Journal for the History of Analytical Philosophy

Volume 1, Number 10

The Role of Naturalness in Lewis's Theory of Meaning

Brian Weatherson

Many writers have held that in his later work, David Lewis adopted a theory of predicate meaning such that the meaning of a predicate is the most natural property that is (mostly) consistent with the way the predicate is used. That orthodox interpretation is shared by both supporters and critics of Lewis's theory of meaning, but it has recently been strongly criticised by Wolfgang Schwarz. In this paper, I accept many of Schwarz's criticisms of the orthodox interpretation, and add some more. But I also argue that the orthodox interpretation has a grain of truth in it, and seeing that helps us appreciate the strength of Lewis's late theory of meaning.

The Role of Naturalness in Lewis's Theory of Meaning

Brian Weatherson

It is sometimes claimed (e.g., by (Sider 2001a; Sider 2001b; Sider 2012 sec 3.2; Stalnaker 2004; Williams 2007; Weatherson 2003)) that David Lewis's theory of predicate meaning assigns a central role to naturalness.[1] Some of the people who claim this also say that the theory they attribute to Lewis is true. The authors I have mentioned aren't as explicit as each other about exactly which theory they are attributing to Lewis, but the rough intuitive idea is that the meaning of a predicate is the most natural property that is more-or-less consistent with the usage of the predicate. Call this kind of interpretation the 'orthodox' interpretation of Lewis.[2] Recently Wolfgang Schwarz (2009, 209ff) has argued that the orthodox interpretation is a misinterpretation, and actually naturalness plays a much smaller role in Lewis's theory of meaning than is standardly assumed.[3] Simplifying a lot, one key strand in Schwarz's interpretation is that naturalness plays no role in the theory of meaning in (Lewis 1969; Lewis 1975), since Lewis hadn't formulated the concept yet, and Lewis didn't abandon that theory of meaning, since he never announced he was abandoning it, so naturalness doesn't play anything like the role orthodoxy assigns to it.

In this article I attempt to steer a middle ground between these two positions. I'm going to defend the following parcel of theses. These are all exegetical claims, but I'm also interested in defending most of the theses that I ultimately attribute to Lewis, so getting clear on just what Lewis meant is of more than historical interest.

1. Naturalness matters to Lewis's (post-1983) theory of sentence meaning only insofar as it matters to his theory of rationality, and the theory of rationality matters to the (pre- and post-1983) theory of meaning.
2. Naturalness might play a slightly more direct role in Lewis's theory of word meaning, but it isn't nearly as significant as the orthodox view suggests.
3. When we work through Lewis's theory of word and sentence meaning, we see that the orthodox interpretation assigns to Lewis a theory that isn't his theory of meaning, but is by his lights a useful heuristic.
4. An even better heuristic than 'meaning = use plus naturalness' would be 'meaning = predication plus naturalness', but even this would be a fallible heuristic, not a theory.
5. When correctly interpreted, Lewis's theory is invulnerable to the challenges put forward by Williams (2007).

I'm going to start by saying a little about the many roles naturalness plays in Lewis's philosophy, and about his big picture views on thought and meaning. Then I'll offer a number of arguments against the orthodox interpretation of Lewis's theory of sentence meaning. After that, I'll turn to Lewis's theory of word meaning, where it is harder to be quite clear about just what the theory is, and how much it might have changed once natural properties were added to the metaphysics.

1. How Naturalness Enters The Theory of Meaning

Most of the core elements of David Lewis's philosophy were present, at least in outline, from his earliest work. The big exception is the theory of natural properties introduced in (Lewis 1983). As he

says in that paper, he had previously believed that "set theory applied to possibilia is all the theory of properties that anyone could ever need" (Lewis 1983, 377n). Once he introduces this new concept of naturalness, Lewis puts it to all sorts of work throughout his philosophy. I'm rather sceptical that there is any one feature of properties that can do all the varied jobs Lewis wants naturalness to do, but the grounds for, and consequences of, this scepticism are a little orthogonal to the main theme of this paper, so I've set it aside.

As the orthodox interpretation stresses, Lewis has naturalness do some work in this theory of content. That he does think there's a connection between naturalness and content is undeniable from the most casual reading of his post-1983 work. But just how they are connected is less obvious. To spell out these connections, let's start with three Lewisian themes.

- Facts about linguistic meaning are to be explained in terms of facts about minds. In particular, to speak a language $\mathscr{L}$ is to have a convention of being truthful and trusting in $\mathscr{L}$ (Lewis 1969; Lewis 1975). And to have such a convention is a matter of having certain beliefs and desires. So mental content is considerably prior to linguistic content in a Lewisian theory. Moreover, Lewis's theory of linguistic content is, in the first instance, a theory of *sentence* meaning, not a theory of *word* meaning.[4]

- The principle of charity plays a central role in Lewis's theory of mental content (Lewis 1974; Lewis 1994). To a first approximation, a creature believes that *p* iff the best interpretation of the creature's behavioural dispositions includes the attribution of the belief that *p* to the creature. And, ceteris paribus, it is better to interpret a creature so that it is more rather than less rational. It will be pretty important for what follows that Lewis adopts a principle of charity that highlights *rationality*, not *truth*. It is also important to Lewis that we don't just interpret the individual creature, but creatures of a kind (Lewis 1980). I'm not going to focus on the social externalist features of Lewis's theory of mental states, but I think they assist the broader story I want to tell.

