Published paper
Williams, J.R.G. (2007) *Eligibility and inscrutability,* Philosophical Review, Volume 116 (3), 361 – 399.

Eligibility and Inscrutability

J. Robert G. Williams
University of Leeds

The philosophy of *intentionality* asks questions such as: in virtue of what does a sentence, picture, or mental state represent that the world is a certain way? The subquestion I focus upon here concerns the semantic properties of language: in virtue of what does a name such as 'London' refer to something or a predicate such as 'is large' apply to some object?

This essay examines one kind of answer to this "metasemantic"[1] question: *interpretationism*, instances of which have been proposed by Donald Davidson, David Lewis, and others. I characterize the "two-step" form common to such approaches and briefly say how two versions described by David Lewis fit this pattern. Then I describe a fundamental challenge to this approach: a "permutation argument" that contends, by interpretationist lights, there can be *no fact of the matter* about lexical

Variants of this article have been presented at talks in St. Andrews, Leeds, Barcelona, and at Cornell University. I learned much from the discussion on these occasions, especially from the excellent responses to the essay presented by Timothy Bays and John Hawthorne at the *Philosophical Review* workshop at Cornell. Particular thanks go to Elizabeth Barnes, Ross Cameron, Kit Fine, Daniel Isaacson, Stephen Leuenberger, Joseph Melia, Andy McGonigal, Brian McElwee, Daniel Nolan, Marcus Rossberg, Daniel Rothschild, Crispin Wright, and two anonymous referees for *Philosophical Review*. The research on which this essay is based was carried out within the Arché AHRC research centre for the philosophy of logic, language, mathematics, and mind in St. Andrews, and I was supported by an AHRC doctoral research studentship.

1. I follow Kaplan 1989 and Stalnaker 1999 in using "metasemantics" as a name for the inquiry into the nature of semantic facts. Other terminology for similar enterprises include "foundational semantics" and "theory of meaning." Though the terminology is not perfect, I have no wish to introduce yet more nomenclature into this area.

Philosophical Review, Vol. 116, No. 3, 2007
DOI 10.1215/00318108-2007-002

content (e.g., what individual words refer to). Such a thesis cannot be sustained, so the argument threatens a reductio of interpretationism.

In the second part of the article, I will give what I take to be the best interpretationist response to the inscrutability paradox: David Lewis's appeal to the differential "eligibility" of semantic theories. I contend that, given an independently plausible formulation of interpretationism, the eligibility response is an immediate consequence of Lewis's general analysis of the theoretical virtue of simplicity.

In the final sections of the article, I examine the limitations of Lewis's response. By focusing on an alternative argument for the inscrutability of reference, I am able to describe conditions under which the eligibility result will deliver the wrong results. In particular, if the world is complex enough and our language sufficiently simple, then reference may be determinately secured to *the wrong things*.

1. Metasemantics, Interpretationism, and Inscrutability

In *Psychosemantics*, Jerry Fodor (1987, 97) urges philosophers to give a broadly reductive account of the nature of intentional and semantic facts:

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear upon their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in the face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or supervenience on?) properties that are themselves *neither* intentional *nor* semantic. If aboutness is real, it must be really something else.

Fodor himself develops a positive reductive account of "aboutness," attempting to describe without appeal to semantic notions a causal relation between words and objects that can be identified as *reference*. With the semantic properties of basic lexical items fixed in this way, one can derivatively account for the semantic properties of more complex expressions, such as sentences.[2] Like identity theories of mental properties, such

2. Field (1972) explicitly proposes this metasemantic account. Fodor's own views are complicated by the fact that he takes the primary task of a causal theory of reference to be to give the semantic properties of the "language of thought" rather than natural language. Compare Field 1978.

an "identity theory" of semantic properties has considerable prima facie appeal to the reductively minded philosopher.

The focus of our inquiry here is a range of theories that take seriously Fodor's challenge to provide a broadly reductive account of the semantic but suggest a different (though equally ambitious) shape for the account to take. Indeed, they reverse the order of explanation suggested by causal theorists of reference. Broadly, these "interpretationisms" adopt a two-step strategy. One begins with facts about the states of the world in which sentences are uttered. One then gives a recipe for extracting data in the form of a correlation of sentences with appropriate contents.[3] Secondly, one maintains that *for an expression to have semantic property P* is just for selected theories (those whose predictions match up with this data) to ascribe that property to the word.[4] Overall, semantic facts—for example, that 'Billy' refers to the person X, or that 'very' is an intensifier that operates on predicates in some characteristic way—emerge because they are part of a simple, finitary theory whose predictions mesh with facts about utterance conditions for sentences.[5]

On interpretationist accounts, there need be no "reference relation," characterizable in nonsemantic terms, onto which our semantic

3. Sentential contents might be taken to be propositions (structured or unstructured), Davidsonian truth conditions, or even truth values. The choice will be determined by the particular interpretationism in view and the semantic framework targeted.

4. It may seem that whereas causal theorists reduce the semantic properties of sentences to those of lexical items, interpretationisms reduce the semantic properties of lexical items to those of sentences. But this is not quite the right description of the situation. For (most extant) interpretationists, it is the *usage properties* of sentences (the famous "patterns of assent and dissent") that have metaphysical priority. But then the *semantic properties* of *all* expressions are fixed holistically by the selected meaning-fixing theory. So there is no priority of *semantic properties* of sentences over words or vice versa for the interpretationist.

5. Though we shall often talk about extensional semantic properties such as reference, the account is intended to generalize to a broad class of representational properties of language. For example, what possible-world intension a word is assigned will be fixed in parallel fashion. My working assumption is that natural language will at most require the "double-indexed" intensions described in Lewis 1980.

Terminology in this area is notoriously contested (see ibid.). I shall use "content" as a neutral term to describe the semantic value of words within whatever framework is salient. So when discussing an extensional semantic theory, the content of a name will be its referent, whereas in the context of a double-indexed semantics, the content of a name will be (say) a function from context-world pairs to truth values. Relative to a different framework, the content of a sentence might be a structured proposition. I shall make no attempt to canvass all these options here; where the distinctive characteristics of the semantic framework being presupposed are important, this will be flagged.

vocabulary latches. Rather, we explain the constitution of semantic facts, such as 'Londres' referring to London, by appeal to holistic properties of a theory in which this claim figures—in particular, its success in generating sentential data on which we have an independent grip.[6] The lack of reductive *identification*, however, should not be thought to vitiate the reductive project. If we can pick out which holistic properties a correct semantic theory should have, in nonsemantic terms, we still have in prospect an overall reductive explanation of semantic truths.[7]

Interpretationist theories sharing the two-step form are various in detail.

1. They differ over how to express the correlation between sentences and states of the world, what the *relata* should be taken to be, and even whether the correlation should be taken to be a *relation* at all: contrast Davidson 1967's use of "T-sentences" with Lewis 1983 [1975]'s pairing of sentences with propositions.
2. They differ about which resources they will allow themselves in characterizing the data. Lewis explicitly appeals to a coarse-grained characterization of the content of propositional attitudes (expectation and preference) in characterizing the pairing of sentences and propositions. This means that his metasemantic account will only constitute a partial answer to the challenge given above: it will afford a reduction of semantic facts to (coarse-grained) intentional facts.
3. They differ as to what kind of semantic theory should be used to explain the data identified: perhaps a possible-worlds semantic theory is favored (Lewis 1970a); or perhaps a David-

6. This contrasts directly with one version of causal theories of reference, whereby the reference relation can be identified with causal relations.

7. There are philosophical accounts of meaning closely linked to interpretationisms, which do not look as if they can play this reductive role. There is a tradition that draws on Davidson's work, which thinks of semantic properties by analogy to *secondary* properties. Just as facts about what color properties an object has are thought by many to be constituted in part by the judgments of well-placed observers, this tradition thinks of semantic properties as constituted in part by the judgments of well-placed interpreters.

The version of interpretationism I will be considering (following Lewis) emphasizes the *interpretation* (i.e., the meaning-fixing semantic theory) rather than the *interpreter*. Unlike its rival, it makes no appeal to interpreters judging this or that, and it promises an account in *nonsemantic terms* of what factors make a given semantic theory the "selected" meaning-fixing one.

> sonian truth-theoretic semantics would be adopted (see Davidson 1967; Larson and Segal 1995).

No matter what exact form they take, there are certain virtues shared by all interpretationist theories. It is a desideratum on metasemantic theorizing that the final account be ("to some extent or other") reductive. This can be respected by interpretationism, but is not allowed to dominate at the expense of other, equally well-founded desiderata.[8] A good metasemantic theory should be *universal*: applicable to all words, rather than focusing on a range of special cases (perhaps names and basic predicates). Moreover, it should be *nonrevisionary*: just as we should look askance at philosophical theories about the nature of mathematics that cannot underpin standard arithmetic, we should look askance at philosophical theories about what fixes the semantics of language that cannot underpin best theory in empirical linguistics, whatever this is.[9]

Interpretationist approaches take universality and nonrevisionism as seriously as the reductive ambition. Universality is secured since all lexical content is treated in the same way: interpretationism gives conditions under which a complete semantic theory is true, and successful semantic theories will provide an account of the systematic contribution to the truth conditions of sentences made by prepositions, connectives, modifiers of various kinds, as well as names and predicates. Since the accomplishments of empirical linguists in developing semantic theories for natural languages are the raw material for the metasemantic accounts of particular languages, nonrevisionism is part of the setup from the beginning.

In what follows we concentrate on two forms of interpretationism described by David Lewis.

1.1. Lewis on Interpretationism

Lewis's overall reductive account of intentionality incorporates an interpretationism about semantic properties.[10] His first step is to identify *linguistic conventions* that govern utterances of sentences: these will provide the pairings of sentences with appropriate sentential contents (in this framework, coarse-grained propositions).

8. As noted earlier, the reductive ambition is a major part of the Lewisian tradition of interpretationism that is our focus here.

9. Compare the complaints in Williamson 2006.

10. For the overall project see Lewis 1974, 1994b. For the specifically interpretationist component see Lewis 1969, 1983 [1975], 1992. For an introduction see Nolan 2005, chaps. 6–7.

A convention in Lewis's sense is a certain kind of regularity in action. The basic data for Lewis's interpretationism will be conventional regularities of *only uttering S if one believes p* (the "convention of truthfulness").[11] The pairing of sentences with states of the world that interpretationism requires is provided for: *S* is paired with the proposition *p* if and only if there is a convention to *only utter S if one believes that p*.

