarks concerning the broader philosophical significance of model-theoretic arguments of the type given by Putnam and (now) by Williams.

1 Three Arguments

Let's start with Putnam's original argument. The *goal* of Putnam's model-theoretic argument is to show that our language is semantically indeterminate—that there's nothing about the way we use our language which forces that language to take on a unique "intended interpretation." So, for instance, there's nothing about our use of the term "cat" which forces it to pick out all and only the cats, and there's nothing about our use of names like "Puffy" and "Fluffy" which forces them to pick out particular cats determinately.

In arguing for this conclusion, Putnam starts with two assumptions. First, he assumes that our language is essentially first-order and that our best overall theory of the world can be captured by a collection of first-order sentences. Call this collection of sentences "T." Second, Putnam assumes that giving an interpretation to our language simply amounts to finding a first-order model which satisfies—in the model-theoretic sense

[*] This paper was originally presented as part of a symposium on J. R. G. Williams' paper, "Eligibility and Inscrutability," which was held at Cornell University in April of 2006. I would like to thank the other participants in that symposium—especially J. R. G. Williams, John Hawthorne and Zoltan Szabo—for helpful comments on this material.

of "satisfies"—all of the sentences in T.[1] Given this, the model-theoretic argument simply consists of using basic theorems of model theory to show that there are, in fact, many different models which satisfy all of the sentences in T. Hence, Putnam claims, there's no *unique* interpretation of our language.

Now, for our purposes, the technical details of Putnam's model theory don't really matter (though, see footnote 5 for a few of these details). Suppose, therefore, that we put this model theory in a big black box. You give Putnam a formalized version of your best theory of the world. Putnam takes your theory, puts it in his box, turns a crank, model-theory happens, and a bunch of models come tumbling out of the box.

These models have three nice properties. First, there are many such models (that's what will eventually ensure that there's no unique intended interpretation of your language). Second, some of these models are quite peculiar. There is a model which uses the term "cat" to pick out a miscellaneous collection of dogs, a model which uses "cat" to pick out a collection of protons, and still another which uses "cat" to pick out a collection of planets. Similarly, there is a model which uses "Puffy" to name a particular dog, a model which uses "Puffy" to name a particular proton, and still another which uses "Puffy" to name a particular planet.[2] Finally, and most importantly, each of these models satisfies all of the sentences in your best overall theory of the world—i.e., all of the sentences in T.

Given all this, Putnam asks the following question: why don't all of these models count as intended interpretations of your language? Granted there are a lot of them, and granted that some of them are quite peculiar. Still, each of them satisfies all of the sentences in your best overall theory of the world, and what more could you ask from an intended interpretation of your language?[3]

[1] So, here's the first place where I'm doing a bit of oversimplifying. There are two things to mention. First, some versions of Putnam's argument work for non-first-order theories—e.g., for second-order theories or for theories formulated in some sublanguage of $\mathcal{L}_{\omega,\omega}$. Still, there are some real restrictions here: the arguments don't work for languages governed by the causal theory of reference and they probably don't work for languages with modal operators. See [5] (chapter 11) for a discussion of the causal case and [8] (section 4) for an interesting analysis of the modal case.

Second, Putnam doesn't simply *assume* than every model of T constitutes an intended interpretation of our language. This is something he argues for—e.g., in his now (in)famous just-more-theory argument. For our purposes, however, this later aspect of Putnam's project isn't really at issue, so I'm going to leave the just-more-theory argument unexplained.

[2] These examples highlight an important point about the model-theoretic argument—namely, that it's intended to establish a fairly *radical* form of semantic indeterminacy. Some arguments for indeterminacy turn on cases of genuine metaphysical controversy and focus on a fairly limited form of indeterminacy. So, for instance, the fact that metaphysicians disagree about whether cats should be thought of as four-dimensional space-time worms or as objects which exist (in their entirety) at many different times might lead us to question whether there's anything about our actual use of the term "cat" which could determine which of these two things the term refers to. In this case, genuine indecision about the metaphysics of cats might motivate an argument for (limited) semantic indeterminacy—i.e., for indeterminacy vis-a-vis the several *metaphysically plausible* candidates for the reference of the term "cat."

In contrast, Putnam's model-theoretic arguments purport to show that *almost any word* can refer of *almost any object and/or property*—e.g., that "cat" may well pick out certain cherries, that "dog" may well pick out certain protons, and that the name "Puffy" may well refer to the moon. It's this kind of radical indeterminacy which makes Putnam's model-theoretic argument so interesting (and, I would argue, so philosophically provoking).

[3] Some references are probably in order here. Canonical formulations of the model theoretic argument can be found in [9], [10], and [11]. The arguments in [9] and [10] apply to ordinary languages and use permutation theorems for their model theory.

This, then, is the model-theoretic argument in a nutshell. Let's turn to David Lewis' response to this argument. Lewis' response turns on a single key idea: some models are *just better* than other models. To understand this idea, we need to ask two questions. First, how do we determine the quality of a model—i.e., what makes one model better than another model? Second, how does this ranking on models interact with the details of Putnam's model-theoretic argument? I'll take these two questions in order.

First, Lewis' answer to our first question comes in two stages. He begins by noting that, for the purposes of physical explanation, some properties are more fundamental than others. So, for instance, the property of being a quark is more fundamental than the property of being a proton, since the properties and behaviors of quarks *determine* the properties and behaviors of protons, while the properties and behaviors of protons don't usually determine the properties and behaviors of quarks. For similar reasons, the property of being a proton is more fundamental than the property of being a molecule, the property of being a molecule is more fundamental than the property of being a cat, and the property of being a cat is more fundamental than the property of being a member of some random collection of quarks, protons, molecules, cats and planets.

Next, Lewis argues that it's a *good thing* for a model to use fundamental properties and relations to interpret the predicates and relation symbols in our language. Suppose that our language contains a unary predicate, P, and suppose also that we have three models: $\mathbb{M}_1, \mathbb{M}_2$, and $\mathbb{M}_3$. $\mathbb{M}_1$ uses P to pick out the collection of quarks, $\mathbb{M}_2$ uses P to pick out the collection of protons, and $\mathbb{M}_3$ uses P to pick out the collection of cats. Then, all other things being equal, $\mathbb{M}_1$ is a better model than $\mathbb{M}_2$ and $\mathbb{M}_3$, and $\mathbb{M}_2$ is a better model than $\mathbb{M}_3$. That's because $\mathbb{M}_1$ uses a more fundamental property to interpret P than than $\mathbb{M}_2$ and $\mathbb{M}_3$ do, and $\mathbb{M}_2$ uses a more fundamental property to interpret P than than $\mathbb{M}_3$ does.[4]

This brings us to our second question: how does this way of ranking of models interact with the details of Putnam's model-theoretic argument? Once again, there are two things to say about this. First, Lewis' ranking helps to explain how there could, in principle, be a *unique* intended interpretation of our language. Suppose that, out of all the many models which satisfy our best theory of the world, one of these models turns out to be uniquely the best model—i.e., "best" on Lewis' fundamental ranking. Then, according to Lewis, this model will count as the intended interpretation of our language. This gives Lewis a non-arbitrary way of explaining how one particular model can be singled out as *the* intended model of our language.