- Lewis's theory of mental content has it that mental contents are (what most of us would call) properties, not (what most of us would call) propositions (Lewis 1979). So a theory of natural properties can easily play a role in the theory of mental content, since mental contents are properties. If you think mental contents are propositions, the connection between naturalness and mental content will be more indirect. Just how indirect it is will depend on what your theory of propositions is. But if mental contents are Lewisian propositions, the connection may be very indirect indeed. After all, propositions that we might pick out with sentences containing words that denote very unnatural properties, such as *All emeroses are gred*, might be intuitively very natural.

Now let's see why we might end up with naturalness in the theory of meaning. An agent has certain dispositions. For instance, after seeing a bunch of green emeralds, and no non-green emeralds, in a large and diverse range of environments, she has a disposition to say "All emeralds are green". In virtue of what is she speaking a language in which "green" means green, and not grue? (Note that when *I* use "grue", I mean a property that only differs from greenness among objects which it is easy to tell that neither our agent, nor any of her interlocutors, could possibly be acquainted with at the time she makes the utterance in question.)

Let's say that $\mathscr{L}_1$ is English, i.e., a language in which "green" means green, and $\mathscr{L}_2$ a language which is similar to $\mathscr{L}_1$ except that "green" means grue. Our question is, what makes it the case that the agent is speaking $\mathscr{L}_1$ and not $\mathscr{L}_2$? That is, what makes it the case that the agent has adopted the convention of being truthful

and trusting in $\mathscr{L}_1$, and not the convention or being truthful and trusting in $\mathscr{L}_2$?

We assumed that the agent has seen a lot of emeralds which are both green and grue. To a first approximation, it is more charitable to attribute to the agent the belief that all emeralds are green than the belief that all emeralds are grue because greenness is more natural than gruesomeness. As Lewis says, "The principles of charity will impute a bias towards believing things are green rather than grue" (1983, 375). And for Lewis, charity requires imputing more reasonable interpretations. But why is it more charitable to attribute beliefs about greenness to beliefs about grueness? I think it is because we need more evidence to rationally form a belief that some class of things are all grue than we need to form a belief that everything in that class is green. And that's because, ceteris paribus, we need more evidence to rationally form a belief that all *F*s are *G*s than that all *F*s are *H*s when *G* is less natural than *H*. The agent has, we might assume, sufficient evidence to rationally believe that all emeralds are green, but not sufficient evidence to believe that all emeralds are grue.

So the first two Lewisian themes notes above, the reduction of linguistic meaning to mental content, and the centrality of a rationality-based principle of charity, push us towards thinking that naturalness is closely connected to mental content and hence to linguistic meaning. And it has pushed us towards thinking that if naturalness is connected to meaning, it is via this connection I've posited between naturalness and rational belief. Note that Lewis doesn't ever endorse anything like that general a connection, but I suspect he had something like this in mind when he wrote the sentence I quoted in the previous paragraph. We'll come back to this interpretative question at some length below.

But the argument I offered was a bit quick, because I ignored the third Lewisian theme: beliefs are relations to properties, not propositions. On Lewis's theory, to believe that all emeralds are green is to self-ascribe the property of being in a world where all emeralds are green. So if a certain body of evidence makes it possible for the agent to rationally believe that all emeralds are green, but not for her to believe that all emeralds are grue, and that's because rationality is constitutively connected to naturalness, then that must be because the first of the following properties is more natural than the second:

- Being in a world where all emeralds are green
- Being in a world where all emeralds are grue

That could still be true, though it is notable how far removed we are from the intuitions that motivate the distinctions between more and less natural properties. It's not like there is some sense, intuitively, in which things that have the first property form a more unified class than things that have the second property. The striking difference between these two properties lies not in metaphysics, but in epistemology. It takes much less evidence to self-ascribe the former than the latter. Perhaps there is an explanation of that in metaphysical terms. But even if so, the explanation will be long and complicated, and the epistemological point simple and direct.

So it's plausible that naturalness is connected to mental content, at least as long as naturalness is connected to rational belief. And since mental content is connected to linguistic content, we're now in the vicinity of the orthodox interpretation. But I don't think the orthodox interpretation can be right. I'll give several reasons for this, starting with the textual evidence for and against it.

2. Textual Evidence about Sentence Meaning

There is some *prima facie* textual evidence for the orthodox interpretation. But looking more careful at the context of these texts not

just undermines the support the text gives to the orthodox interpretation, but actually tells against it. (This part of the paper is indebted even more than the rest to Wolfgang Schwarz's work, and could be easily skipped by those familiar with that work.)

I'll focus on the last seven pages of "New Work for a Theory of Universals". This is the part of "New Work" that uses the notion of naturalness, as introduced in the paper, to respond to Putnam's model-theoretic arguments for massive indeterminacy of meaning. Lewis actually responds to Putnam twice over. First, he responds to Putnam directly, by showing how adding naturalness to a use-based theory of sentence meaning avoids the 'just more theory' objection that's central to Putnam's argument. And when Lewis describes this direct response, he says things that sound a lot like the orthodox interpretation.