Lewis 1983 [1975] exploits linguistic conventions to develop an interpretationist metasemantic account. Whether or not a semantic theory (what he calls a "grammar") is correct, relative to a given population, turns on which assignment of propositions to sentences (what he calls a "language") is correct for that population. This in turn is fixed by the linguistic conventions that prevail:

> I would say that a grammar Γ is used by *P* if and only if Γ is a best grammar for a language Ł that is used by *P* in virtue of a convention in *P* of truthfulness and trust in Ł; and I would define the meaning in *P* of a constituent or phrase . . . accordingly. (Lewis 1983 [1975], 177)

In Lewis's account of what makes a "language" correct, characterized in terms of conventions, we find the characteristic first step of an interpretationist metasemantics—identification of sentential data. In his account of the correctness of a "grammar" (semantic theory), we find the second component—lexical content-facts fixed by the best theory of this data.

Lewis 1999b [1984] describes a different interpretationism—*global descriptivism*. The first interpretationist step is to construct, for language as a whole, a "term-introducing" or "total" theory. Lewis is extremely unspecific about exactly what this "total theory" is.[12] Global descriptivism

11. This is the formulation in Lewis 1969. In Lewis 1975 he complicates the account by adding "conventions of trust": the conventional regularity of *forming the belief that p in response to hearing someone utter S*.

Further, more complex, approaches to specifying appropriate linguistic conventions are possible. Griceans, for example, might wish to adopt interpretationism about *timeless meaning* of linguistic items by means of conventions of individuals to speaker—mean *p* when uttering *S* (cf. Schiffer 1972). Avramides 1997 endorses this kind of proposal, combined with interpretationism, in addressing lacunae within the Gricean framework. Not only does it extract a notion of the "timeless meaning" of sentences from the representational properties of sentence-tokens the Gricean has available, but it will underpin ascriptions of subsentential meaning, which otherwise have no obvious place in the Gricean framework.

12. He is perhaps not himself endorsing the theory but rather offering it as a reconstruction of the view described as "standard" by Putnam 1980.

therefore delimits a *class* of views, distinguished by the characterization of total theory they offer.

One version in particular is often associated with Lewis. This sees it as a generalization of the so-called Ramsey-Carnap-Lewis treatment of theoretical terms (Lewis 1970b). This treatment is restricted to the terms of some particular domain of inquiry: say, heat, or mentality, or inheritance of behavioral traits. In the case of mentality, the first step might be to gather platitudes such as: "If someone hits you hard, and you are paying attention, you will feel pain"; and "If you are feeling pain, then unless distracted you will tend to wince and groan." The introduction of theoretical terms is supposed to be accomplished by formulating such platitudes efficiently and then transforming the resulting *folk theory* into a definition, say, of mental vocabulary, by the technique of "Ramsification."[13] The associated version of global descriptivist metasemantics takes "total theory" to be "global folk theory," the sum total of all the platitudes gathered from every walk of life—all the sentences that we take to be too obvious to question. In what follows, I shall assume this is the form of global descriptivism under discussion.

Next:

> The intended interpretation will be the one, if such there be, that makes the term-introducing theory come out true. (Lewis 1999b [1984], 60)

Again we find a two-step strategy. The first interpretationist step is to identify an appropriate set of sentences—global folk theory. The second interpretationist step is to find a semantic theory that renders all (or: enough) of these sentences true. (Equivalently, we could formulate the data as a pairing of sentences with truth values and require the semantics to match this.) If the language is sufficiently simple, the semantic theory can just specify a first-order model of the target sentences.[14]

13. For an account of this, see Lewis (1970b).

14. There are well-known worries about the adequacy of such models to handle some very simple languages, such as the language of first-order set theory. The problem is that standardly models are taken to be certain set-theoretic constructions, and in particular, the domain of a given model has to be set sized. But since there is no set of all sets, there will be no model corresponding to the intended interpretation of set theory. One way of addressing this point would be to dispute the possibility of languages with absolutely unrestricted first-order quantification (Lewis 1970a). Another would be to give non-set-theoretic explications of interpretations and models of a language—we could instead use the resources of type theory, as suggested by Williamson (2003). Thanks to an anonymous referee for highlighting this point.

1.2. Paradoxical Inscrutability

An *inscrutability argument* seeks to demonstrate that one can account for the patterns in linguistic usage by means of radically deviant assignments of lexical content. If all that is needed for an assignment of reference to be correct is for it to generate the right truth conditions for sentences, then a successful inscrutability argument will show that *there is no fact of the matter* which object an ordinary name refers to. For example, according to the "permutation" argument popularized by Hilary Putnam, 'London' might as well pick out Sydney, Australia, as London, England. If the argument is sustained, reference would be "inscrutable"—there would be *no fact of the matter* whether the word 'Londres' refers to London, England, or to Sydney, Australia.[15]

As I present them, permutation arguments for the inscrutability of reference will depend essentially on interpretationism.[16] More specifically, I discuss inscrutability arguments directed against global descriptivism. Following the methodology of Lewis 1999b [1984], I will initially focus on a first-order fragment of natural language.[17]

Recall once more that the core pattern of interpretationist metasemantic theories incorporates two elements. The semantic properties of a language are fixed by assigning contents to sentences. Facts about lexical content are those given by an appropriate theory that fits this sentential data.

15. I would prefer to use the term 'indeterminacy' rather than 'inscrutability' to describe the theses argued for, to avoid any connotation that the theses at hand are epistemic in force. Unfortunately, terminology has split so that 'indeterminacy of meaning' is standardly taken to refer to *sentential* meaning.

There are those who argue that all indeterminacy should be seen as an epistemic phenomenon (Williamson 1994). I do not wish to rule out this (surprising) contention: but nothing in the inscrutability arguments commits one to such a strong view.

16. With many others (e.g., Lewis 1999b [1984]), I hold that moves intended to make the permutation arguments universally applicable (for example, to argue for radical inscrutability in the context of a causal theory of reference) are unsuccessful. For instructive discussion, see Field 1975. (This is not to say that *other* ways of arguing for [limited] inscrutability will not work: arguably Quine's "argument from below" [Quine 1960, chap. 2] poses a threat even given a causal theory of reference.)

17. The arguments can be adapted to (a) more sophisticated semantic theories and (b) to different versions of interpretationism. On (a), an appendix to Putnam 1981 sketches an extension of the permutation argument to modal languages. This is further developed in Hale and Wright 1997, including treatment of languages involving higher-order resources. Elsewhere, I have worked through the details of adapting the permutation argument to a (double-indexed) multiply intensional higher-order language of the kind described in Lewis 1970a, 1980. This work allows one to answer the objections

Assume the data provided by the interpretationist is given. We observe (a) there are many theories that match the sentential data; (b) we know that not all of them are correct. If the correctness of a theory just *consists* in accounting for sentential data, then it seems that we cannot maintain (a) and (b). I here review a simple argument for (a), in informal terms.

Suppose we are working with a language whose only nonlogical vocabulary consists of names and predicates. What Putnam (1981) and others, such as Field (1975), Wallace (1977), and Davidson (1979) observe is that crazy assignments of reference to names can be "canceled out" by a compensating assignment of extensions to predicates, so that, overall, the truth value of sentences is unaffected.

Take a crazy assignment of reference, on which 'Billy' refers, not to Billy, but to the Taj Mahal, and, correspondingly 'Taj Mahal' refers to Billy. We show how to assign extensions to predicates so that every sentence that was true on the standard interpretation will still be true.

In setting up our interpretation, we shall use the phrase 'the image of *x*' to pick out *x* whenever *x* is anything other than Billy or the Taj Mahal, to pick out Billy if *x* is the Taj Mahal, and to pick out the Taj Mahal if *x* is Billy. Thus, we can describe our crazy reference scheme as follows: *N* crazy-refers to *x* just in case *N* standardly-refers to *y* and *x* is the image of *y*. Take any atomic predicate *P*. We adjust for the crazy assignment of reference in the following way: we let *P* crazy-apply to *x* if and only if *P* standardly applies to some *y*, such that *x* is the image of *y*.

Let the crazy-interpretation of our language be one where reference of names and the extensions of predicates coincide with crazy-reference and crazy-application. What we note is that the twists cancel out—the distribution of truth values to sentences is the same on both interpretations. All this can be spelled out formally and extended to any "permuted" reference scheme we choose.[18] What we find is that, at the

---

made in recent work by McGee (2005), who focuses on alleged difficulties in applying the permutation arguments in the context of a language allowing quantification into modal contexts. The issue revolves around the treatment of variables within the rich semantic theory: I discuss this in Williams 2005, chap. 5, appendix C.

On (b), some complex issues arise in adapting permutation arguments to a Tarskian truth-theoretic semantics or a semantic theory formulated in terms of structured propositions. I discuss these in Williams 2005, chap. 5, and Williams forthcoming a.

18. Thus: let ϕ be a permutation of the domain of quantification. Suppose *N* refers to *a*. Then let it crazy-refer to $\phi(a)$. Suppose *P* refers to the function from objects to truth values *f*. Then let it crazy-refer to the function $f \circ \phi^{-1}$, which again takes objects to truth values. We note that

level of sentences, all is as it was on the standard interpretation. This simple observation suffices to establish (a), above, if it is global descriptivism, rather than one of the more sophisticated interpretationist proposals, that is in view.

According to interpretationism, there is nothing more to semantic properties than being part of a theory adequate for a certain range of data that is characterized *at the level of sentences*. If generating the appropriate sentential data is all that is required, then the "standard" theory does no better than any one of the crazy theories. It appears that the interpretationist is committed to there being *no fact of the matter* about which of the theories is "correct" and, consequently, is committed to there being no fact of the matter about whether 'Billy' refers to Billy rather than the Taj Mahal.[19]

The argument for inscrutability just given attempts to establish (a): that there are many semantic theories adequate to the data that the interpretationist provides. But for this to constitute a reductio of interpretationism, we would need, in addition, (b): that not all of them can be correct. That is, one needs to show why one should not simply endorse the conclusion of the radical inscrutability argument: as does Davidson (1977, 1979). This requires substantive argument, which it is beyond the scope of this article to develop.[20] In what follows, I will assume such radical inscrutability theses are not only intuitively repugnant but also theoretically unacceptable.