Second, Lewis' ranking helps to explain why the models generated in Putnam's box *can't be* the intended models for our language. If you look at the model theory in Putnam's box, you'll notice that the models it generates are almost guaranteed—simply in virtue of the way they are constructed—to be *bad models* on Lewis' fundamental ranking. (For convenience, I'll call any model theory which has this particular property

These are the arguments which are most relevant to the present discussion. (I'll say a bit more about the technical details of these arguments in footnote 5.) The argument in [11] is unusual in that it focuses primarily on mathematical language and uses the Löwenheim-Skolem theorem for its model theory. I will say almost nothing about this later argument here. For Putnam's more recent thoughts on his model-theoretic argument, see the comments in [12] and [13].

[4] As it stands, this description may make Lewis' position seem a bit *ad hoc.* In his new paper, though, Williams does a nice job of showing that Lewis' ranking on models actually flows quite naturally from his deeper views on, e.g., properties, universals and the nature of physical explanation. See section 2 of Williams' paper for a more thorough discussion of this matter.

bad-making model theory.) Given the way Lewis singles out intended models, therefore, the models generated in Putnam's box can't be the intended models of our language.[5]

This, then, is the gist of Lewis' response to Putnam's argument. Lewis starts by arguing that some interpretations of our language are more natural—or more fundamental—than other interpretations. He then suggests that the intended interpretation of our language is simply the most natural interpretation which satisfies our best overall theory of the world. Finally, Lewis argues that it's very unlikely that the models constructed in Putnam's model-theoretic argument will turn out to *be* the most natural interpretations of our language. Hence, contrary to what Putnam himself suggests, these models can't constitute intended interpretations of our language.

So much then for Putnam and Lewis. Let's turn Williams' new paper, [14]. The first thing to notice is that Williams' response to Lewis involves, not so much defending Putnam's original model-theoretic argument, as simply replacing that argument with a new one of his own. In particular, Williams' argument turns on replacing all of the model theory in Putnam's back box with some new model theory of his own. (To use the formal jargon, Williams takes *permutation theorems* out of the box, and puts *Henkin constructions* in.)

[5] Let me say a little more about the technical details of this argument. The argument to which Lewis is responding starts by assuming that we can treat the actual world as a *model* for our theory T. That is, we can regard ordinary objects as constituting the domain of our model, and we can then fix a set-theoretic interpretation function which "connects" our language to this domain so as to mimic the ordinary relations of reference and predication. So, for instance, we set $i(\text{“Puffy”}) = \text{ Puffy}$, $i(\text{“cat”}) = \{x \mid x \text{ is a cat }\}$, $i(\text{“dog”}) = \{x \mid x \text{ is a dog }\}$, etc.

Next, we notice that any permutation of our domain carries with it a canonical method for redefining our interpretation function so that the resulting model 1.) has the same domain as our original model, 2.) satisfies the same sentences as our original model, but 3.) has a radically different interpretation function from our original model. So, for instance, let ϕ be a permutation of our domain. Then we can define i' by setting:

$i'(c) = \phi(i(c))$ for any name c,
$i'(P) = \{\phi(x) \mid x \in i(P)\}$ for any unary predicate P,
$i'(R) = \{\langle \phi(x), \phi(y)\rangle \mid \langle x, y\rangle \in i(R)\}$ for any two-place relation R,
etc., etc.

In particular, then, suppose that our permutation switches Spot the dog with Puffy the cat. Then the model induced by this permutation will use "Spot" to name Puffy and "Puffy" to name Spot; it will also think that Spot falls under the predicate "is a cat" and Puffy under the predicate "is a dog."

Now, what Lewis notices is this. A property like being a cat doesn't rank very highly in Lewis' fundamental ordering, since it would be very complicated to define this property in terms of fundamental physical properties—i.e., to define it "from the quarks up." But the property which our new model uses to interpret the predicate "is a cat" does even worse on Lewis' ordering. To define this new property, we would need to understand *both* the ordinary property of being a cat *and* the permutation function, ϕ. Hence, this new property is more complicated than the original property, and so it is even less eligible to serve as the intended interpretation of the predicate "is a cat."

In short: because the properties which Putnam's new models use to interpret our language are *parasitic* on the properties with which we ordinarily interpret our language—i.e., parasitic via the function ϕ—these new properties are ineligible to serve as the intended interpretations of our predicates/relations. Hence, Putnam's models can't be the intended models of our language. For a more detailed discussion of this argument—and for some important hedges and qualifications—see section 2 of Williams' new paper. For Lewis' own development of the argument, see [7] (which rests on [6]).

Now, just as before, the details of these new constructions don't matter very much.[6] Instead, we can simply focus on two, purely-philosophical consequences of Williams' new model theory. First, Williams' model theory isn't *bad-making.* That is, there's nothing about this model theory which makes it obvious that the models it generates will be more complicated—and so less eligible—than the ordinary intended model of our language. Hence, one of Lewis' responses to Putnam's version of the model-theoretic argument simply doesn't work against Williams' new version of the argument.[7]

Second, and more importantly, Williams argues that in at least some cases Lewis' ranking actually picks out the *wrong* model for our language—i.e., the wrong model from an intuitive perspective. To follow this argument—which is, I think, the most interesting argument in Williams' paper—we need to understand a few technical facts about Williams' models. So, suppose that Williams turns the crank on his black box and generates a model for our theory T. Since I want to talk about this model at some length, it's useful to give it a name: I'll call this model "Charlie." Charlie has four nice properties.

First, just as in Putnam's argument, Charlie satisfies all of the sentences in our best overall theory of the world—i.e., all of the sentences in T. Second, Charlie is a finite model—there are only finitely many objects in Charlie's domain. Now, I should note here that these first two properties depend on the assumption that T itself isn't committed to the existence of infinitely many things. So, T doesn't say that there are infinitely many cats or infinitely many planets; nor (and this will be crucial later on) does T say that there are infinitely many abstract objects—say, infinitely many sets or infinitely many propositions.[8] If we are willing to grant Williams these assumptions, though, then he can indeed generate a finite model which satisfies all of the sentences in T.

Third, Charlie is a *numerical* model. The objects in Charlie's domain are simply the natural numbers $1, \ldots, n$ for some finite n, and the properties and relations which Charlie uses to interpret the predicates and relation symbols in our language are number-theoretic properties and relations. Some of these properties and relations may be quite natural—e.g., x is even or x is prime—while others may be more awkward and disjunctive—e.g., x is either 2, or 3, or 27, or 32, or etc. Still, however complicated these properties may be, they remain perfectly good number-theoretic properties.