> I would instead propose that the saving constraint concerns the referent - not the referrer, and not the causal channels between the two. It takes two to make a reference, and we will not find the constraint if we look for it always on the wrong side of the relationship. Reference consists in part of what we do in language or thought when we refer, but in part it consists in eligibility of the referent. And this eligibility to be referred to is a matter of natural properties. (Lewis 1983, 371)

But after this direct response is finished, Lewis notes that he has conceded quite a lot to Putnam in making the response.

> You might well protest that Putnam's problem is misconceived, wherefore no need has been demonstrated for resources to solve it. Where are the communicative intentions and the mutual expectations that seem to have so much to do with what we mean? In fact, where is thought? … I think the point is well taken, but I think it doesn't matter. If the problem of intentionality is rightly posed there will still be a threat of radical indeterminacy, there will still be a need for saving constraints, there will still be a remedy analogous to Merrill's suggested answer to Putnam, and there will still be a need for natural properties. (Lewis 1983, 373)

I noted earlier that Schwarz makes much of a similar passage in "Putnam's Paradox", and I think he is right to do so. Here's a crucial quote from that paper.

> I shall acquiesce in Putnam's linguistic turn: I shall discuss the semantic interpretation of language rather than the assignment of content to attitudes, thus ignoring the possibility that the latter settles the former. It would be better, I think, to start with the attitudes and go on to language. But I think that would relocate, rather than avoid, the problem; wherefore I may as well discuss it on Putnam's own terms. (Lewis 1984, 222)

That passage ends with a footnote where he says the final section of "New Work" contains a version of how the 'relocated' problem would be solved. So let's turn back to that. The following long portmanteau quote from pages 373 to 375 captures, I think, the heart of my interpretation.

> The problem of assigning content to functionally characterised states is to be solved by means of constraining principles. Foremost among these are principles of fit. … A state typically caused by round things before the eyes is a good candidate for interpretation as the visual experience of confronting something round; and its typical impact on the states interpreted as systems of belief ought to be interpreted as the exogenous addition of a belief that one is confronting something round, with whatever adjustment that addition calls for. … Call two worlds equivalent iff they are alike in respect of the subject's evidence and behaviour, and note that any decent world is equivalent inter alia to horrendously counterinductive worlds and to worlds where everything unobserved by the subject is horrendously nasty. … We can interchange equivalent worlds ad lib and preserve fit. So, given any fitting and reasonable interpretation, we can transform it into an equally fitting perverse interpretation by swapping equivalent worlds

> around … If we rely on principles of fit to do the whole job, we can expect radical indeterminacy of interpretation. We need further constraints, of the sort called principles of (sophisticated) charity, or of 'humanity'. [A footnote here refers to "Radical Interpretation".] Such principles call for interpretations according to which the subject has attitudes that we would deem reasonable for one who has lived the life that he has lived. (Unlike principles of crude charity, they call for imputations of error if he has lived under deceptive conditions.) These principles select among conflicting interpretations that equally well conform to the principles of fit. They impose *apriori* – albeit defeasible - presumptions about what sorts of things are apt to be believed and desired … **It is here that we need natural properties.** The principles of charity will impute a bias toward believing that things are green rather than grue … In short, they will impute eligible content … They will impute other things as well, but it is the imputed eligibility that matters to us at present. (Lewis 1983, 373-5, my emphasis)

I think that does a reasonably clear job of supporting the interpretation I set out in the introduction over the orthodox interpretation. Naturalness matters to linguistic meaning all right. But the chain of influence is very long and indirect. Naturalness constrains what is reasonable, reasonableness constrains charitable interpretations, charitable interpretations constrain mental content, and mental content constrains linguistic content. Without naturalness at the first step, we get excessive indeterminacy of content. With it, the Putnamian problems are solved. But there's no reason here to think naturalness has any more direct role to play at any level in the theory of linguistic content.

In short, Lewis changed what he thought about rationality when he adopted the theory of natural properties. Since rationality was a part of his theory of mental content, and mental content determines linguistic content, this change had downstream consequences for what he said about linguistic content. But there wasn't any other way his theory of linguistic content changed, nor, contra orthodoxy, any direct link between naturalness and predicate meaning.

Moreover, when we look at the closest thing to a worked example in (Lewis 1983), we don't get any motivation for the orthodox interpretation. Here's the example he uses, which concerns mental content. Let f be any one-one mapping from worlds to worlds such that the agent has the same evidence and behaviour in w and $f(w)$. Extend f to a mapping from sets of worlds to sets of worlds in the following (standard) way: $f(S)=\{f(w):w\in S\}$. Then the agent's behaviour will be rationalised by her evidence just as much if she has credence function C and value function V, as if she has credence function C' and value function V', where $C'(f(S))=C(S)$, and $V'(f(S))=V(S)$. To relate this back to the familiar Goodmanian puzzle, let f map any world where all emeralds are green to nearest world where all emeralds are grue, and vice versa, and map any other world to itself. Then the above argument will say that the agent's behaviour is rationalised by her evidence just as much as if her credences are C as if they are C'. That is, her behaviour is rationalised by her evidence just as much if she gives very high credence to all emeralds being green as to all emeralds being grue. So understanding charity merely as rationalizing behaviour leaves us without a way to say that the agent believes unobserved emeralds are green and not grue.