- *PN* is true (under the crazy interpretation) ↔
- The function *P* crazy-refers to takes the crazy-reference of *N* to the true; ↔
- $f \circ \phi^{-1}$ maps $\phi(a)$ to the true; ↔
- $f(\phi^{-1}\ (\phi\ (a)))$ = the true; ↔
- fa = the true; ↔
- the function *P* refers to takes the reference of *N* to the true; ↔
- *PN* is true (on the standard interpretation)

Such arguments show that atomic truth values are invariant under the permuted interpretations. A simple induction shows this to be the case for all sentences whatsoever. Hale and Wright 1997 gives details. A generalization to multiply intensional type theories is given in Williams 2005, chap. 5, appendix C, and in Williams forthcoming b, appendix.

19. Of course, the interpretationist can fix things so that *sentences* about reference such as "'Billy' refers to Billy" is true. But this is irrelevant to the philosophical claim being made.

20. In Williams 2005, chaps. 6–7, I examine putative costs of accepting inscrutability of reference. The most persuasive of these—concerned with damaging interactions between radical inscrutability and deductive inference—is further developed in Williams forthcoming b.

Lewis grants (b). He is, however, committed to an interpretationist metasemantic theory. If the inscrutability arguments are not to engender paradox in his position, he must find fault with the argument for (a). It is to his response that we now turn.

2. The Eligibility Response to Inscrutability

Lewis is committed both to interpretationism and to rejecting radical inscrutability of lexical content. So he must find a flaw in the inscrutability argument just sketched. His favored response should be seen as composed of three moves. First, the correct semantic theory is the *best* one that accounts for the relevant data. Second, *fitting* with the data is but one kind of theoretical virtue a theory can have; other virtues, such as *simplicity*, can make one theory better than another. Third, simplicity is to be (at least partially) analyzed by appeal to objectively natural ("elite" or "sparse") properties. Lewis is independently committed to each component, and we shall see that when put together, his response to the inscrutability puzzles—the "eligibility constraint"—follows.

We begin by setting out the second and third components, as they emerge in Lewis's discussion of laws of nature.[21]

2.1. Additional Junk and Scientific Laws

Consider the ultimate theory T of microphysics, one which gives accurate predictions of the behavior of all subatomic particles. Contrast T with the theory T', which is just like T except for the addition of a "redundant" natural law: one that generates no new predictions about particular matters of fact. Suppose, for example, that it governs the behavior of particles under nomologically impossible circumstances: what an atom would do if it traveled faster than light, for example. Since no actual particles meet the conditions (we will suppose), both the putative law and its negation are consistent with all the local matters of fact that the world supplies.

If there were basic facts about the world that did not concern local matters about the distribution of fundamental properties in space and time—if there were robust "law-making facts" of the kind that Armstrong

21. Lewis (1999a [1983]) himself notes the analogy between his treatment of laws of nature and his eligibility response to inscrutability arguments, but this is not pursued in much detail. Making the analogy exact would require discussion of Humeanism about special sciences and the standing of novel vocabulary within Humean accounts (cf. Lewis 1994a).

(1983) postulates—then whether T or T' is the correct theory of the world would be settled by correspondence to reality.

But some theorists do not wish to postulate law-making exotica. For these theorists, the truth-makers for scientific theories must be found in the arrangement of matters of particular fact, rather than in abstruse ontology. Call this view *Humeanism*.

As we have already indicated, T and T' fit the matters of particular fact *exactly as well as one another*. If what it is for a scientific theory to be true, given Humeanism, is just to fit with matters of particular fact, then there is no distinguishing T and T'. Call this the *argument from additional junk*: it threatens to show that Humeanism about laws of nature will make it indeterminate whether or not the redundant law included in T' holds.

The response of the Humeans is that there is more to being a good theory of some range of data than simply being consistent with that data. Fitting with the appropriate range of data is a virtue of a theory, but there are other considerations besides. Among the additional virtues, for example, are simplicity—how economical and parsimonious the theory is; and strength—how many claims the theory commits itself to and how much of the data it predicts. For Humeans such as Lewis (1986b, introduction), the correct scientific theory of a range of data is the *best theory* of that data, where the best theory is one that has the optimal combination of simplicity, strength, and fit.

With the generalized notion of "best theory" in place, we have a recipe for resolving the puzzle over T and T'. Since T' is just T plus additional junk, it is less simple than T. Moreover, this loss of simplicity is not compensated by any gain in predicative power (strength) or descriptive adequacy (fit). T is the better theory, hence (if these are the only candidates we are to consider) it is the correct theory. The threat of indeterminacy from "additional junk" vanishes.

2.2. Simplicity and Naturalness

But what makes for simplicity? Prima facie, the notion seems subjective since it is tempting to think of simplicity as a "projection" of facts about what *strikes us* as simple. One might reasonably think, however, that incorporating a subjective notion in a crucial role within an account of laws of nature is suspect: it threatens to undermine the objectivity of scientific laws. In response to this worry, Lewis (partially) *analyzes* the notion of theoretical simplicity in terms of objective features of the theory.

The first thing that strikes one about the "junky" theory T' is simply that it contains an extra axiom. That is, the *axiomatized theory* is syntactically more complex. To address this, count an axiomatized theory as simpler than another if it has fewer, and syntactically less complex, axioms.[22]

This alone cannot resolve our puzzle. Consider, for example, the single property: *being such that T′ holds*. Let the predicate '*P*' denote this property. Now consider the theory T'', which consists of the single axiom, $\exists xPx$. In syntactic terms, it is clearly simpler than T and arguably matches it for strength and fit.[23] But clearly we do not want it to be the best overall scientific theory, or the whole enterprise will be trivialized.

To address this concern, we start by distinguishing between relatively *natural* properties: having spin 1/2, being green, being an animal, and so on; and relatively *unnatural* properties: being thought of by somebody, being grue (being green prior to *t*, blue after *t*), being the mereological sum of the left half of a human and the right half of a donkey, being such that T' is true. At the limit, we can distinguish the *perfectly natural* or *fundamental* properties from all the rest. (With Lewis, we shall continue to say that even the most unnatural such phrase denotes a property: properties in this sense are "abundant.")[24]

Lewis insists that, in evaluating a scientific theory for simplicity, the primitives of a scientific theory must pick out *fundamental* properties—the basic furniture of the world. Once this is done, we can fairly compare theories according to their syntactic complexity.

In order for this to contribute to an *objective* (partial) analysis of simplicity in terms of syntactic complexity, we need to postulate an *objective* distinction between the natural/nonnatural—between elite properties and abundant rubbish. There are good questions about the meta-

22. For the moment, we set aside the question of how to measure the syntactic complexity of an axiomatized theory.

23. See Lewis 1999a [1983]. The case is not decisive, however. For although the single axiom $\exists xPx$ *entails* T', it has little *deductive* power. The above point, therefore, requires that we characterize the "power" of a theory through its *entailments* rather than its *logical consequences*.

24. On Lewis's views, properties are sets of *possibilia*. More neutrally, I shall take abundant properties to be whatever entities play the role of semantic value of predicates, be this sets of actual individuals, sets of actual and nonactual individuals, functions from worlds to sets of individuals, or some other construction. In a context where we are committed to whatever ontology is required to underpin semantic theory, properties in this sense involve no new ontological commitments.

physics of this distinction, discussion of which we will not go into here.[25] If the metaphysics stands up, we can cash out at least some of our claims about one theory's being simpler than another in nonsubjectivist, nonrelativist terms. All else equal, one theory will be simpler than another if the first is syntactically less complex than the second, when each is spelled out in primitive terms. In particular, then, we can make the case that T' is less simple than T.

2.3. Naturalness and Eligibility

Now let us turn back to the case of semantic facts. The situation is analogous to the one we found in science. Various theories were available, all of which fit the relevant range of basic facts, and an unacceptable indeterminacy threatens. It is attractive to respond just as we did in the case of scientific theories: to hold that *fit* is but one among several theoretical virtues. To be the best theory of sentential data, a theory needs also to optimize the other theoretical virtues: in particular, simplicity. When offering accounts of the constitution of intentional facts, more than ever we must be wary of appealing to agents' *attitudes* to theories in characterizing simplicity—for our ambition is a *reductive* account of the intentional.[26]

Now recall Lewis's objectivistic (partial) analysis of the simplicity of a theory. Strictly, we look at how syntactically complex the theory is *when spelled out in primitive terms*, where "primitive terms" are required to

25. Armstrong (1978a, 1978b) provides a classic defense of the need for an objective distinction and advocacy of one particular form that such a distinction might take. Lewis (1999a [1983], 1986a, sec. 1.5) argues for the utility of the distinction and canvasses several forms that it might take without endorsing any particular account. Armstrong 1989 is a more recent survey and evaluation of the options. Van Fraassen (1989) disputes the entire framework of inegalitarianism concerning properties. Taylor (1993) argues for a *theory-relative* version of the distinction that will not secure the objectivity we need here.

26. There is perhaps wriggle room here. If one adopts what Lewis (1994b) calls the "headfirst" strategy to the intensional—giving first a characterization of *mental* content in nonintensional terms and then allowing appeal to the mental content within a foundational account of linguistic content—then there is no immediate circularity in appealing to agents' judgments of the simplicity of theories within a metasemantic theory. However (a) Lewis (1999b [1984]) wishes his metasemantic theory to be compatible with non-headfirst strategies, and (b) Lewis appeals to the same resources to deal with the threat of indeterminacy within his accounts of mental and of linguistic content. So both for Lewis himself, and for several other important versions of interpretationism, appealing to subjectively constituted constraints would induce *direct* circularity.

pick out the basic notions of fundamental microphysics. We might formulate a semantic theory by including axioms such as:

'is an atom' applies to something if and only if it is an atom.

But to evaluate the simplicity of a theory by Lewisian lights, we have to replace the term 'atom' as it is *used* by the theory (i.e., on the right-hand side of the above biconditional) by a characterization of this property in terms of the fundamental properties of physical science. This must be done, not only for scientific discourse, but also for the general run of natural language expressions: 'is red', 'is a human', 'is running', and so forth.[27]

Equivalently, we could assign a "degree of eligibility" to each semantic value for a lexical item featuring in the semantic theory—a measure that reflects the syntactic complexity of a clause that assigns that semantic value to an expression. For example, the degree of eligibility of the property *being human* gives a measure of how much syntactic complexity is added to semantic theory by a clause assigning that property to the predicate 'is human'. The overall eligibility of a theory is thus just another way of measuring the syntactic complexity of that theory *when spelled out in primitive terms*.[28] Measuring simplicity of a theory by its syntactic com-

27. We shall set aside doubts about whether finite definitions of macroscopic properties in microscopic terms are available (cf. Sider 1995). Also, we set aside worries based on the *vagueness* of natural language expressions in contrast with the precision of microphysical descriptions. I consider these and other objections in Williams 2005, chap. 8 and argue that no *decisive* objection to the proposal results. I regard the "Pythagorean" argument to be described below as bringing sharp focus to the general concerns here: if we can resolve that puzzle, it is likely that in so doing we will be able to lay to rest the general worries just mentioned.