Finally, the fact that Charlie is both finite and numerical allows us to obtain nice *bounds* on the definitional complexity of the number-theoretic properties and relations which Charlie uses to interpret our language. Consider, for instance, the one-place predicate "cat." In defining the property which Charlie uses to interpret this predicate—a property which I'll call "NCat" for "numerical cat"—the worst thing we'll have to do is to

[6] For more on these details, see sections 3 and 4 (esp. 3) of Williams' paper.

[7] To be more precise, the properties and relations which Williams' models use to interpret our language aren't defined *in terms of* the properties and relations with which we ordinarily interpret our language. Hence, to use the terminology from footnote 5, Williams' models aren't *parasitic* on the intended model of our language. As a result, we can't use a simple term-by-term examination of the ways properties and relations are defined to ensure that Williams' models rank poorly in Lewis' fundamental order.

[8] For Williams own discussion of this assumption see pp. 25 and 30 of his paper.

give straightforward disjunctive definition of the form:

$$x \text{ is an NCat } \iff x = 2 \vee x = 3 \vee x = 27 \vee \ldots.$$

Since Charlie contains only n elements, this disjunction will have at most n disjuncts.[9] Similarly, worst thing we'll have to do in defining a two-place relation is to give a disjunction with n^2 disjuncts; for a three-place relation, we'll need n^3 disjuncts; for a four-place relations, n^4 disjuncts; etc.; etc.

These, then, are four technical fact about Charlie. Given these facts, Williams proceeds to make two, somewhat more philosophical, assumptions. First, Charlie exists in every possible world. Even though we have constructed Charlie in the actual world, the fact that Charlie is abstract ensures that Charlie exists in every other possible world. Second, the *complexity* of properties like NCat—i.e., their rank in Lewis' fundamental ordering—doesn't change as we move from possible world to possible world. As we've just seen, properties like NCat can be given uniform definitions in terms of basic number-theoretic properties. So, if we simply assume that these basic properties are perfectly natural—i.e., that they count as fundamental properties in every possible world—then properties like NCat will have also have a fixed place in Lewis' overall ranking.[10]

Finally, Williams notices that Lewis' theory allows the complexity of ordinary *physical* properties to vary as we move from one possible world to another. Suppose that in our world quarks are the most fundamental physical entities. Then a property like being a cat won't rank very highly in Lewis' ordering, since it would be very complicated to define this property in terms of fundamental physical properties—i.e., to define it "from the quarks up." But now consider a possible world that's just like ours, except that it has extra layers of subatomic particles which live "below" the quarks—say, layers of sub-quarks, sub-sub-quarks and sub-sub-sub-quarks. In *that* possible world, the property of being a cat would be even worse than it is here, since the definition of cats would have to pass through several extra layers of subatomic particles—i.e., it would have to be a definition "from the sub-sub-sub-quarks up."[11]

[9] If we work in a language which contains only the non-logical symbols 1 and $'$ (for successor), then a simple computation shows that this definition will contain at most $\frac{n(n+7)}{2} - 1$ symbols. Similar computations provide nice bounds for n-ary relations.

[10] A few comments on this argument are probably in order. First, Williams measures the complexity of a property syntactically—by looking at the minimal length of a definition of that property using only terms referring to fundamental properties as primitives. In the case of a property like NCat, therefore, the assumption that basic number-theoretic properties are fundamental, combined with the fact that we have a uniform way of defining NCat in terms of such basic properties, is what gives us a stable bound on the complexity of NCat.

Second, when I say that NCat has a fixed place in Lewis' ranking of properties, I'm talking about NCat's *absolute* position in this ranking—i.e., the position given *simply* by measuring the length of NCat's definition. NCat's *relative* position in the ranking—i.e., its position vis-a-vis other properties—may well change as we move from possible world to possible world. Indeed, this is part of what Williams' larger argument ultimately turns on.

Third, the assumption that basic number-theoretic properties are perfectly natural goes by rather quickly in Williams' paper (see fn. 50 on p. 27). While I'm happy enough to grant Williams this assumption, I should note that it *does* rule out certain substantial positions in the philosophy of arithmetic—e.g., positions which view number-theoretic properties as simply general and/or highly derived physical properties. So, this is one place where pressure might be brought to bear on Williams' argument.

[11] A qualification is in order here. On some views, the property of being a cat wouldn't exist—or, at least, wouldn't be

At this point, the careful reader should see where this whole argument is going. Consider a series of possible worlds, $W_0, W_1, W_2, \ldots$, in which W_0 is just the actual world and the move from W_n to W_{n+1} adds one more layer of particles beneath the bottom layer in W_n. (So, the move to W_1 adds sub-quarks, the move to W_2 adds sub-sub-quarks, the move to W_3 adds sub-sub-sub-quarks, etc.) As we move from world to world in this sequence, the property of being a cat gets worse and worse. But, as we saw a few paragraphs ago, the property of being an NCat remains stable throughout the sequence. Eventually, therefore, we'll get to a world where the property of being an NCat actually comes out *better* on Lewis' fundamental ranking than the property of simply being a cat.

Clearly, this this result isn't specific to NCat. If we go far enough out our sequence, the property of being an NDog will be better than the property of being a dog, the property of being an NProton will be better than the property of being a proton, and the property of being an NPlanet will be better than the property of being a planet.[12] In fact, if we simply go far enough, we'll find a world, W_n, in which *all* of the properties which Charlie uses to interpret our language come out better than their ordinarily intended counterparts.[13] In W_n, therefore, Charlie will provide a better overall interpretation of our language than the ordinary intended interpretation does. *That,* Williams suggests, is a real problem for David Lewis.

This, then, is the heart of Williams' critique of Lewis. Using the fact that Charlie's complexity doesn't change as we move from one possible world to another—while the complexity of ordinary physical properties *does* change—Williams builds what he calls *Pythagorean worlds.* These worlds are just like our own world except that 1.) there are extra layers of subatomic particles living underneath the quarks, and as a result, 2.) the language that people speak in these worlds manages to refer only to numbers and to number-theoretic properties and relations. Since it's pretty silly to think that simply adding extra layers of subatomic particles could lead to this kind of massive shift in reference, Williams concludes that Lewis' theory of reference-fixing is inadequate.

Now, before looking at some problems with this argument, I want to highlight two assumptions that my presentation has essentially glossed over (assumptions which will turn out to be important when we get to section 3). First, Williams' argument presupposes that our language is essentially finite. To see the issue here, suppose that our language isn't finite and that, in particular, it includes an infinite list of

exemplified—in worlds with extra layers of subatomic particles (since the animals which look like cats in those worlds are "made up" of different stuff than our own cats are). For our purposes, however, this kind of issue isn't really all that important. After all, it's still the case that the things which people in those other worlds call "cats" are harder to define than the things which *we* call "cats"; so, the intended interpretation of "cat" in *their* world is still worse than the intended interpretation of "cat" in *our* world. This is enough for Williams' argument. See p. 29 n. 52 for William's own thoughts on this matter.