Lewis's solution is to say that charity requires more than that. In particular, it requires that we assign natural rather than unnatural beliefs to agents where that is possible. I've argued above that this makes perfect sense if we connect naturalness with rationality. The crucial thing to note here is that this all happens a long time before we can set out the way that a sentence is used, since the way a sentence is used on Lewis's theory of linguistic content includes the beliefs that are formed on hearing it. So the discussion in "New Work" suggests that naturalness matters for content, but

not in a way that can be easily factorised out. And that's exactly what I think is the best way to understand Lewis's theory.

3. Textual Evidence on Naturalness and Rationality

A major part of my argument above was that naturalness affected Lewis's theory of rationality. In particular, once he had naturalness to work with, he seemed to think that it was more rational to project natural rather than unnatural properties. The textual evidence for this is, I'll admit, fragmentary. But it is fairly widespread. Let's start with a quote we've already seen.

> The principles of charity will impute a bias toward believing that things are green rather than grue (Lewis 1983, 375)

As noted above, I assume this isn't a special feature of green and grue, but rather that there is a general principle in favour of projecting natural properties. But it would be good to have more evidence for that.

Lewis returns to the example of the believer in grue emeralds a few times. Here is one version of the story in *Plurality*.

> We think that some sorts of belief and desire … would be unreasonable in a strong sense … utterly unintelligible and nonsensical. Think of the man who, for no special reason, expects unexamined emeralds to be grue. … What makes the perversely twisted assignment of content incorrect, however well it fits the subject's behaviour, is exactly that it assigns ineligible, unreasonable content when a more eligible assignment would have fit behaviour equally well. (Lewis 1986, 38-9)

And a little later, when replying to Kaplan's paradox, he says,

> Given a fitting assignment, we can scramble it into an equally fitting but perverse alternative assignment. Therefore a theory of content needs a second part: as well as principles of fit, we need 'principles of humanity', which create a presumption in favour of some sorts of content and against others. (Lewis 1986, 107)

He returns to this point again in "Reduction of Mind".

> [Folk psychology] sets presumptive limits on what our contents of belief and desire can be. Self-ascribed properties may be 'far from fundamental', I said – but not *too* far. Especially gruesome gerrymanders are *prima facie* ineligible to be contents of belief and desire. In short, folk psychology says that we make sense. It credits us with a modicum of rationality in our acting, believing and desiring. (Lewis 1994, 320 in reprint)

The running thread through these last three quotes is that our theory of mental content rules out gruesome assignments, and it does this because assigning rationality is constitutive of correctly interpreting. This can only work if naturalness is connected to rationality. I've attributed a stronger claim to Lewis, that not only is naturalness connected to rationality, but that the connection goes through projection.[5]

One piece of evidence for that is that Lewis says, in "Meaning Without Use" that Kripkenstein's challenge was "formerly Goodman's challenge" (Lewis 1992, 109). He goes on to say that the solution to this challenge (or should that be 'these challenges') involves "carrying more baggage of primitive distinctions or ontological commitments than some of us might have hoped" (Lewis 1992, 110). A footnote on that sentence cites "New Work", in case it isn't obvious that the baggage here is the distinction between natural and unnatural properties. So somehow, Lewis thinks that natural properties help solve Goodman's puzzle. I think that the simplest such solution is the right one to attribute to Lewis; natural properties are *prima facie* more eligible to be projected.

A referee noted that this passage is a little odd; it appears to simply conflate a meta-semantical paradox with an epistemologi-

cal paradox. But I think that just shows how much, for Lewis, meta-semantical questions are epistemological questions. Words get their meanings in virtue of our conventions. Our conventions consist of our beliefs and desires. And facts about rationality are, in part, constitutive of what we believe and desire.

Finally, consider the way in which the papers on natural properties are introduced in *Papers in Metaphysics and Epistemology*. Lewis says that "I had been persuaded by Goodman and others that all properties were equal: it was hopeless to try to distinguish 'natural' properties from gruesomely gerrymandered, disjunctive properties." (Lewis 1999, 1-2) A footnote refers to *Fact, Fiction and Forecast*. Of course, the point of "New Work" is that Lewis abandons this, explicitly Goodmanian, view. Now that he had learned property egalitarianism from Goodman of course doesn't show that once he became a property inegalitarian, he applied this to Goodman's own paradox. But it does seem striking that the only citation of an egalitarian view is of *Fact, Fiction and Forecast*. I take that to be some, inconclusive, evidence that Lewis did indeed think natural properties were related to Goodman's paradox.

Ultimately, it seems the textual evidence is this. There are many different occasions where Lewis makes clear there is a connection between naturalness and rationality, and in particular, between naturalness and the kind of rationality that is relevant to content assignment. There are hints that this connection goes via naturalness playing a role in solving Goodman's paradox. Notably, there is no other obvious way in which naturalness could connect to rationality. At least, I can neither think of another connection, nor see any evidence for another connection in the Lewis corpus. So I conclude, a little tentatively, that Lewis thought natural properties had a role to play in solving Goodman's paradox.

4. Word Meaning and Naturalness

In "Languages and Language", Lewis doesn't say that human linguistic practices merely determine truth conditions for the spoken sentences. That is, our linguistic practices don't merely determine which **language**, in Lewis's sense, we speak. They also determine, to some extent, a **grammar**, which specifies the truth conditional contribution of the various parts of the sentence. The grammar determines the "fine structure of meaning" (Lewis 1975, 177) of a sentence or phrase.