28. I'm here supposing that the *only* aspect of variation between semantic theories will be in the clauses assigning semantic values to expressions. This works best in frameworks for semantic theory where almost all the work is done by the assignment of semantic values to expressions (e.g., the general semantics of Lewis 1970a). In such treatments, the only other part of semantic theory that potentially contributes to overall syntactic complexity is the compositional axiom.

In alternative settings, things vary somewhat. In traditional presentations of the semantics for a first-order language, for example, the interpretation of quantifiers is given through dedicated "syncategorematic" axioms. In Davidsonian truth theories, *all* clauses are syncategorematic.

In settings where the syncategorematic aspects of semantic theory vary among those in the running for being the "meaning-fixing theory," then rather than talking solely of the eligibility of *semantic values*, we need in addition to factor in the eligibility of the syncategorematic *clauses*—again, measuring how much syntactic complexity would be involved in writing these out in primitive terms.

plexity when spelled out in primitive terms, or by overall eligibility of the semantic values assigned, are thus one and the same thing.

We can then restate the result as follows: a semantic theory will be simpler to the extent that it is overall more *eligible*.

What we have reached is exactly Lewis's "eligibility" response to the inscrutability arguments.[29] First, this response views interpretationism as selecting the "best" (i.e., highest-scoring) theory, where *eligibility* is one of the factors relevant to gaining a high score:

> We have no notion how to solve the problem of interpretation while regarding all properties as equally eligible to feature in content. For that would be to solve it without enough constraints. Only if we have an independent, objective distinction among properties, and we impose the presumption in favour of eligible content *a priori* as a constitutive constraint, does the problem of interpretation have any solution at all. . . . [C]ontenthood just consists in getting assigned by a high-scoring interpretation, so it's inevitable that contents tend to have what it takes to make for high scores. . . . I've suggested that part of what it takes is naturalness of the properties involved. (Lewis 1999a [1983], 54–55)

Lewis focuses on the naturalness or eligibility of the *properties* assigned to predicates. The treatment above is a generalization of this, talking of the naturalness or eligibility of the semantic values assigned to any predicate.

Second, this response views eligibility as determined by the syntactic complexity of definitions of a properties in perfectly natural terms:

> Physics discovers which things and classes are the most elite of all; but others are elite also, though to a lesser degree. The less elite are so because they are connected to the most elite by chains of definability. Long chains, by the time we reach the moderately elite classes of cats and pencils and puddles; but the chains required to reach the utterly ineligible would be far longer still. (Lewis 1999b [1984], 66)

Lewis's idea that the eliteness (naturalness/eligibility) of a property is a matter of the length of *definitions* of those properties is exactly what we should expect if syntactic complexity of a theory T is measured by the

29. Lewis states that he owes the idea to G. H. Merrill.

"length" of a theory (say, the number of connectives it configures[30]) in primitive terms.[31]

(There are other passages, particularly in the paper, where a rather different account of the eligibility of theories may be thought to be in play. In particular, one might think that one's account of the naturalness of properties should directly assign properties *degrees of naturalness* that can be identified with their eligibility. It is doubtful that this is Lewis's view, and moreover it is open to independent objection. By contrast with the present interpretation, which sees Lewis's eligibility response as a special case of a general analysis of simplicity, brutely imposing an eligibility constraint of this form on choice of semantic theory seems an ad hoc "monster-barring" move.)[32]

In terms of the arguments for inscrutability, the upshot is this: a semantic theory may *fit* the relevant range of data about the content of sentences, and yet still not be the *best theory* of that data.

The kind of constructions required in order to generate the inscrutability paradox intuitively involve extremely *unnatural* properties. For in order to leave the semantic values of sentences invariant under a crazy reference scheme (e.g., one that assigns as reference to 'Billy' the ϕ-image of Billy), we need to assign a crazy extension as semantic value of the predicate 'runs' (e.g., saying that it applies to all x such that $\phi(x)$ runs—that is, it applies to those things which are such that their ϕ-image runs). But typically this will be a much less eligible extension than the set of

30. Harold Hodes made the nice point that there may be ways of measuring syntactic complexity of a theory other than by simply "counting the connectives" in the way alluded to above. For example, it is natural to reach for the resources of Komolgorov complexity theory at this point. Altering the details here should not change anything essential to the discussion, so long as "length of definitions" is understood throughout as a way of referring to the *complexity of definitions*.

31. Notice that here we will need to restrict the range of connectives in terms of which we formulate the theory or else the account will be trivialized. It is in keeping with the general spirit of the eligibility response to postulate an objective demarcation of "elite" logical functions for this purpose: for example, there may be universals of conjunction and negation, but no universal corresponding to a complex twenty-five-adic truth function. For critical discussion of this way of handling "syntactic complexity," see Sider 1995.

32. On the exegetical point: Lewis is officially neutral as to whether primitive naturalness is to be preferred to primitive contrastive resemblance or a theory of universals. But only the first of these accounts seems able to deliver "degrees of naturalness" directly. If the eligibility response was incompatible with a theory of universals or resemblance nominalism, it surely would have been noted by the author.

runners.[33] Interpretationists who claim that semantic facts are given by the best theory about the data can resist the arguments for radical inscrutability by arguing that the permuted interpretations will typically be *less eligible* than their "natural" rivals.

There are many areas in which more detail can be sought but which we have no space to pursue here. What are the metaphysics of these natural properties—is 'natural' to be taken as primitive, or can we define it, as Lewis at one point suggests, in terms of a theory of sparse universals or tropes? Is the characterization of syntactic complexity offered above tenable and appropriate? Is it plausible that properties like 'being human' are less than infinitely unnatural, in Lewis's sense (i.e., are finitely definable from a range of perfectly natural microphysical properties)?[34]

Some further questions apply equally to Humeanism about scientific laws as to interpretationism: Can we find a way of "adding together" individual virtues like simplicity, strength, and fit to get an overall virtue: the "goodness" of a theory?[35] Even given a distinction between more or less natural properties, are we entitled to the assumption that there are *perfectly natural* properties?[36]

2.4. Dialectical Success and the Lack of a Safety Result

Let us, pro tem, grant Lewis all the resources he wants. In particular, let us accept both his constraint that, all else equal, a good semantic theory must ascribe to predicates extensions that are at least as natural as its rivals, and his characterization of the naturalness of properties in terms of the minimal syntactic complexity of a definition of that property, in microphysical terms. Waive worries about the definability of the properties: suppose that every basic property (or property candidate) is capable of definition in a finite number of steps. Furthermore, grant that Lewis's response covers not just empirical cases, but also mathemati-

---

I am convinced by the arguments formulated in Sider 1996 that the primitive naturalness account faces severe objections (this can be avoided, it seems, only by the unattractive option of committing oneself to sui generis abundant relations). So attempting to reformulate the eligibility response in terms of primitive degrees of naturalness is objectionable, as well as ad hoc.

33. Recall that we were using 'property' (in the abundant sense) for the semantic values of predicates, so in the current context, we may speak of the eligibility of properties and extensions interchangeably.

34. Sider (1995) raises this concern.

35. See the introduction to Lewis 1986b.

36. Armstrong (1978b, chap. 15) denies that we are so entitled.

cal cases—the relation of addition, for example, is more natural than Kripke's quaddition.[37]

I take it that, given these concessions, Lewis has given us an effective answer to the challenge from the specific permuted reference schemes that we have mentioned. The crazy-reference schemes above will lead to a less natural assignment of extensions to predicates such as 'is a person'. Standardly, the entities falling under this predicate have the comparatively natural property of *being a person*; afterward, the entities are the ϕ-images of people, and this collection is likely to have no such unifying properties. The kinds of permutations we have considered are thus likely to lead to a less eligible overall semantic theory than the standard reference scheme.

The dialectical point generalizes. The point is that the inscrutabilist's constructions are *parasitic* on the original semantic theory: obtained by carefully chosen modifications. In virtue of this, prima facie the inscrutabilist's deviant semantic theory inherits all the complexity of the original theory, and then adds some more. Suppose we start from a "natural" interpretation of a language, which characterizes the extension of a primitive predicate P by *using* the metalinguistic expression f. Now consider the permutated interpretation based on ϕ. In the first instance, we can characterize P's new extension by using the metalinguistic compound $\phi^{-1} \circ f$. Suppose that there is a minimal definition of f in perfectly natural terms with n symbols. To escape the accusation that the permuted theory is less eligible than the original, the inscrutabilist will need to argue that (on average) there will be a characterization of the function $\phi^{-1} \circ f$ that has no more than n symbols. But the only characterization we have available, in the general case, will use at least $n + 1$ symbols. The general inscrutabilist argument thus fails to establish its intended conclusion, given the eligibility constraint.

Notice, though, that Lewis's response does not provide any *safety result*. The Lewisian cannot rule out the possibility that there is some characterization of the permuted extensions that is as economical as any available for the standard interpretation. Since nothing in the Lewis account rules out this possibility, nothing in Lewis's story gives a guarantee that every crazy permutation of the standard reference scheme will lead to a less eligible overall assignment of extensions to predicates. Might there not be some "symmetry" in the space of properties, so that when we map an object to its symmetrical twin, we find a pattern of

37. Kripke 1982; for Lewis's eligibility response, see Lewis 1999a [1983], 52–55.

instantiation of relatively eligible properties that exactly matches the original situation? That the only characterizations that the inscrutabilist has available will always be slightly more complex than the standard one is a good *dialectical* point, but it offers no guarantees against eligible unintended interpretations.