For convenience, I'm going to go ahead and talk as though it's the properties themselves which change their complexity as we move from possible world to possible world. The scrupulous reader should feel free substitute phrases like "the property which provides the intended interpretation of the predicate 'is a __' in world W" wherever they so desire.

[12] The notation here should be obvious. For any particular predicate, P, NP is simply the number-theoretic property which Charlie uses to interpret P.

[13] At any rate, that's the hope. See the qualification in the final three paragraphs of this section.

unary-predicates: $P_1, P_2, \ldots$. Suppose also that we need move to world W_1 in order to ensure that Charlie's interpretation of P_1 beats the ordinary intended interpretation of P_1, we need move to W_2 in order to ensure that Charlie's interpretation of P_2 beats the intended interpretation of P_2, and so on and so forth. In this case, there won't be any W_n in which Charlie beats the intended interpretation of our language for *every* predicate P_i. In fact, for any particular W_n, the intended interpretation of our language will beat Charlie for almost every P_i (i.e., for every P_i such that $i > n$). To avoid this kind of problem, Williams needs to ensure that there's some fixed stage at which *all* of the P_i's have been dealt with; so, he needs to build some kind of finiteness assumption into his argument.

Second, Williams needs some method for comparing the *overall* complexity of models which can't be compared on a simple predicate-by-predicate basis. That is, he needs the ability to compare models $\mathbb{M}_1$ and $\mathbb{M}_2$, where $\mathbb{M}_1$ beats $\mathbb{M}_2$ on the interpretation of some predicates, but $\mathbb{M}_2$ beats $\mathbb{M}_1$ on the interpretation of other predicates.[14] This is because, despite what I said a moment ago, Williams can't really ensure that there's any W_n in which Charlie beats the intended interpretation of our language on *all* predicates.

To see the problem here, suppose that our theory T includes just a little bit of set theory. (Of course, it can't include *very much* set theory since it's not committed to the existence of infinitely many sets. Still, we can assume that it thinks that there are *some* sets, and that these sets are abstract objects.) Further, suppose that in the actual world the property of being a set is more fundamental than the property of being an NSet, where NSet is just the property which Charlie uses to interpret the ordinary English predicate "is a set." Then the very same considerations which led us to think that properties like NCat and NSet don't change their complexity as we move from possible world to possible world should also lead us to think that the property of being a set doesn't change *its* complexity as we move from world to world. If this is right, then there won't be *any* possible world where Charlie's interpretation of "is a set" beats the ordinary intended interpretation.

To deal with this, Williams needs some way of summing up—or averaging—the overall complexity of a model. The hope is that, if we can simply make properties like NCat and NDog beat their ordinarily intended counterparts *really badly,* then any advantage which the intended interpretation of our language gets from abstract predicates like "is a set" or "is a proposition" will be *swamped out* by the advantage which Charlie get from physical predicates like "is a cat" or "is a dog."[15] So, even if Charlie doesn't beat the intended

[14] The simplest case here might be one where $\mathbb{M}_1$ uses P to pick out the protons and Q to pick out the quarks, while $\mathbb{M}_2$ uses Q to pick out the protons and P to pick out the quarks. So, $\mathbb{M}_1$ does a better job of interpreting Q than $\mathbb{M}_2$ does, while $\mathbb{M}_2$ does a better job of interpreting P than $\mathbb{M}_1$ does.

[15] Note that there's no problem making NCat and NDog beat their intended counterparts *really badly.* As we go out our sequence of W_is, the complexity of the intended interpretations of "is a cat" and "is a dog" gets worse and worse, while the complexity of NCat and NDog remains fixed. So, if we simply go out far enough, we can make the differences between these complexities as large as we want. In contrast, the difference between the complexities of NSet and the property of being a set remains fixed as we move from world to world. The only real question, therefore, is whether our methods of summing and/or averaging will allow the advantages Charlie gets from the interpretation of physical predicates to *cancel out* the advantages which the intended interpretation gets from abstract predicates.

interpretation of our language on *all* predicates, Charlie will still come out ahead on an all-things-considered basis. At any rate, that's what the summing and/or averaging is supposed to accomplish.[16]

2 A Quick Dilemma

In this section, I examine what I take to be the most obvious objection to Williams' new argument. To motivate this objection, we should start by noticing two features of that argument which may initially seem to be in tension. First, Williams' construction of Charlie turns on the assumption that our best overall theory of the world is compatible with there being only finitely many things. Second, Charlie's domain consists of natural numbers.

Now, on the surface, it's hard to see who would find the *combination* of these two features very plausible. On the one hand, suppose that you don't believe in the natural numbers (perhaps because you're a nominalist, and you don't believe in *any* abstract objects). Then you're not going to find Williams' new model-theoretic argument very threatening. After all, you have a perfectly good explanation as to why Charlie can't be the intended interpretation of our language: on your view, Charlie doesn't even exist! On the other hand, suppose that you *do* believe in the natural numbers. Then you almost certainly think that there are infinitely many such numbers, and so your best overall theory of the world isn't compatible with the assumption that there are only finitely many things. Once again, then, you have a perfectly good explanation as to why Charlie can't be the intended interpretation of our language: on your view, Charlie doesn't satisfy T.

This, then, is the motivating dilemma for this section. Depending on your views about the natural numbers, it seems that either you should think that Charlie is *too small* to provide an intended interpretation of our language (because Charlie is only finite) or you should think that Charlie doesn't even exist (because Charlie is numerical). Of course, as it stands, one horn of this dilemma depends on calling into question one of the background assumptions in Williams' argument—i.e., his assumption that we're not committed to the existence of infinitely many abstract objects. Hence, as natural as the dilemma may be, it may also seem to beg the question against Williams. What we want to know, therefore, is whether the dilemma can be sharpened up so as to be made consistent with Williams' background assumptions.

I think that the answer to this question is "yes." To see why, suppose that Williams' model, Charlie, has exactly n elements for some large, finite n. Then there are two cases. On the one hand, perhaps you don't think that there *are* n natural numbers. Maybe you're a nominalist who doesn't believe in numbers at all, or maybe you're just a pragmatist who thinks that the numbers run out somewhere around where your checkbook does. In either case, you won't think that Charlie exists, and so you won't be all that worried about Williams' new argument.

[16] For Williams' own thoughts on this summing/averaging issue, see pp. 16–17 (esp. p. 17 n. 31), 27–28 and 30 of his new paper. Note here that this summing/averaging issue gives us another reason for insisting that our underlying language be finite: it's a lot easier to compute finite sums and averages than infinite sums and averages. Finally, though I don't care *which* method of summing/averaging we use, the argument of section 3 will assume that *some such* method is available.