In comments on an earlier draft of this paper, an anonymous referee stressed that naturalness could enter directly into a theory of meaning once we stopped focussing on sentence meaning, and started looking on word meaning. I don't mean to say the referee was endorsing any particular role for naturalness in the theory of word meaning. But the point that we need to say more about the Lewisian approach to word meaning before we conclude that naturalness is only indirectly related to meaning is right. And I'm grateful for the encouragement to discuss it further.

Lewis has a short discussion of grammars in "Languages and Language", and another in "Radical Interpretation". It's worth looking at both of these in turn. I'll take "Languages and Language" first, since even though it has a slightly later publication date, in the respects we're discussing here it closely resembles the theory in *Convention*.

On pages 177-8 of that paper, Lewis notes three ways in which there may be indeterminacy in the grammar.

1. A subject's behavioural dispositions and anatomy might underdetermine their beliefs and desires.

2. The beliefs and desires might underdetermine the truth conditions of their language.

3. The truth conditions of the language might underdetermine the meanings of the individual words.

While Lewis does not think the second is actually a source of indeterminacy, he does think that the third is.

> My present discussion has been directed at the middle step … I have said … that the beliefs and desires of the subject and his fellows are such as to comprise a fully determinate convention of truthfulness and trust in some definite language. … I am inclined to share in Quine's doubts about the determinacy of the third step. (Lewis 1975, 178)

Lewis gives reasons for this inclination a few paragraphs earlier. He says that while we can say what it is for a community to speak one language rather than another, we can't say what it is for a community to speak one grammar rather than another. He says that we don't have any objective measures for evaluating grammars. And he says Quine's examples of indeterminacy of reference show that languages can have multiple good grammars, even if these disagree radically about the meaning of some constituents.

Notably, Lewis doesn't take to show that there is anything wrong with the notion of word meaning. He says it would be "absurd" (177) to conclude that. His conclusion here is more one of modesty rather than philosophical scepticism. We don't know how to extend the theory of sentence meaning he offers to a theory of word meaning, so we should do what we can without talking about word meaning.

The approach in "Radical Interpretation" has a bit more of a hint for how to restore semantic determinacy. The subject matter of that paper is how to solve for the mental and linguistic contents of a speaker, called Karl, given the physical facts about them. Lewis uses **M** for "a specification, in our language, of the meanings of expressions of Karl's language." (Lewis 1974, 333) He lists a number of constraints on a solution, including early versions of his principles of constitutive rationality. But the most notable constraint, from our perspective, is this:

> *The Principle of Generativity* constrains **M**: **M** should assign truth conditions to the sentences of Karl's language in a way that is at least finitely specifiable, and preferably also reasonably uniform and simple. (Lewis 1974, 339)

There's something very odd about this. Lewis, in 1974, didn't have a theory of what made an assignment simple. He needed his theory of natural properties to do that. Or, at least, once he had the theory of natural properties, it did all the work he ever wanted out of an account of simplicity.

Be that as it may, it does suggest that Lewis did think that simplicity of assignments could be used as a way of cutting down the third kind of semantic indeterminacy discussed in "Languages and Language". He doesn't think it would generate a fully determinate interpretation of Karl's language.

> It seems hopeless to deny, in the face of such examples as those in [Quine's "Ontological Relativity", pp. 30-39], that the truth conditions of full sentences in **M** do not sutfice to determine the rest of **M**: the parsings and the meanings of the constituents of sentences. At least, that is so unless there is something more than our Principle of Generativity to constrain this auxiliary syntactic and semantic apparatus. (Lewis 1974, 342-3)

It's notable that some of the examples Quine gives in "Ontological Relativity" are not cases where the alternative meanings are by any measure equally natural. This positive allusion to Quine's examples suggests a link to this comment in "Languages and Language"

> We should regard with suspicion any method that purports to settle objectively whether, in some tribe, "gavagai" is true of temporally continuant rabbits or time-slices thereof. You can give their language a good grammar of either kind—and that's that. (Lewis 1975, 177)

Note that he doesn't say 'equally' good. And note also how this contrasts with the attitude he takes towards the prospects of indeterminacy in sentence meaning. I earlier quoted him saying that part of the point of "Languages and Language" was to show how the second type of indeterminacy didn't arise. He ends "Radical Interpretation" with this 'credo'.

> Could indeterminacy of beliefs, desires, and truth conditions also arise because two different solutions both fit all the constraints perfectly? Here is the place to hold the line. This sort of indeterminacy has not been shown by convincing examples, and neither could it be shown–to me–by proof. *Credo*: if ever you prove to me that all the constraints we have yet found could permit two perfect solutions, differing otherwise than in the auxiliary apparatus of **M**, then you will have proved that we have not yet found all the constraints. (Lewis 1974, 343)

So that's where things stood before 1983. Lewis thought he had a theory that eliminated, or at least minimised, indeterminacy at the level of truth conditions. But he didn't think his theory eliminated indeterminacy, even quite radical indeterminacy, in word meanings. And he didn't seem bothered by this aspect of the theory; indeed, he thought Quine's arguments showed that we shouldn't eliminate this kind of indeterminacy.