The lack of a safety result does not necessarily reinstitute a worry for the interpretationist. There are cases, indeed, where the gap is exploitable, and where arguably the permutation argument does go through. Two possible examples are, first, mathematical theories containing certain "automorphic" symmetries: the complex numbers are a case in point; second, in empirical scenarios where there is some symmetry in the universe—for example, where the universe is periodic.[38] However, even if these scenarios do generate inscrutability, it is not obvious that this would constitute an objection to Lewis's program. Maybe we should be happy with localized indeterminacy in abstruse mathematical cases.[39] Maybe we should not worry over *contingent* determinacy of interpretation, as one is confident that our universe is not symmetrical in one of these ways.[40] (For what it is worth, it seems to me that the eligibility constraint has the resources to rule out a great range of such interpretations.)[41] But even if

38. For the first case, see Brandom 1996. For the second, see Strawson 1959, chap. 1.

39. This might extend beyond mathematics: Quine's "argument from below" (Quine 1950, chap. 2), and even the ordinary vagueness that affects almost all of our terms, might be thought to provide reasons for thinking that certain restricted inscrutability results hold.

40. John Hawthorne has pressed these "restricted inscrutability" worries (in his comments on this essay for the *Philosophical Review* workshop). One of the examples he describes (the modal version of the "belief world" argument) differs from the indeterminacy induced by Strawsonian symmetries in threatening to generate damaging results for the semantic properties in the *actual world*.

41. The key thoughts are, first, that (at least as I have presented matters) the Lewisian should measure the *overall eligibility of semantic theory* rather than simply the eligibility of the properties assigned to predicates. Consequently, the eligibility of the semantic values assigned to singular terms is as relevant as the semantic values assigned to predicates. Second, the Lewisian should insist that the semantic values of, say, indexical expressions such as 'I' are their Kaplanian *characters* rather than their Kaplanian *contents*: they are functions from contexts to objects. Third, *the function from C to the image under symmetry S of whoever is speaking in C* is parasitic on *the function from C to whoever is speaking in C*, in the sense that, prima facie, any specification of the former will take just that little bit longer to write out in primitive terms than a specification of the latter.

Putting these together, the semantic value of 'I' on the sensible interpretation will be more *eligible* in Lewis's sense than the Strawsonian rival semantic value for 'I' (by 2 and 3); ipso facto the sensible semantic theory will be overall more eligible than its Strawsonian semantic theory (by 1). Such considerations generalize to Hawthorne's cases, I believe, but not to the Brandom mathematical case.

arguments for restricted inscrutability of the kinds mentioned are either undamaging or avoidable, however, the point remains that nothing in Lewis's story rules out there being some way of assigning extensions to predicates that will:

(A) match, or exceed, the simplicity of the "standard" assignments;
(B) fit the data that is given by the first part of the interpretationist story; and
(C) are crazy.

The examples given above illustrate (A) and (B), but they may not be crazy enough to worry the Lewisian. But I will argue that we can state exactly conditions under which all of (A), (B), and (C) are met.

3. An Alternative Argument for Radical Inscrutability

We now turn to an alternative way of arguing for inscrutability that is not easily deflected by the eligibility constraint. The result at the heart of the argument for radical inscrutability, due to Henkin (1949, 1950), is the following: if T is a syntactically consistent theory, then T has a model—that is, an interpretation under which every sentence is true.

Such a result is a promising resource for those interested in indeterminacy of interpretation, and they are discussed as such in Putnam 1980. Consider, for example, the global descriptivist variety of interpretationism described earlier. Its demand on successful interpretation is that it render our "folk theory of the world" true—that is, that the interpretation be a model of this theory. So, if the folk theory itself is consistent, Henkin's theorem guarantees that it will have a model.

These models need be nothing like the "intended" interpretation for discourse about medium-sized dry goods. Indeed, in the usual presentation of the result, the elements within the domain of the model are certain equivalence classes of expressions—clearly not the intended interpretation. There is, moreover, nothing to stop us setting up the model within the Henkin construction with any domain we choose, so long as the *size* is appropriate. So we could just as well build the model out of a domain of atoms, or of fish, or of abstract objects such as numbers.[42] Thus, the models that Henkin's constructions provide can be

42. Even if the domain of objects of the Henkin construction is that of the intended interpretation, we have a guarantee that the model construction by Henkin's methods is not the intended interpretation if Tarski's theorem on the undefinability of truth

clearly "unintended," but at least prior to the imposition of something like Lewis's eligibility constraint, the global descriptivist has no resources to disqualify them. Each such construction is "as good as" every other, when it comes to fitting with the total folk theory that constitutes the global descriptivist's data.

It is not surprising that when radical inscrutability of reference is in view, more attention has been paid to the *permutation arguments* discussed earlier than on arguments for inscrutability based on the metalogical results just mentioned. The permutation arguments are an almost trivial result, model-theoretically speaking, whereas the model-existence theorems are comparatively difficult to prove.[43] It is a mistake to ignore these alternative ways of arguing for inscrutability, however, for as we shall see, in the context of the Lewisian "eligibility" constraint, they form the basis for much more robust objections to interpretationisms than that resulting from permutation.

3.1. The Construction

I will now sketch how the Henkin model-existence theorem can be proven and describe how this can be turned into an argument for radical inscrutability of reference. The theorem is part of standard metalogic, so those prepared to accept the result on trust can safely skip this section and move on. I shall first describe abstractly how the construction works and then illustrate the ideas with a toy example.

Recall that the challenge is to construct a model for an arbitrary consistent (first-order) theory T. The Henkin procedure is as follows.

First of all, one modifies the theory T to obtain a consistent theory θ, which extends T and has various other useful properties. In particular, it is "negation complete" (for any sentence S in the language of the

---

holds for the targeted theory. For by a result of Bernays, the Henkin construction can be arithmetized, and we can extract a Δ^0_2 definition of truth in that model. Ex hypothesi, truth in the intended interpretation is undefinable, so the constructed model cannot be the intended interpretation. Thanks here to the anonymous referee for pointing toward these results and to Daniel Isaacson for discussion.

43. When radical inscrutability of reference is in question, permutation arguments might *appear* to do just as well as those based on more complicated methods—though this is disputed below. For other purposes, we might need to exploit the additional power of other argument forms. For Putnam, arguing that "ideal theory" will always be true, the permutation arguments will be of no help: the model-existence theorems are what is needed. Equally, for the arguments in favor of *quantificational* inscrutability given in Putnam 1981, the powerful, downward Löwenheim-Skolem theorem is required.

theory, θ contains *either S* or its negation), and it is "fully witnessed" (for every existential formula $\exists x \phi x$ it contains, it contains a "witnessing" formula ϕc, for some constant c).

To obtain a fully witnessed theory, we might have to extend the language in which *T* was given by adding new constants (consider a case where *T* is formulated in a language with no constants: if *T* contains existential formulae, the need to extend the language in order to get θ is obvious).

Suppose we could find a model *M* for θ (i.e., a domain + interpretation that makes every sentence of θ true). Since θ extends *T*, any model for the former is a model for the latter, so *M* is a model for *T*. Now, *M* provides an interpretation for a language that contains constants that didn't appear in the original language in which *T* was presented. But by restricting the interpretation to the original language, we obtain a "restricted" model for *T* in the original language.

It suffices, therefore, to, first, construct θ and, second, construct a model for this theory. To illustrate the general idea, let us consider a toy theory. Let us consider a "total theory" formulated in a constant-free language containing a single nonlogical predicate 'Dog'. The theory will consist just of the pair of sentences:

$\exists x$ Dog x
$\exists x \neg$Dog x

First, we introduce witnessing constants for the existentials, 'Fido' and 'Betsy', say. The resulting extended theory is (the logical closure of):

$\exists x$ Dog x
Dog (Fido)
$\exists x \neg$Dog x
¬Dog (Betsy)

Next, we extend the theory in a more-or-less arbitrary way to a theory *T**, such that for every sentence φ in the language containing the two new constants, *T** contains either φ or ¬φ. (Obviously, we choose one that will not introduce any inconsistency.) If new existentials are introduced, the language may no longer be fully witnessed, so we would have to iterate the procedure. For simplicity, suppose we choose an extension where this doesn't happen.[44]

44. In the general case, we may have to iterate the procedure *infinitely many times*. Effectively, we have two procedures: *closing* a theory and *witnessing* a theory. If we apply

By such means, we arrive at a theory θ that has the nice properties mentioned earlier: it extends our original theory *T*, and it is fully witnessed and negation complete. We then start to build a model for it. Again, I first describe the abstract framework and then illustrate it with our toy example.

The general case takes all *equivalence* classes of constant symbols, under the relation "$c = c'$ is a member of θ."[45] The set of all such equivalence classes will be the domain of the model we are constructing. The question then is simply how to assign elements of the domain to the constants, extensions to the predicates. The answer is simply to assign to each constant '*c*' the equivalence class of which it itself is a member, and to let an equivalence class γ be in the extension of a predicate '*F*' if and only if for some $c \in \gamma$ '*Fc*' is a member of θ. A few quick lines of proof then verifies that this construction is well defined and does indeed render each sentence in θ true.

In the context of our toy example, the domain that we build consists of two equivalence classes, $f = \{$'Fido'$\}$ and $b = \{$'Betsy'$\}$. We then let "Fido" refer to *f* and "Betsy" refer to *b*. Extensions are assigned to predicates in accordance with the description above. In this case, "Dog" will be assigned $\{f\}$, since "Fido is a dog" is in the theory and "Betsy is a dog" is not. It is easy to see in this particular case we have constructed a model for our extended theory, and that, restricted to the original language, we have a domain and interpretation that make true:

$\exists x$ Dog x
$\exists x \neg$Dog x

The above argument, generalized and formalized—a standard piece of metalogic—vindicates the bare model-existence theorem.[46] Paying attention to the details of how this is proved, we can see that we can strengthen the stated result in a small way, since there is nothing mathematically significant about the use of a linguistic ontology. We could use any set

these one after the other infinitely many times, the "summation" of all the resulting theories will turn out to be both fully witnessed and negation complete.

45. That this *is* an equivalence relation is guaranteed by the fact that θ is negation complete and consistent. Given this, we can prove that θ is "deductively closed with respect to finite sets of formulae." That is, if $\phi, \psi \vdash \chi$, and $\phi, \psi \in \theta$, then χ in θ. (Proof: since θ is negation complete, either χ or $\neg\chi$ must be in θ. But if $\neg\chi$ were in θ, then $\{\phi, \psi, \neg\chi\}$ would form an inconsistent subset of θ, contradicting consistency. So χ must be in θ.)

46. Details can be found in the section on "compactness theorems" in Zilber 2000.

of objects *C* whose elements can be mapped in a 1–1 fashion onto the equivalence classes of terms described above. By paralleling Henkin's construction, we can then build up a model for θ whose domain of quantification consists exactly of the members of *C*. In our toy example, we could start with {0, 1} rather than {*f*, *b*}. Within the construction, we then put 0 wherever *f* features above, and 1 wherever *b* is. Indeed, for *any* putative assignment of objects to constant symbols that doesn't directly contradict the identities featuring in θ, we can build a model that depicts that assignment as the reference scheme.