On the other hand, perhaps you *do* think that numbers exist and that there are at least n of them. Then you probably also think that there are some things which aren't natural numbers—say, three cats or four philosophy papers. On the whole, then, you think that there are at least $n+3$ things in the world (n numbers plus three or more other things). But this means that Charlie doesn't satisfy your best overall theory of the world, since Charlie satisfies a sentence which says that there are only n things in total. Once again, therefore, you have every reason to deny that Charlie satisfies your theory T, and so you have no reason to be worried about Williams' new argument.

In short: Williams' argument still faces essentially the same dilemma that it faced a few paragraphs ago. Either you think that Charlie is too small to provide an intended interpretation of your language (because Charlie says that there are only n things in the world, while your theory T says that there are more than n things in the world), or you think that Charlie simply doesn't exist (because there aren't enough natural numbers to make up Charlie's domain). In neither case will you find Williams' overall argument very troubling.[17]

This, then, is the basic dilemma that Williams' argument needs to overcome. Before examining some strategies for overcoming it, I want to make three remarks concerning the ways this dilemma interacts with the details of Lewis' and Williams' papers. First, we should note that both horns of this dilemma "transfer" from our world to Williams' Pythagorean worlds—i.e., that whichever horn of the dilemma holds in the actual world also holds in all of the Pythagorean worlds. This is important because Williams only claims that Charlie beats the intended interpretation of our language *in the Pythagorean worlds.* He doesn't also argue that our own world *is* a Pythagorean world (although he does want leave this open as a "non-skeptical epistemic possibility," p. 26). For my dilemma to have any real effect on Williams argument, therefore, it needs to hold in those Pythagorean worlds where Charlie is actually doing some real work.

Fortunately, it's pretty clear that my dilemma *does* transfer cleanly. On the one hand, Williams' argument assumes that if Charlie exists at all, then he exists in every possible world. Presumably this also means that if Charlie doesn't exist, then he doesn't exist in any possible world. So, the first horn of the dilemma transfers nicely. On the other hand, the structure of Williams' Pythagorean worlds ensures that our best overall theory of the world is identical to the theory held by our Pythagorean counterparts. So, if Charlie doesn't satisfy *our* theory, then he doesn't satisfy *their* theory either. This shows that the second horn of

[17] This dilemma is similar to a dilemma concerning the Löwenheim-Skolem version of Putnam's model-theoretic argument which I've discussed at some length in [3]. Let S be some particular axiomatization of set theory. Then Gödel's theorems show that Putnam cannot both 1.) use S as his background set theory and 2.) prove the existence of a non-standard model which satisfies S. In particular, then, Putnam cannot use the very set theory which is accepted by a particular realist to construct a non-standard model of that realist's "theoretical constraints." Hence, when faced with one of Putnam's non-standard models, the realist will always have two options: 1.) she can reject Putnam's model outright (because she doesn't accept the background set theory used in constructing the model), or 2.) she can question whether the model really satisfies her theoretical constraints (because Putnam's background set theory isn't strong enough to *prove* that his non-standard models satisfy these theoretical constraints). Clearly, this problem is quite similar to the does-not-exist/does-not-satisfy-T dilemma discussed above. For more on the Löwenheim-Skolem issue, see sections 2 and 5 of [3]; see also section 2 of [1].

our dilemma also transfers. At the end of the day, then, whichever horn of the dilemma holds in the actual world also holds in all of Williams' Pythagorean worlds.

Second, we should note that both horns of this dilemma are compatible with Lewis' own approach to interpretation—i.e., that neither horn involves the introduction of any new constraints on interpretation which are foreign to Lewis' original analysis. Lewis argues that the intended interpretation of our language is simply the most natural interpretation which happens to satisfy our best overall theory of the world. My argument, in turn, shows that we have no reason to believe that Charlie satisfies these two conditions: either because there is no Charlie, or because Charlie doesn't satisfy our best overall theory. Given Lewis' account of interpretation, therefore, we have no reason to think that Williams' argument raises any real problems concerning the interpretation our own language.[18]

Finally, I should emphasize that the mere fact that Williams' argument doesn't raise problems concerning the interpretation of *our* language doesn't mean that it fails to raise genuine difficulties for Lewis. To see why, note that my dilemma presents an essentially *dialectical* problem for Williams. It shows that, as long as Williams limits himself to assumptions that we ourselves accept as part of our best overall theory of the world, he cannot produce a version of Charlie which is compatible with that very theory. Hence, he cannot give *us* a reason for thinking that *our* languages and theories are susceptible to his new argument.

That being said, our languages and theories aren't the only ones which are relevant for assessing Lewis project. When we look at other, slightly simpler, languages, then I think we'll find that Williams' argument still raises an important theoretical difficulty. Consider a possible world containing a tribe of cavemen whose theoretical commitments are quite modest: they don't count past 20, they don't know about sets, functions or propositions, they don't believe in too many—or too recherché of—material objects, etc. Even though these cavemen themselves aren't in a position to appreciate something like Williams' construction of Charlie, *our* theoretical commitments are rich enough that *we can* follow the construction of a version of Charlie which applies to the cavemen's language/theory (call this model "CaveCharlie"). So, if we simply pass to a Pythagorean counterpart of the *cavemen's* world, then we'll find that CaveCharlie provides a better overall interpretation of cave speech than the ordinary intended interpretation does.

But surely something has gone wrong here. From a theoretical standpoint, it's just as problematic to think that our cavemen's talk about spears and mastodons really refers to sets of natural numbers as it is to think that *our* talk about cats and dogs really refers to NCats and NDogs. Even if the above dilemma can save *our* language from Williams' argument, therefore, the fact that slightly simpler languages are susceptible to that argument is still a real problem for Lewis.[19] Further, and as we'll see in the next section,

[18] In light of this, I'm not inclined to worry too much about the specific problems which Williams raises on pp. 30–31 of his new paper. These problems all concern the possibility that our *own* language might refer only to numbers and to number-theoretic properties and relations (e.g., if our own world turned out to be Pythagorean). While I agree that Williams' argument raises some real problems for Lewis (see below), I don't, for the reasons sketched above, find these kinds of "close to home" problems very threatening.

[19] To put this point another way, the fact that we accept a lot of *arithmetic* isn't what makes our use of terms like "cat" and "dog" semantically determinate. So, since our use of arithmetic *is* what saves our language from Williams' argument—i.e.,

there are some natural ways of modifying Williams' argument which may allow it to avoid our dilemma completely—i.e., to avoid it even in the case of our *own* languages. Hence, although I still think that this dilemma shows something interesting about the dialectics of Williams' new argument, I don't think that it *refutes* that argument or that it saves Lewis from the theoretical difficulties which the argument is supposed to uncover.

3 BigCharlie and BigCharlie*

In this section, I want to outline two ways of modifying Williams' argument so as to avoid the dilemma presented in the last section. To provide some intuitive motivation for these constructions, it's useful to begin by recalling the features of Williams' argument which gave rise to that dilemma in the first place. At the most basic level, the dilemma arises from the combination of two facts: 1.) the fact that Charlie is a finite model and 2.) the fact that our background language can talk about finite cardinalities in a fairly precise way. The combination of these facts was what enabled us to find a single *sentence*—the one which says that there are exactly n things in the world—about which Charlie and T had to disagree.