This attitude towards Quinean arguments for indeterminacy is obviously a striking contrast to the forcefulness, and rapidity, with which he responded to Putnam's arguments for indeterminacy. That shouldn't be too surprising once we attend to Lewis's threefold distinction between kinds of indeterminacy. Quine was arguing that indeterminacy of the third kind was rampant. Putnam was arguing that indeterminacy of the second kind was rampant. And, as Lewis announced in "Radical Interpretation", he wasn't going to believe any such argument.

Still, we might wonder whether the resources he brought to bear in responding to Putnam also help respond to Quine. Or, perhaps more importantly for exegetical reasons, we might wonder whether Lewis thought they were useful in responding to Quine. The evidence from "New Work" seems to suggest a negative answer to the latter question. Lewis never says that one of the things you can do with the distinction between natural and unnatural properties is respond to arguments for Quinean indeterminacy. And that's despite the fact that "New Work" has a very survey-like feel; the bulk of the paper is a long list of philosophical work that a theory of universals can do.

In "Putnam's Paradox" there is a brief footnote on Quine's arguments for indeterminacy. It reads

> It is not clear how much indeterminacy might be expected to remain. For instance, what of Quine's famous example? His rabbit-stages, undetached rabbit parts, and rabbit-fusion seem only a little, if any, less eligible than rabbits themselves. (Lewis 1984, 228n)

As I've stressed repeatedly, following Schwarz, taking the disclaimers at the start of "Putnam's Paradox" seriously means that we have to be careful in interpreting what Lewis says about how words acquire determinate meaning in that paper. But even before we adjust for the disclaimers, this is hardly a ringing rejection of Quine's indeterminacy arguments. The contrast to Lewis's attitude towards Putnam's arguments is striking. Since it is the very same contrast that we saw in both "Languages and Language" and "Radical Interpretation", I think it is fair to assume that he continued to think Quine's arguments were considerably stronger than Putnam's.

But there is, perhaps, a change of view in "Meaning Without Use". Here's the problem Lewis addresses at the end of that paper. Let $\mathscr{L}_1$ once again be English as we currently understand it, and let $\mathscr{L}_3$ be just like English, except that it doesn't assign any truth conditions to sentences over a thousand words long.[6] Do our actual linguistic practices manifest a convention of trust in $\mathscr{L}_1$, or trust in $\mathscr{L}_3$? Lewis argues that it is more like a convention of trust in $\mathscr{L}_3$. If someone utters a very long sentence, we expect some kind of performance error, at best. We don't, in general, believe what they say. So the theory of "Languages and Language" seems to predict that these long sentences have no truth conditions. But that's wrong, so the theory must be corrected.

Lewis's correction appeals, it seems, to natural properties in fixing a grammar. He says that linguistic practice determines truth conditions for a fragment of the language that is widely used. Those truth conditions determine meanings of words. This determination requires natural properties; without them the Quinean problems multiply indefinitely. We then use those word meanings to determine the meaning of unused sentences. A long footnote suggests that the procedure might not be restricted to unused sentences. As long as there is a large enough fragment in which there are conventions of truthfulness and trust, we can extrapolate from that to other parts of the language that are used.

This is a marked deviation from anything Lewis had said until then. From the earliest writings, he had stressed a step-by-step approach to content determination. Behavioural dispositions plus physical and biological constraints determine mental content; mental content determines sentence meaning; and sentence meaning determines word meaning. In "Meaning Without Use", it seemed the last two steps were being somewhat merged.

But we shouldn't overstate how much the third step was allowed to encroach on the second. Lewis does think we need to rule out 'bent' grammars, which don't assign any truth conditions to sentences over a thousand words long, or which give sentences different meanings to what we'd expect if the word 'cabbage' appears forty times. But he doesn't think we need to rule out any 'straight' grammar, which includes "any grammar that any linguist would actually propose." (Lewis 1992, 109)

So Lewis's focus here is to rule out unnatural *compositional rules*, not unnatural assignments of content to individual words. The reference to linguists here might be useful. Linguists tend to spend much more time on compositional rules than they do on the contents on individual predicates. Notably, Quine didn't argue for indeterminacy by positing indeterminacy in the compositional rules of the language; his non-standard interpretations all share a standard syntax. If we posit that Lewis thought that there was little syntactic indeterminacy in the language, like there is little indeterminacy at the level of truth conditions of sentences, we can tell a story that doesn't involve too many unsignalled changes of view. Here's how I would tell that story in some more detail.

Lewis's early view, expressed clearly in "Radical Interpretation" and "Languages and Language", and not retracted before, I think, 1992, has the following parts:

1. Conventions of truthfulness and trust determine (very sharply) truth conditions for sentences in a speaker's language.

2. Any reasonably good grammar, i.e., assignment of word meanings and compositional rules, that is consistent with the truth conditions is not determinately wrong. There is potentially substantial indeterminacy in the meaning of any given word, because there are many reasonably good grammars consistent with the truth conditions. Any grammar that has excessively complex compositional rules is not reasonably good.

After 1983, 'complexity' was understood in terms of naturalness, but otherwise the story doesn't change a lot.

The later view, which goes by somewhat more quickly in "Meaning Without Use", has the following parts:

1. Conventions of truthfulness and trust in (the bulk of) the used fragment of the language determine truth conditions for that fragment.