These latter modifications are trivial variants on the Henkin-style construction: just picking isomorphic copies of the model that the Henkin procedure provides. One effect is to enable us to find models embedding arbitrary reference schemes so that we can use the result within an argument for the inscrutability of reference.

3.2. Why It Pays Not to Be a Parasite

We judged Lewis's response to the inscrutability arguments to be *dialectically effective* against permutation inscrutability arguments. Our only grip on the complexity of the permuted interpretation is parasitic on an assumed "standard" interpretation. So the description of the new interpretation will, in general, involve additional complexity.

The dialectical situation changes if we use arguments for radical inscrutability that do not have the parasitic character of the permutation arguments. The Henkin strategy sketched in this section builds "deviant" interpretations matching the interpretationist's data independently of what the "standard interpretation" may be.[47] Thus we have two interpretations, one built up out of a domain of arbitrary elements, and we have no grip on how the syntactic complexities of the two compare.

This should itself disturb the interpretationist. For the inscrutability argument at this point threatens to issue in stand-off—though the inscrutabilist has as yet no grounds for claiming his or her twisted theories match the intended theory on grounds of eligibility, neither does his or her opponent have any grounds for *denying* they do so. The stand-off is unhappy for the interpretationists, for their aim was to *secure* scrutable reference, not leave the matter undecided.

47. We can, for example, deploy it to build models of consistent but "ω-inconsistent" theories, such as arithmetic with the negation of the Gödel sentence adjoined, which arguably *have no* intended interpretation.

In fact, on the global descriptivist version of interpretationism, we can make the worry sharp. The final section of this article will exploit this to make a case that, unless substantial metaphysical concessions are granted, the eligibility constraint on interpretation will lead to absurd results when applied to the kind of Henkin constructions just described.

4. Pythagorean Worlds

Suppose we adopt a global descriptivist form of interpretationism. The data constraining semantic theory in this case is a "total theory" that the semantics must render true—or equivalently, a pairing of sentences with truth values that the semantics must match. "Total theory," recall, was the sum total of all the platitudes gathered from every walk of life—all the sentences that are too obvious to question. If total theory is consistent, we know that we can find *models* for that discourse since we have just seen how a standard metalogical result—the Henkin construction—gives us a recipe for constructing such models. For simplicity let us suppose that total theory can be formulated in first-order terms.

We will be considering a case where this "total theory" is consistent with there being only finitely many things. For example, the total theory will not commit itself to a plenitude of abstracta, or to infinitely divisible regions of substantival space or time. (If one maintains that *our* folk theory of the world contains such commitments, consider a more cautious community who are at best agnostic about such matters.) Given this, we may consider a consistent theory that results from *adding* "there are exactly n things" for an appropriate n, to total theory.

The Henkin construction gives us a model for this theory whose domain contains n objects. As mentioned earlier, the Henkin construction is indifferent to the identities of the objects within the domain of the model: we can, for example, take them to be (an initial segment of) the natural numbers. Hence, a Henkin construction enables us to find a model for the above theory whose domain consists of the numbers less than n. Call this the *arithmetical model* of total theory.[48]

48. Timothy Bays, in discussion at the *Philosophical Review* workshop, highlighted what appears to be a tension in this presentation. The "total theory" of the world on which the interpretationist goes to work has to be finitely satisfiable, yet the model that the interpretationist constructs appeals to *numbers* (albeit finitely many of them). So prima facie the interpretationist appears to use a theory—arithmetic—which commits her to the existence of infinitely many things.

As anticipated above, we now have a situation where two interpretations of our language—the intended one and the arithmetical one—are in competition to be the "reference-fixing theory." Again, as anticipated, the eligibility constraint has nothing *directly* to say about this situation. Unlike the deviant interpretations generated by permutations we considered earlier, our deviant interpretation is not specified *derivatively* from the intended interpretation, so there is no quick prima facie argument for the intended interpretation beating the deviant one in point of eligibility. On the one hand, we can expect the extensions assigned to predicates on the arithmetical interpretation will in general not seem very "naturally unified." On the otherhand, the intended interpretation will not typically assign to the predicates of natural language perfectly natural properties, by Lewis's lights: the "lengths of definitions" required to get from the fundamental properties of microphysics to such macroscopic properties as *being a table*, *being human*, and *being a stone* are enormous (supposing such definitions to be possible at all).

But the situation does not remain a stand-off. In what follows, I shall argue for the existence of *Pythagorean* twins of the actual world: worlds that are similar to the actual world at macroscopic levels, but in which the arithmetical interpretation *beats* the "intended" interpretation in point of eligibility.[49] I then argue that the existence of such Pythagorean worlds leads to deep problems: ones sufficient to reject the eligibility constraint if major alterations or concessions are not made.

I think this highlights *dialectical* limitations in the argument as presented. One may not be able to use the arithmetical model to convince *oneself* that *one's own* language has an arithmetical model of the kind just sketched—absent subtle argument, either you reject arithmetic and thus arithmetical construction, or else your "total theory" commits you to too many things for the construction to be applicable. For exactly the same reason, the argument may not be able to persuade *another* that the arithmetic interpretation applies to their language. But it seems to me that when it is the language of a *third person* that is under consideration, we need not worry: you and I may agree that invoking arithmetic to construct models for the language of a third party is legitimate, even if that third party is at best agnostic about whether arithmetical models exist.

49. See Quine 1964 for discussion of Pythagorean interpretations of language in the context of criteria for ontological reduction. Notice that I will not argue for worlds where the arithmetical interpretation is the most eligible interpretation, only that it beats the any "sensible" interpretation on this basis. Whatever the most eligible interpretation ends up being, in those worlds, it will be hugely unintended.

4.1. The Henkin Construction as Giving a Benchmark

Let us look again at what the arithmetical model provides. It gives a way of interpreting folk theory, but more than this, any community with the same "global folk theory" as us are interpretable via the same arithmetical model, no matter how their world otherwise differs from ours. In particular, it seems clear that a world that has the same macroscopic structure as ours, or one that duplicates ours "from the quarks up," will be a world where the relevant "global folk theory" is the same. The arithmetical model is a candidate for interpreting their language just as it is for ours. The effect of the Henkin construction, therefore, is to enable us to provide an interpretation whose fit with total theory is *quite independent* of what might or might not be going on at extreme microscopic levels.

Now recall that we were considering a case where the total theory was compatible with there being only finitely many things, and consequently the arithmetical model could be chosen so as to include only finitely many things (numbers) in its domain. The extension of each predicate in the language is finite, therefore, and could in principle be "brutely" specified simply by enumerating the items in the domain falling under it. Since there are n objects in the domain, we can specify the extension of an arbitrary predicate under the arithmetical model in a clause with no more than n conjuncts. Repeating this for *all* the nonlogical vocabulary of the theory, we could in principle give in this way a finite overall specification of the arithmetical interpretation. We can suppose that this long "brute" specification of the model features m connectives.

On the Lewis account sketched earlier, we measure the eligibility of a theory by looking at how complex the interpretation would be when specified in perfectly natural terms. That amounted to measuring how syntactically complex the theory would be when spelled out in perfectly natural terms. What we have just seen is that, in the case of the arithmetical model, we can find an upper bound for this complexity: m.[50]

The Henkin construction and the intended interpretation are in a race for being the "best theory." For the global descriptivist, they both meet the constraints of "fit" and "predicative power," since ex hypothesi both provide a model for the "folk theory" of the world. Which theory is

50. This supposes the mathematical vocabulary involved to be "perfectly natural." This might be questioned, but so long as there is *some* finite specification of mathematical vocabulary in perfectly natural terms, which does not vary from world to world, the overall point will not be affected.

better comes down to which is simpler—under Lewis's handling of this, we are to look to their relative *eligibility*. We have just seen that we can fix the eligibility of the arithmetical model provided by the Henkin construction (i.e., the minimal number of connectives in a specification of the interpretation in fundamental terms) as *m*, say. Whether or not the intended interpretation "beats" the deviant one is then entirely a matter of whether or not the number of connectives in a minimal specification of the intended interpretation in perfectly natural terms is fewer than *m*.

Now we are in a position to see the danger for the global descriptivist. The Henkin construction sets a benchmark for eligibility that the intended interpretation must match. The threat is no longer inscrutability but something worse: if the complexity of the intended interpretation is greater than the benchmark, then it will *determinately* be the case (on Lewis's account) that the intended interpretation is not appropriate to our language. At the extreme, if the arithmetical construction is *optimal*, we would secure relative *determinacy* of interpretation of our language—but our names for each other, for example, would determinately refer to *numbers*.

4.2. The Existence of Pythagorean Worlds

One feature of Lewis's picture of the world is that "perfectly natural" or "fundamental" properties are found only at the *microphysical* level—on this view fundamental properties correspond to the primitive notions of a completed microphysics.[51] Expressed in terms of a theory of universals, Lewis's view is that universals are *ultra sparse*—not to be found at any macroscopic level, or even at most of the microscopic ones. There are no universals holding of all and only humans, or of all and only instances of a particular chemical kind, for example (cf. Lewis 1986b). Exploiting this assumption, we can argue positively for the existence of Pythagorean worlds.

In our world, let us suppose, atomic nuclei are made out of protons and neutrons, which are in turn made out of quarks. Only this bottom layer is a repository for universals, on the ultra-sparse conception. Presumably, that the microstructure of the world takes this shape is metaphysically contingent. Consider therefore a possible world that replicates ours from the "quarks up" but that has quark-counterparts composed of

51. Presumably, though, fundamental relations in metaphysics will have to be allowed also: for example "constitution," "part of," and so forth.

yet more basic particles, the subquarks. The ultra-sparse conception now places the universals *below* the quark-counterparts.