Next, we should recall why Williams wanted Charlie to be finite in the first place. Basically, the fact that Charlie is finite is what ensures that we can get good bounds on the complexity of the number-theoretic properties and relations which Charlie uses to interpret our language. It is, for instance, what ensures that properties like NCat and NDog can be given nice disjunctive definitions. In turn, it's the fact that we can give these definitions and find these bounds which ensures that there are possible worlds in which properties like NCat and NDog beat their ordinarily intended counterparts *really badly.* Hence, it's what ultimately ensures that there are worlds in which Charlie beats the intended interpretation of our language.

Given all this, my basic strategy for avoiding our dilemma turns on two key ideas. First, I'll try to construct an infinite version of Charlie. Since first-order languages are notoriously bad at pinning down the sizes of infinite sets, this will ensure that there's no first-order sentence concerning cardinality on which Charlie and T have to disagree.[20] Indeed, for convenience, both of my constructions will go a bit further and steal a page from Putnam's original argument by making their new versions of Charlie *isomorphic* to the intended model of our language. This will ensure that these new versions of Charlie satisfy *exactly* the same sentences as the intended model does (and so, on the assumption that the intended model satisfies T, that these new versions of Charlie also satisfy T).

Second, I'll try to effect this modification of Williams' argument without losing the nice bounds which Williams' originally obtained by making Charlie finite. My strategy for doing this splits into two parts. For purely physical predicates and relation symbols—like "cat" and "dog"—I'll just follow the lead of Williams' original argument: I'll assume that there are only finitely many *physical* objects, I'll replace these objects

in the dilemma just sketched—the dilemma itself probably doesn't get at the philosophical heart of Williams' argument. For discussion of a similar issue in the case of Putnam's Löwenheim-Skolem argument, see section 2.2 of [3] and sections 2–3 of [1].

[20] Assuming, of course, that T itself is compatible with the existence of infinitely many things.

with some nicely definable counterparts, and I'll then give the same kinds of disjunctive definitions that Williams himself originally gave. For more-abstract predicates and relation symbols—like "set" or "natural number"—I'll simply make sure to use only abstract primitives in defining the corresponding properties and relations. This will ensure that the complexity of these properties and relations remains fixed as we move from one possible world to another.[21] As before, then, if we simply move to a world where my new models beat the intended interpretation of our language *really badly* on physical predicates, then these models will also come out ahead on an all-things-considered basis.

So, there's the basic strategy. From a formal standpoint, both of my constructions begin with the same four assumptions. First, I assume that there are (only) three kinds of things in the world: physical objects, natural numbers, and other kinds of abstract objects. Second, I follow Williams in assuming that there are only finitely many physical objects, but I allow there to be infinitely many abstract objects, and my first construction will actually *require* the existence of all the natural numbers. Third, I join Williams in assuming that our background language is essentially finite. Finally, I need a somewhat technical assumption about definability: in the intended model of our language, any property/relation which is definable on the class of abstract objects is definable using *only* abstract primitives.[22]

Given these assumptions, the idea behind my first construction is quite simple. Let n be the number of material objects in the actual world. Then we can use the natural numbers, $1 \ldots n$, to stand proxy for these material objects, we can use the remaining numbers, $n + 1 \ldots \infty$, to stand proxy for *all* of the natural numbers, and we can let other abstract objects stand for themselves. Having done this, we can obtain a new model for our language by simply reinterpreting our names, predicates and relation symbols so as to respect these different substitutions—in, e.g., the manner of Putnam's original permutation argument. For convenience, let's call the model which results from this construction "BigCharlie."[23]

[21] The case of *mixed* predicates and relations—i.e., predicates which apply to both physical and abstract objects or relations which hold *between* physical and abstract objects—is a bit tricky. I'll say more about this case in footnotes 22 and 26.

[22] So, for instance, we can use the relation "is liked by" to define the set of numbers liked by Tim Bays. Then our assumption says that this set of numbers can also be defined in purely number-theoretic terms (perhaps I like only prime numbers or only numbers between 10 and 10,000). What the assumption rules out is the case where we use mixed predicates and relations—plus, perhaps, reference to some specific physical objects—to pick out some otherwise undefinable collections of abstract objects.

[23] More formally, let $f : \{x \mid x \text{ is material}\} \rightarrow \{1, \ldots, n\}$ be an arbitrary bijection, and let σ be the following function:

$$\sigma(x) = \begin{cases} f(x) & \text{if } x \text{ is material} \\ x + n & \text{if } x \text{ is a natural number} \\ x & \text{if } x \text{ is abstract, but not a natural number.} \end{cases}$$

Then the domain of our new model is just the class of all abstract objects, and our new interpretation function is given as follows (where i is just the interpretation function from the intended model of our language):

$i'(c) = \sigma(i(c))$ for any name c,
$i'(P) = \{\sigma(x) \mid x \in i(P)\}$ for any unary predicate P,
$i'(R) = \{\langle\sigma(x_1), \ldots, \sigma(x_n)\rangle \mid \langle x_1, \ldots, x_n\rangle \in i(R)\}$ for any n-place relation R.

Note that this construction makes σ into an *isomorphism* between BigCharlie and the intended model of our language.

BigCharlie has three nice properties. First, because BigCharlie's domain consists entirely of abstract objects, BigCharlie exists in all possible worlds. Second, BigCharlie satisfies exactly the same sentences as does the intended model of our language; so, if the intended model satisfies our theory T, then BigCharlie will also satisfy T.[24] Finally, if we let BCCat and BCDog be the properties which BigCharlie uses to interpret the predicates "cat" and "dog," then we can give these properties *exactly* the same kinds of disjunctive definitions that we earlier gave to NCat and NDog.[25] Hence, if we move to possible worlds with extra layers of subatomic particles living below the quarks, then we can make BCCat and BCDog beat their ordinary counterparts—i.e., cat and dog—as badly as we like.

Of course, BigCharlie still won't beat the intended interpretation of our language on *all* predicates (even when we move to other possible worlds). So, for instance, when it comes to interpreting a purely number-theoretic predicate like "is prime," BigCharlie is going to have to use a parasitic definition like the following:

$$x \text{ is BCPrime } \iff x - n \text{ is prime.}$$

But this isn't a problem. Any finite advantage which the intended interpretation of our language gets from the (repeated) "$-n$" factor in the definition of number-theoretic properties and relations will be swamped out by the advantage which BigCharlie gets from the interpretation of ordinary physical predicates. And this point is perfectly general. All of the properties and relations which BigCharlie uses to interpret our language can be defined using only abstract primitives.[26] Hence, their complexity won't change as we move from possible world to possible world. As before, then, if we simply move to a world where BigCharlie beats the intended interpretation of our language *really badly* on physical predicates like "cat" and "dog," then we can ensure that BigCharlie will also come out ahead on an all-things-considered basis.