2. Naturalness considerations determine the compositional rules for the language by extrapolation from that grammar.

3. Word meanings are determined, so far as they are determinate, by the truth conditions for sentences, plus the compositional rules.

4. Truth conditions for sentences outside the used fragment are determined by the word meanings and the compositional rules.

Neither of these views look much like the orthodox view. Remember that the orthodox view has it that considerations of naturalness can be used to resolve debates in metaphysics. That's certainly the use that Sider (2001a) makes of the orthodox view. But on the early view, simplicity considerations only come in after the truth conditions for every sentence have been determined, and hence so that all metaphysical debates are settled. And on the later view, simplicity considerations primarily are used to settle truth conditions for unused, or at least unusual, sentences.

Now if you thought the salient fragment in point 1 of the later view was small, and if you thought naturalness had a major role to play in step 3 of the later view, you would get back to something like the orthodox view. But I don't see the textual evidence for either of those positions. Lewis says that "the used fragment is large and varied." (Lewis 1992, 110) It doesn't look like he is positing wholesale changes to his view on the determination of truth conditions. He is positing some changes; the last two pages of the paper are clearly marked as deviations from his earlier position. But both the examples he uses and the rhetoric around them suggests that the bulk of the changes happen at point 2. Naturalness considerations constrain the syntax of a language much more tightly than they constrain the assignment of meaning to a given word. In sum, at no point in the evolution of his views did Lewis seem to endorse the orthodox interpretation, even as a theory of word meaning.

5. An Argument for the Orthodox Interpretation

So far I've argued that there is no solid textual support for the orthodox interpretation. My rival interpretation relied on there being a connection between naturalness and induction, and as we've just seen, there is some textual evidence for this. But perhaps there is a more indirect way to motivate the orthodox interpretation of Lewis. The orthodox interpretation attributes to Lewis a theory that is quite attractive as a theory of semantic determinacy and indeterminacy. Call that theory the **U&N Theory**, short for the **U**se plus **N**aturalness theory of meaning. Since Lewis was clearly looking for such a theory when he discussed naturalness in the context of his theory of content, it is reasonably charitable to attribute the **U&N Theory** to him, as the orthodox interpretation does.

My response to this will be in three parts. First, I'll argue in this section that my rival interpretation attributes to Lewis a theory of semantic determinacy and indeterminacy that does just as well at capturing the facts Lewis wanted a theory to capture, so there's no charity based reason to attribute the **U&N Theory** to him (And, as we saw in the previous section, there's no direct textual reason to attribute it to him either.) Second, the **U&N Theory** is subject to the criticisms in (Williams 2007), while the theory I attribute to Lewis is not. Third, the **U** part of the **U&N Theory** is hopelessly vague; it isn't clear how to say what 'use' is on a

Lewisian theory that makes it suitable to add to naturalness to deliver meanings. Either use is so thick that naturalness is unneeded, or it is so thin that naturalness won't be sufficient to set meaning. So actually it isn't particularly charitable to attribute this theory to him.

Still, let's start with the attractions of the **U&N Theory**. On the one hand, agents are inclined to say "All emeralds are green" both in situations where they've seen a lot of green emeralds (and no non-green ones) and in situations where they've seen a lot of grue emeralds (and no non-grue ones). That's because, of course, those are exactly the same situations. So at first glance, it doesn't look like the way in which "green" is used will determine whether it means green or grue. On the other hand, once we add a requirement that terms have a relatively natural meaning, we do get this to fall out as a result. Moreover we can even see how this falls out of a recognisably Lewisian approach to meaning.

Consider again our agent who says "All emeralds are green" after seeing a lot of emeralds that are both green and grue. And remember that for her to speak a language, she must typically conform to conventions of truthfulness and trust in that language. Now if the agent was speaking $\mathscr{L}_2$, she would have to think that she's doing an OK job of being truthful in $\mathscr{L}_2$ by saying "All emeralds are green". But that would be crazy. Why should she think that all emeralds are grue given her evidence base? To attribute to her that belief would be to gratuitously attribute irrational beliefs to her. And on Lewis's picture, gratuitous attributions of irrationality are false. So the agent doesn't have that belief. So she's not speaking $\mathscr{L}_2$.

Things are even clearer from the perspective of hearers. A hearer of "All emeralds are green" would be completely crazy to come to believe that all emeralds are grue. The hearer knows, after all, that the speaker has no acquaintance with the emeralds that would have to be blue for all emeralds to be grue. So the hearer knows that this utterance could not be sufficient evidence to believe that all emeralds are grue. Yet if she speaks $\mathscr{L}_2$, she is disposed to believe that all emeralds are grue on hearing "All emeralds are green". She isn't irrational, or at least we shouldn't assign irrationality to her so quickly, so she doesn't speak $\mathscr{L}_2$.

So it looks like in this one case at least, we have a case where use plus naturalness gives us the right theory. Agents are disposed to use "green" to describe emeralds that are green/grue. But the fact that greenness is more natural than gruesomeness makes it more appropriate to attribute to them a convention according to which "All emeralds are green" means that all emeralds are green and not that all emeralds are grue.