We can iterate this procedure, each time imagining underlying "layers" to reality. On the ultra-sparse conception of universals, such underlayering pushes the domain of perfectly natural properties further and further away from the macroscopic. Moreover, each such iteration decreases the eligibility of properties, such as *being human*, since the long chains of definition formulated in perfectly natural terms now require extra clauses.[52]

One strange effect of the Lewis characterization of eligibility, thought of as a measure of the simplicity of a theory, is that a theory will become *less simple* as its *subject matter* becomes more complicated. Indeed, the process described above allows us to decrease the eligibility of all of the properties that feature in the "intended interpretation" *without limit*. For any number *N*, we can find a world like ours from the "quarks up" with sufficient microstructure that the syntactic complexity of the intended interpretation, presented in fundamental terms, is more than *N*. In particular, we can choose a world of sufficient microstructure that the syntactic complexity of the "intended" interpretation exceeds the syntactic complexity *m* of the arithmetical interpretation. All the worlds we are discussing are like ours from the "quarks up," so, as discussed above, the same "total theory" will feature in each. Hence, such worlds will be Pythagorean.

The argument for the existence of Pythagorean worlds relies on the Henkin construction, and we have here formulated that only for first-order languages. Such languages are relatively unexpressive, though: arguably, to interpret natural language, we would require higher-order

52. It is a nice question in metaphysics whether a world whose fundamental microstructure "goes further down" in this way can really still be composed of *quarks*. Supposing, for the sake of argument, that *being a quark* is a fundamental property of the actual world, supposing it to be instantiated by the quark-counterparts in the underlayered world would seem to require that the naturalness or fundamentality of *being a quark* could vary from world to world. But one might well be skeptical about whether the naturalness of a property can be a contingent matter in this way. However, nothing in what follows need wait on an answer to this question; we can assume for the sake of argument that the (nonspatiotemporal) properties instantiated in each world are alien to every other world in the series. Strictly speaking, then, perhaps we should say, not that the eligibility of the property *being human* decreases in the series of worlds described, but that the eligibility of the property that plays the *human* role decreases from world to world in the series. I continue to write in the loose way, for the sake of simplicity. I'm grateful to John Hawthorne for raising this issue in the *Philosophical Review* workshop.

and multiply intensional resources. (Cf. Lewis 1970a, Cresswell 1973, Montague 1970, Partee 1996.)

On the other hand, the only feature of the Henkin construction that we required was the construction of a model for "total theory" T that could be specified by clauses of finite complexity, where the recipe for constructing the model for T was independent of the contingent setup of the world.

To begin with, then, we can appeal to the Henkin constructions that are possible for simple type theories (Henkin 1950). The techniques involved are generalizations of those described here, and the effect is the same for our purposes. So the need for higher-order resources in describing natural languages does not alter our conclusion.

We do need to strengthen our assumptions slightly in order to handle the intensional cases. In order to deal with, for example, positive possibility claims 'possibly p' featuring in total theory, our model will need to appeal to an extra "world" relative to which p is true. The semantic value assigned to a predicate would then need to be a *possible-world intension* rather than simply its extension at the actual world.

Our result will not be significantly changed so long as we can still specify this intension in a finite way by enumeration. To secure this, however, we may need to place an additional constraint on total theory. In the original case, we require agnosticism over whether or not there are infinitely many things; the analogous constraint is to require that total theory is agnostic whether or not there *could have been* more than N things, for some N: only given this condition will we have finite extensions for our predicates with respect to each world. Since all we are asking for is *agnosticism* on such points, I regard this as reasonable.[53]

The substantive point is that so long as we *somehow* obtain a finite specification of a model in a way *independent* of how the world is set up at a microlevel, we will always be able to consider macroscopic duplicates of our world, where the "intended interpretation," given in perfectly natural terms, requires a longer specification.

53. We also require that the theory have a model with only finitely many "indices" (for example, worlds).

4.3. Three Problems Arising

I shall now suppose that the existence of Pythagorean twins of the actual world is established. We now weigh the costs.

The result is clearly disturbing: worlds that are indiscriminable from ours at macroscopic levels, but where our words refer to *numbers* rather than what we would standardly take to be their referents. I see at least three direct problems.

First, and most obviously, we have a violation of an extremely intuitive principle that (some) semantic facts supervene on the *macroscopic* structure of the world. Worlds that are structurally like ours, but have different microphysical constituents might well engender twin-earth-style differences in representational content, but there is no precedent for the kind of content change here envisaged.[54] Here we have a people behaving just as we do, within a world that is "well behaved" and just like ours in every detail from the quarks up; yet they refer to numbers where we refer to the objects around us. The suggestion beggars belief.

Second, the actual world may itself be Pythagorean. The complexity of the arithmetical interpretation is enormous—but so are the lengths of definitions that would relate macroscopic structures to fundamental microphysics. I see no grounds for thinking that the first of these large numbers turns out, in the actual case, to be greater than the second. Clearly *this* result would be a reductio of the interpretationist case: it is even more incredible than the inscrutabilist theses from which we started.

Third, even if the actual world turns out *in fact* to be non-Pythagorean, its being Pythagorean is a nonskeptical epistemic possibility. One route to this conclusion is just by noting the observation made above: that we ourselves have no grounds for confidence that the intended interpretation will beat the arithmetical interpretation. Another way of putting this: imagine a string of worlds, where each successive world has an additional "layer" of microscopic substructure. By the above argument, at *some* point in this string, the distance between macroscopic properties and ultimate microphysical properties becomes

54. In particular, it is natural to take it that the *A*-intensions or *linguistic meaning* of linguistic items should supervene on macroscopic structure. But this would be violated by the present case. In the Pythagorean world, *A*-intensions as well as "horizontal content" (*C*-intensions) are abandoned in favor of extensions drawn from the natural numbers. The point here is that the *whole semantic theory*—including the functions from contexts to horizontal content—is to be determined by best fit with appropriate data. In Pythagorean worlds, arithmetical extensions beat all such theories.

Thanks to Bob Stalnaker for pressing me on this point.

great enough that the arithmetical interpretation beats the intended interpretation on grounds of eligibility. However, as things stand, we have no knowledge where this cut-off point comes; in our current epistemic state, it might be that a world that bottoms out at the level of quarks has a large enough gap between the macroscopic and the fundamental level to be Pythagorean.[55]

This last conclusion is extremely disturbing. Part of our reason for rejecting radical inscrutability in the first place was to preserve epistemic access to semantic facts. We have just seen considerations that lead to the conclusion that, even if semantic facts *do* happen to obtain, we have no reason to think that they do, and consequently no knowledge of any such facts.

5. Conclusion

This article falls into three parts. In the first, we outlined both the attractions of an *interpretationist* metasemantic theory and the ways that inscrutability puzzles create problems within this setting. In the second, we looked at Lewis's eligibility response. As presented, this is no ad hoc paradox-barring maneuver, but a constraint on best interpretation motivated directly from general ideas about how to give an objective analysis of the theoretical virtue of simplicity. The framework already in place in Lewis's Humean account of laws of nature, when applied to the case of semantic theory, yields the eligibility response directly.

The eligibility response thus has much going for it: it is an independently motivated way of blocking the paradoxes while remaining faithful to the guiding interpretationist idea that semantic facts are fixed by best theory of a suitable range of data characterized at the level of sentences. Strategically, then, it is a sound line for the interpretationist to take. Moreover, it is a dialectically effective response to the most popular argument for radical inscrutability, one based on permutations of the intended interpretation.

55. An alternative route to this conclusion would be to argue for the claim that we do not, and could not, ever know whether we have reached the fundamental level. If, for all we know, any one of the string of "underlayered" worlds described above could be actual, then the epistemic possibility of Pythagorean worlds is established. Schaffer (2003) argues that in our current state of knowledge we have no reason to think there even is a fundamental level. His arguments can be adapted to my purpose here, though I think the kind of world in which *he* is primarily interested—a world of "infinite descent" with no fundamental level at all—raises quite different issues.

As we have seen, there are other ways to argue for radical inscrutability for which the eligibility response gives no such easy response. The model-existence theorems of Henkin (1949, 1950) allow us to argue for results just as disturbing as radical inscrutability, even granted constraints of eligibility. It is an epistemic possibility that the actual world is *Pythagorean*: that an interpretation that depicts our words as picking out integers matches the intended interpretation on grounds of fitting with "total theory" and beats it on grounds of eligibility.

I have argued that Lewis's global descriptivist is committed to Pythagorean worlds, and the resulting costs are unsustainable. But what resources has the interpretationist in general for avoiding the conclusion?

The first response is simply to move to a different form of interpretationism. I have not argued here that Lewis's convention-based interpretationism faces these problems. It is not altogether obvious that such a theorist can *avoid* analogous arguments for Pythagoreanism described above; but the issues are involved, and I will not examine them here.[56]

More directly, we can note the role played in the argument above by Lewis's use of *perfectly natural properties*, which he assumes inhabit only the "fundamental level" of reality: in the framework of a theory of universals, this is the assumption that universals are "ultra sparse"—corresponding only to the framework notions of a completed microphysics.[57] This assumption is crucial to the argument for Pythagorean worlds, for if there can be ontologically emergent universals at a relatively macrolevel, there is no longer an argument that "underlayering" reality will increase the logical distance between standard macroproperties and those that are perfectly natural or that correspond to universals.

56. However, I argue that there is especial interest in the global descriptivist form of interpretationism. In particular, it is not committed to the "headfirst" strategy that Lewis favors—Lewis's convention-based approach makes essential appeal to contentual mental attitudes. The flexibility of global descriptivism—not shared by its rivals—is especially evident when we begin to consider alternative ways of picking out the "total theory" on which it relies. One might, for example, regard total theory as the set of sentences of the language of thought that are (stably) in the belief box. Such resources are among those admitted by Fodor (1987) in framing the problem of intentionality for thought. If we had an adequate form of global descriptivism, then, it seems it could fit into an (otherwise) broadly Fodorian theory of mind and language. It is hard to see how one could adapt other forms of interpretationism to provide a theory of the semantic properties of a language of thought.

57. Together with, presumably, basic logical and metaphysical notions such as "part of," "constituted by," and so on.

Thus we have a remedy, but it comes at a severe cost: what looks like an ontologically extravagant appeal to emergent property ontology or its analogue within other ways of explicating the divide between perfectly natural properties and the rest.[58] The appeal would, of course, be less extravagant if one were independently committed to "macrolevel" universals or something equivalent—perhaps in order to secure an analysis of macroscopic similarity, causation, and laws.[59] However, it is embarrassing, to say the least, that what looks to be a local problem for the metaphysics of meaning should require such substantial commitment.[60]

We can expect a familiar dialectic to ensue. If one wishes to preserve the attractive package of interpretationist metasemantic theories and the eligibility response to inscrutability arguments, a prima facie case has been made for the need for a certain kind of additional ontology. One now looks around for something that can do the same work, at a cheaper cost.