[24] This follows from the existence of an isomorphism between the two models (i.e., the σ discussed in the last footnote).

[25] Note that this isn't way these properties were initially defined back in footnote 23. But, since BigCharlie simply uses a finite set of natural numbers to interpret predicates like "cat" and "dog," these alternate (disjunctive) definitions will capture the very same collections.

[26] To see this, we need to consider three cases. The case of purely physical predicates and relations can be handled using the finite-disjunction trick discussed above. The case of purely abstract predicates and relations—i.e., predicates and relations which *only* apply to abstract objects—can be handled via definitions of the form:

$$\begin{aligned}
\text{BCP}(x) &\iff [x \text{ is a number and } P(x-n)] \vee [x \text{ is not a number and } P(x)]. \\
\text{BCR}(x,y) &\iff [x \text{ and } y \text{ are numbers and } R(x-n, y-n)] \vee [x \text{ and } y \text{ are not numbers and } R(x,y)] \vee \\
&\qquad [x \text{ is a number and } y \text{ is not and } R(x-n, y)] \vee [y \text{ is a number and } x \text{ is not and } R(x, y-n)].
\end{aligned}$$

Finally, mixed predicates and relations can be "factored" into a finite number of abstract pieces which are then combined disjunctively. So, for instance, for an m place relation, there are only $(n+1)^m$ ways to insert material objects into (some of) that relation's argument places. Of these, n^m insert material objects into *all* of the relation's argument places and so can be dealt with using our standard finite-disjunction trick. The remaining $(n+1)^m - n^m$ leave some argument places open and so give rise to sub-relations on the abstracts. Given our definability assumptions, these sub-relations can themselves be defined in purely abstract terms. In the end, therefore, the whole relation can be captured using a long, but still finite, disjunction in which each disjunct uses only abstract primitives.

This, then, gives us one technique for modifying Williams' argument so as to overcome the dilemma sketched in the last section. In turns, essentially, on the observation that there's an asymmetry in the ways Williams' argument needs to treat physical predicates and abstract predicates. For physical predicates, something like the finite-disjunction trick seems to be necessary if we want to get nice bounds on the complexity of the properties which our models use to interpret these predicates. Hence, it's important that we start with only finitely many physical objects and that we replace these objects with some nicely definable counterparts (e.g., numbers).[27] For abstract predicates, we simply need to ensure that our models only use abstract properties to interpret these predicates, and we can then use some kind of summing and/or averaging technique to overcome any local advantages which this interpretation gives to the intended model. Hence, there's no need to limit ourselves to using only finitely many abstract objects, and there's "room" to make use of Putnam-style isomorphism tricks.

Of course, once we've understood the basic strategy behind this example, it's straightforward to come up with other, analogous examples. So, for instance, instead of building a model in the actual world and then showing that this model causes trouble in other possible worlds, we could simply look for a possible world which contains a troublesome model. Consider a world which is just like ours except that 1.) it includes extra layers of subatomic particles living beneath the quarks and 2.) it includes a fundamental physical relation which linearly orders the "bottom" layer of these particles (perhaps based on the time at which individual particles come into existence). Next, construct an analog of BigCharlie which uses the first n of these subatomic particles to stand proxy for ordinary physical objects, and which leaves all of the abstract objects—including the natural numbers—to stand for themselves. Call this model "BigCharlie*."

Now, just as in the BigCharlie example, BigCharlie* is isomorphic to the ordinary intended model of our language; so, on the assumption that the intended model satisfies T, so too does BigCharlie*. Further, because the individual elements in BigCharlie*'s domain are so basic—and because the properties which BigCharlie* uses to interpret physical predicates can all be defined in terms of a *fundamental* ordering on those elements—BigCharlie* constitutes a more eligible interpretation of our language than the ordinary intended interpretation does. As before, then, BigCharlie* creates problems for Lewis' argument without running aground on the dilemmas presented in the last section. So, it once again strengthens the overall force of Williams' original argument.

4 Some Philosophical Lessons

In the last two sections, I looked at some relatively technical issues concerning Williams' new version of the model-theoretic argument. In this section, I want to step back and make a few remarks concerning the

[27] In fact, the assumption that there are only finitely many physical objects isn't really essential. It's enough to assume that there's a possible world which looks just like our own world but which contains only finitely many physical objects. Because the people in that world have the same theory as we do—i.e., T—we can simply use their world as the basis for a BigCharlie-like construction. Since, the resulting model is isomorphic to the intended model of *their* language, it has to satisfy *our* theory T.

broader philosophical lessons which I think we can learn from this argument. To bring these lessons out, it's useful to begin by going back and reexamining Putnam's original argument. That argument starts with two things: a first-order language and a collection of mathematical models for that language. (Note that at this point I haven't yet introduced any *sentences* which these models are supposed to satisfy; for now, we simply have a language and a collection of models for that language.) Given this, Putnam challenges us to *pare down* this collection of models until we have a single model which will constitute the "intended model" of our language.

To accomplish this "paring down," Putnam gives us two tools. First, he allows us to fix the interpretation of certain logical constants in an "absolute" manner—i.e., by specifying the intended interpretation of these constants in the definition of the first-order satisfaction relation.[28] Second, Putnam allows us to stipulate that certain sentences are supposed to count as true. He then suggests that the intended models for our language are just those models which happen to satisfy all of the sentences which we have chosen to specify (using, of course, the notion of "satisfies" which was defined in the first part of this exercise).

Given this way of framing things, the model-theoretic side of Putnam's argument simply shows that these two tools don't get us very far. As long as we use the ordinary first-order satisfaction relation to formulate our background semantics, no mere collection of *sentences* will allow us to fix the intended interpretation of non-logical terms like "cat" and "dog." Instead, the significance of these terms will vary quite widely as we move from one model to another. This is true *even if* we restrict ourselves to models which satisfy some antecedently given collection of sentences. Hence, unless we can find some other tools for fixing intended interpretations, semantic indeterminacy looks to be a serious threat.

Now, for our purposes, there are three things to notice about this conclusion. First, the conclusion itself should be utterly unsurprising. To think that we *could* fix the intended interpretation of our language in the way Putnam suggests would amount thinking that ordinary predicates like "cat" and "dog" could be defined in purely logical terms. As I have noted elsewhere, this thought would commit us to an extremely strong form of logicism: not just logicism about mathematics, but logicism about zoology, logicism about astronomy, logicism about animal husbandry, etc.[29] If, like me, you find this kind of "global logicism" pretty implausible, then you shouldn't be at all surprised by the model-theoretic side of Putnam's argument.[30]

Second, we wouldn't really *want* first-order model theory to fix the interpretation of non-logical terms like "cat" and "dog." From a mathematical standpoint, model theory is *designed* to allow substantial variation in the models at which particular sentences are interpreted (and, indeed, in the models at which these sentences come out true). If our model theory pinned things down too tightly—e.g., by fixing the interpretation of

[28] So, for instance, the interpretation of → and = doesn't get fixed by the interpretation functions of particular models; instead, it's fixed *directly* by the recursion clauses in the definition of satisfaction. For more on the significance of this fact, see section 2 of [2].