Recall the original need for perfectly natural properties: it was to avoid triviality results when explicating the "elegance" or "simplicity" of an axiomatized theory in terms of syntactic complexity. Appeal to perfectly natural properties puts limits on what predicates we could use in measuring the complexity of the theory. If we had some other objective way of distinguishing the "interesting," relatively macroscopic properties from gerrymandered ones, we could have it play this role within Lewis's account, and the need for perfectly natural properties at a macrolevel would be alleviated. I leave it to the ingenuity of others to think of ways of drawing this distinction without falling into circular appeal to intentional notions, or extravagant appeal to additional metaphysical resources.

58. Lewis 1999a [1983] discusses several ways of making out a property "inegalitarianism." On the side of property ontology, there are theories of sparse universals or sparse tropes; without additional ontology one could appeal to metaphysically primitive ideology: a primitive predicate 'perfectly natural' applied to properties, or a primitive contrastive resemblance relation between individuals. In each case, we can distinguish ultra-sparse versions of the view from versions that allow "perfectly natural" properties at a macrolevel. For example, on the resemblance nominalist version, we either allow or do not allow macroscopic objects to be related by the primitive resemblance relation.

59. Schaffer (2004) argues for emergent universals on exactly this kind of basis.

60. Moreover, there are knock-on effects of endorsing macroscopic, perfectly natural properties. For example, in the light of considerations adduced by Barnes (2005), it is hard to see how one could avoid metaphysical vagueness other than by restricting perfectly natural properties to the ultra-sparse level—and the coherence of metaphysical vagueness is hotly disputed.

What is clear is that substantial reworking is needed if the line of response to inscrutability is to be saved. Unless we patch the eligibility response, Lewis's remedy for inscrutability threatens to be worse than the disease.

References

Armstrong, D. M. 1978a. *Nominalism and Realism*. Vol. 1 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press.

———. 1978b. *A Theory of Universals*. Vol. 2 of *Universals and Scientific Realism*. Cambridge: Cambridge University Press.

———. 1983. *What Is a Law of Nature?* New York: Columbia University Press.

———. 1989. *Universals: An Opinionated Introduction*. Boulder, CO: Westview.

Avramides, A. 1997. "Intention and Convention." In *A Companion to the Philosophy of Language*, ed. C. Wright and B. Hale, 60–86. Blackwell: Oxford.

Barnes, E. 2005. "Vagueness in Sparseness: A Study in Property Ontology." *Analysis* 65:315–21.

Brandom, R. 1996. "The Significance of Complex Numbers for Frege's Philosophy of Mathematics." *Proceedings of the Aristotelian Society* 96:293–315.

Cresswell, M. 1973. *Logics and Languages*. London: Methuen.

Davidson, D. 1967. "Truth and Meaning." *Synthese* 17:304–23. Reprinted in D. Davidson, *Inquiries into Truth and Interpretation*, 17–36. Oxford: Oxford University Press, 1980.

———. 1977. "Reality without Reference." *Dialectica* 31:247–53. Reprinted in D. Davidson, *Inquiries into Truth and Interpretation*, 215–26. Oxford: Oxford University Press, 1980.

———. 1979. "The Inscrutability of Reference." *Southwestern Journal of Philosophy* 10:7–19. Reprinted in D. Davidson, *Inquiries into Truth and Interpretation*, 227–42. Oxford: Oxford University Press, 1980.

Field, H. H. 1972. "Tarski's Theory of Truth." *Journal of Philosophy* 69:347–75. Reprinted in H. H. Field, *Truth and the Absence of Fact*, 3–29. Oxford: Oxford University Press, 2001.

———. 1975. "Conventionalism and Instrumentalism in Semantics." *Noûs* 9:375–405.

———. 1978. "Mental Representation." *Erkenntnis* 13:9–61. Reprinted in H. H. Field, *Truth and the Absence of Fact*, 30–67. Oxford: Oxford University Press, 2001.

Fodor, J. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Hale, B., and C. Wright. 1997. "Putnam's Model-Theoretic Argument against Metaphysical Realism." In *A Companion to the Philosophy of Language*, ed. C. Wright and B. Hale, 427–57. Oxford: Blackwell.

Henkin, L. 1949. "The Completeness of the First-Order Functional Calculus." *Journal of Symbolic Logic* 14:159–66.

———. 1950. "Completeness in the Theory of Types." *Journal of Symbolic Logic* 15:81–91.

Kaplan, D. 1989. "Demonstratives." In *Themes from Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein, 481–563. New York: Oxford University Press.

Kripke, S. A. 1982. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

Larson, R. K., and G. Segal. 1995. *Knowledge of Meaning*. Cambridge, MA: MIT Press.

Lewis, D. K. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

———. 1970a. "General Semantics." *Synthese* 22:18–67. Reprinted with postscript in D. K. Lewis, *Philosophical Papers*, 1:189–229. Oxford: Oxford University Press, 1983.

———. 1970b. "How to Define Theoretical Terms." *Journal of Philosophy* 67:427–46. Reprinted in D. K. Lewis, *Philosophical Papers*, 1:78–95. Oxford: Oxford University Press, 1983.

———. 1974. "Radical Interpretation." *Synthese* 23:331–44. Reprinted in D. K. Lewis, *Philosophical Papers*, 1:108–18. Oxford: Oxford University Press, 1983.

———. 1980. "Index, Context, and Content." In *Philosophy and Grammar*, ed. S. Kanger and S. Öhman, 79–100. Dordrecht: Reidel. Reprinted in D. K. Lewis, *Papers on Philosophical Logic*, 21–44. Cambridge: Cambridge University Press, 1998.

———. 1983 [1975]. "Language and Languages." In D. K. Lewis, *Philosophical Papers*, 1:163–88. Oxford: Oxford University Press. Originally published in *Minnesota Studies in the Philosophy of Science*. Vol. 7, *Language, Mind, and Knowledge*, ed. Keith Gunderson, 3–35. Minneapolis: University of Minnesota Press.

———. 1986a. *On the Plurality of Worlds*. Oxford: Blackwell.

———. 1986b. *Philosophical Papers*, vol. 2. Oxford: Oxford University Press.

———. 1992. "Meaning without Use: Reply to Hawthorne." *Australasian Journal of Philosophy* 70:106–10. Reprinted in D. K. Lewis, *Papers on Ethics and Social Philosophy*, 145–51. Cambridge: Cambridge University Press, 1999.

———. 1994a. "Humean Supervenience Debugged." *Mind* 103:473–90. Reprinted in D. K. Lewis, *Papers on Metaphysics and Epistemology*, 224–47. Cambridge: Cambridge University Press, 1999.

———. 1994b. "Reduction of Mind." In *A Companion to the Philosophy of Mind*, ed. S. Guttenplan, 412–31. Oxford: Blackwell. Reprinted in D. K. Lewis, *Papers on Metaphysics and Epistemology*, 291–324. Cambridge: Cambridge University Press, 1999.

———. 1999a [1983]. "New Work for a Theory of Universals." In D. K. Lewis, *Papers on Metaphysics and Epistemology*, 8–55. Cambridge: Cambridge University Press (originally published in *Australasian Journal of Philosophy* 61: 343–77).

———. 1999b [1984]. "Putnam's Paradox." In D. K. Lewis, *Papers on Metaphysics and Epistemology*, 56–77. Cambridge: Cambridge University Press (originally published in *Australasian Journal of Philosophy* 62:221–36).

McGee, V. 2005. "Inscrutability and its Discontents." *Noûs* 39:397–425.

Montague, R. 1974 [1970]. "Universal Grammar." In *Formal Philosophy: Selected Papers of Richard Montague*, ed. R. Thomason, 222–46. New Haven, CT: Yale University Press (originally published in *Theoria* 26:373–98).

Nolan, D. 2005. *David Lewis*. Montréal: McGill-Queen's University Press.

Partee, B. H. 1996. "The Development of Formal Semantics in Linguistic Theory." In *Handbook of Contemporary Semantic Theory*, ed. S. Lappin, 11–38. Oxford: Blackwell.

Putnam, H. 1980. "Models and Reality." *Journal of Symbolic Logic* 45:421–44. Reprinted in *Philosophy of Mathematics: Selected Readings*, 2nd ed., ed. P. Benacerraf and H. Putnam. Cambridge: Cambridge University Press, 1983.

———. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.

Quine, W. V. 1950. *Methods of Logic*. New York: Holt.

———. 1960. *Word and Object*. Cambridge, MA: MIT Press.

———. 1964. "Ontological Reduction and the World of Numbers." *Journal of Philosophy* 61:209–16. Reprinted with substantial changes in W. V. Quine, *The Ways of Paradox and Other Essays*, rev. and enlarged ed., 212–20. Cambridge, MA: Harvard University Press, 1976.

Schaffer, J. 2003. "Is There a Fundamental Level?" *Noûs* 37:498–517.

———. 2004. "Two Conceptions of Sparse Properties." *Pacific Philosophical Quarterly* 85:92–102.

Schiffer, S. R. 1972. *Meaning*. Oxford: Oxford University Press.

Sider, T. 1995. "Sparseness, Immanence, and Naturalness." *Noûs* 29:360–77.

———. 1996. "Naturalness and Arbitrariness." *Philosophical Studies* 81:283–301.

Stalnaker, R. 1999. *Context and Content*. Oxford: Oxford University Press.

Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen.

Taylor, B. 1993. "On Natural Properties in Metaphysics." *Mind* 102:81–100.

van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Clarendon.

Wallace, J. 1977. "Only in the Context of a Sentence Do Words Have Any Meaning." In *Midwest Studies in Philosophy*. Vol. 2, *Studies in the Philosophy of Language*. Morris: University of Minnesota Press.

Williams, J. R. G. 2005. "The Inscrutability of Reference." PhD diss., University of St. Andrews.

———. Forthcoming a. "Permutations and Foster Problems: Two Puzzles or One?" *Ratio* 21, no. 1.

———. Forthcoming b. "The Price of Inscrutability." *Noûs*.
Williamson, T. 1994. *Vagueness*. London: Routledge.
———. 2003. "Everything." *Philosophical Perspectives* 17:415–65.
———. 2006. "Must Do Better." In *Truth and Realism*, ed. P. Greenough and M. Lynch. Oxford: Oxford University Press.
Zilber, B. 2000. "Lecture Course on Model Theory." Mathematical Institute, University of Oxford, www.maths.ox.ac.uk/~zilber/lect.pdf.