[29] For a more-detailed discussion of this point, see pp. 9–10 of [2].

[30] Of course, other parts of Putnam's argument are genuinely more surprising. In particular, Putnam's just-more-theory argument—and argument which purports to show that global logicism is, in fact, the only real option for interpreting our language—is substantially (and deservedly) more controversial than the purely model-theoretic side of Putnam's argument.

ordinary terms like "cat" and "dog"—then it would make our language mathematically uninteresting. From a more philosophical standpoint, the fact that model theory lets us vary the interpretation of our terms is part of what makes the subject so philosophically fruitful—e.g., it's what lets us give model-theoretic analyses of notions like *logical consequence* and it's what lets us use models as formal proxies for possible worlds in certain metaphysical arguments. Given this, neither philosophers nor mathematicians should even want first-order model theory to fix the interpretation of non-logical terms like "cat" and "dog."[31]

Finally, the underlying reason that first-order model theory doesn't pin down the interpretation of terms like "cat" and "dog" is that the model-theoretic machinery itself isn't defined in terms of the ways people actually use these terms. Clearly, neither the definition of a model nor the definition of satisfaction makes any *explicit* reference to the ways people use their language (as opposed, for instance, to the case where we simply build some version the causal theory of reference into our specification of the class of admissible interpretation functions). Nor does information concerning the ordinary use of terms like "cat" and "dog" play any *implicit* role in defining the model-theoretic machinery (as opposed, once again, to the role that the ordinary use of phrases like "if. . . then" and "is equal to" plays in motivating the definition of first-order satisfaction).[32] Given this, it's not at all surprising that our model theory doesn't capture the intended interpretation of ordinary terms like "cat" and "dog." Since the model-theoretic machinery makes no reference to ordinary usage, it's only natural that there will be cases where the usage and the model theory start to come apart.

This brings me to Williams' new argument. What Williams' argument shows is that this third point generalizes rather broadly. If the mechanisms we use to pick out the intended interpretation of our language aren't connected to our actual *use* of that language, then they're unlikely to capture the ordinary sense of "intended interpretation." In Lewis' case, the fact that one property is more fundamental than another property doesn't itself have anything to do with the ways people use their language.[33] (Indeed, when we rank properties in other possible worlds, we don't even need to know whether there *are* any people who use language in those worlds.) As a result, Lewis overall ranking on models is also independent of usage; it's just an abstract ranking based on the physical significance of the properties which a model uses to interpret the predicates and relations in a given formal language. Hence, it's not surprising that Williams can find artificial examples—like his Charlie example—where Lewis' ranking starts to come apart from ordinary usage.[34]

[31] For a more-detailed discussion of this point, see pp. 499–500 of [4].

[32] Actually, this isn't quite true. In giving a model-theoretic interpretation of our language, we allow ordinary usage to fix basic grammatical categories—e.g., to distinguish between names and predicates or between one-place predicates and two-place relations. But we don't allow our model-theoretic machinery to reflect more-detailed facts about usage—e.g., the fact that we use "cat" to talk about animals and "proton" to talk about sub-atomic particles. As far as our model theory is concerned, "cat" and "proton" are just two, essentially interchangeable, unary predicates.

[33] This independence is quite clear in Lewis' own discussion of the matter. See pp. 64–66 of [7].

[34] Another problem here is that Lewis-style eligibility constraints aren't really designed to deal with the kinds of radical indeterminacy that are at issue in Williams' new argument. In general, eligibility constraints are supposed to help us distinguish between different, *metaphysically plausible,* competitors for the reference of a given term—e.g., to determine decide whether the term "cat" refers to cats or to cat stages or to undetached cat parts. They're not really designed to deal with the possibility that "cat" refers to sets of *natural numbers.* I'm grateful to Zoltan Szabo for helping me to focus on this point.

This, then, is the general point which Williams' new argument brings out so nicely. Unless the mechanisms used to fix the "intended model" of our language are connected to the ways we actually use that language, it's unlikely that they will reliably pick out the models which we would intuitively want to count as "intended." Putnam's original model-theoretic argument used the fact that the definitions of "model" and "satisfaction" don't refer to the ways people use their language to show that a purely model-theoretic notion of interpretation won't match up with our intuitive understanding of interpretation. Williams' new argument uses the fact that Lewis' ranking on models *also* doesn't refer to the ways people use their language to show that the notions of interpretation generated by that ranking still don't match up with our intuitive understanding of interpretation.

Now, as far as I can see, this point should generalize pretty widely. For an account of interpretation to be plausible—and, in particular, for it to give the right answers for many different languages and across a wide range of counterfactual contexts—it's going to have to make some fairly explicit reference to the ways people actually use their languages. As nice as model-theoretic semantics may be, and as useful as eligibility constraints are for eliminating certain close competitors to the intended interpretation of our language (cf. fn. 34), they're not nearly enough to provide a *complete* theory of interpretation. This, I think, is the key insight of Williams' new paper, and I want to end by thanking him for developing it so elegantly.

References

[1] Timothy Bays. More on Putnam's models: A response to Bellotti. (In Preparation).

[2] Timothy Bays. Two arguments against realism. (In Preparation).

[3] Timothy Bays. On Putnam and his models. *The Journal of Philosophy*, XCVIII:331–50, 2001.

[4] Timothy Bays. The mathematics of Skolem's paradox. In Dale Jacquette, editor, *Philosophy of Logic*, pages 485–518. Elsevier, London, 2006.

[5] Michael Devitt. *Realism & Truth.* Princeton University Press, Princeton, 1984.

[6] David Lewis. New work for a theory of universals. *Australasian Journal of Philosophy*, 61:343–377, 1983.

[7] David Lewis. Putnam's paradox. *Australasian Journal of Philosophy*, 62:221–236, 1984.

[8] Van McGee. Inscrutability and its discontents. *Noûs*, 39:397–425, 2005.

[9] Hilary Putnam. Realism and reason. In *Meaning and the Moral Sciences*, pages 123–138. Routledge, New York, 1978.

[10] Hilary Putnam. *Reason, Truth and History.* Cambridge University Press, New York, 1981.

[11] Hilary Putnam. Models and reality. In *Realism and Reason*, pages 1–25. Cambridge UP, Cambridge, 1983.

[12] Hilary Putnam. Model theory and the 'factuality' of semantics. In Alexander George, editor, *Reflections on Chomsky*, pages 213–231. Blackwell, Cambridge, 1989.

[13] Hilary Putnam. A defense of internal realism. In *Realism with a Human Face*, pages 30–42. Harvard UP, Cambridge, 1990.

[14] J. R. G. Williams. Eligibility and inscrutability. *The Philosophical Review*, ??:??–??, 2006.