But more carefully, what we should say is that the **U&N Theory** gives us the right result in this case. It doesn't follow that it will work in all cases, or anything like it. And it doesn't follow that it works for the right reasons. As we'll see, neither of those claims are true. In fact, just re-reading the last three paragraphs should undermine the second claim. Because we just saw a derivation that the agents are not speaking $\mathscr{L}_2$ that didn't even appeal to the **U&N Theory**. Rather, that derivation simply used the theory of meaning in *Convention* and the theory of mental content in "Radical Interpretation". It's true that the latter theory assigns a special role to rationality, and the theory of rationality we used has, among other things, a role for natural properties, but that is very different to the idea that naturalness feeds directly into the theory of meaning in the way the orthodox interpretation says. As I said at the start, I think the best interpretation of Lewis is that he changed his theory of *rationality* in 1983, but that's the only change to his theory of *meaning*.

Put another way, these reflections on "green" and "grue" are consistent with the view that the **U&N Theory** is a false *theory*, but a useful *heuristic*. It's a useful heuristic because it agrees with the true Lewisian theory in core cases, and is much easier to apply.

That's exactly what I think the **U&N Theory** is, both as a matter of fact, and as a matter of Lewis interpretation.

6. Indeterminacy and Radically Deviant Interpretations

If the **U&N Theory** is a heuristic not a theory, we should expect that it will break down in extreme cases. That's exactly what we see in the cases discussed in (Williams 2007). Those cases highlight the fact that a Lewisian theorist needs to be careful that we don't end up concluding that normal people, such as the agent in our example who says "All emeralds are green", speak $\mathscr{L}_4$. $\mathscr{L}_4$ is a language in which all sentences express claims about a particular mathematical model (essentially a Henkin model of the sentences the agent accepts), and it is set up in such a way that ordinary English sentences come out true, and about very natural parts of the model. On the **U&N Theory**, it could easily turn out that ordinary speakers are speaking $\mathscr{L}_4$, since the assigned meanings are so natural. We can see this isn't a consequence of *Lewis's* theory by working through the case from first principles. I have two arguments here, the first of them relying on some slightly contentious claims about the epistemology of mathematics, the second less contentious.

Assume, for reductio, that ordinary speakers are speaking $\mathscr{L}_4$. So, for instance, when O'Leary says "The beer is in the fridge", what he says is that a certain complicated mathematical model has a certain property. (And indeed it has that property.) Now this won't be a particularly rational thing for O'Leary to say unless he knows more mathematics than ordinary folks like him ordinarily do. So if O'Leary has adopted a convention of truthfulness and trust in $\mathscr{L}_4$, then uttering "The beer is in the fridge" would be irrational, even if he is standing in front of the open fridge, looking at the beer. That's a gratuitous assignment of irrationality, and gratuitous assignments of irrationality are false, so O'Leary doesn't speak $\mathscr{L}_4$.

Perhaps that is too quick. After all, the mathematical claim that $\mathscr{L}_4$ associates with "The beer is in the fridge" is a necessary truth. And Lewis's theory of content is intentional, not hyperintentional. So O'Leary does know it is true. (And when he is standing in front of the fridge, there's even a sense that he knows that "The beer is in the fridge" expresses a truth, if $\mathscr{L}_4$ is really his language.) I think that's probably not the right sense of "rational", and I'm not altogether sure how much hostility to hyperintensionalism we should attribute to Lewis. But so as to avoid these questions, it's easier to consider a different argument that focusses attention on O'Leary's audience.

When O'Leary says "The beer is in the fridge", Daniels hears him, and then walks to the fridge. Why does Daniels make such a walk? Well, he wants beer, and believes it is in the fridge. That looks like a nice rational explanation. But why does he believe the beer is in the fridge? I say it's because he's (rationally) adopted a convention of truthfulness and trust in $\mathscr{L}_1$, and so he rationally comes to believe the beer is in the fridge when O'Leary says "The beer is in the fridge". On the assumption that O'Leary and Daniels speak $\mathscr{L}_4$, none of this story goes through. But we must have some rational explanation of why O'Leary's statement makes Daniels walk to the fridge. So O'Leary and Daniels must not be speaking $\mathscr{L}_4$.

Michael Morreau pointed out (when I presented this talk at CSMN) that the preceding argument may be too quick. Perhaps there is a way of rationalising Daniels's actions upon hearing O'Leary's words consistent with the idea that they both speak $\mathscr{L}_4$. Perhaps, for instance, Daniels's walking to the fridge constitutes saying something in a complicated sign language, and that thing is the rational reply to what O'Leary said. If this kind of response works, and I have no reason to think it won't, the solution is to

increase the costs to Daniels of performing such a reply. For instance, not too long ago I heard Mayor Bloomberg say "Lower Manhattan is being evacuated because of the impending hurricane", and I (and my family) packed up and evacuated from Lower Manhattan. Even if one could find an interpretation of our actions in evacuating that made them constitute the assertion of a sensible reply to Bloomberg's mathematical assertion in $\mathscr{L}_4$, it would be irrational to think I made such an assertion. Evacuating ahead of a storm with an infant is not fun - if it was that hard to make mathematical assertions, I wouldn't make them! And I certainly wouldn't make them in reply to someone who wouldn't even see my gestures. So I think at least some of the actions that are rationalised by testimony, interpreted as sentences of $\mathscr{L}_1$, are not rationalised by testimony, interpreted as $\mathscr{L}_4$. By the kind of appeal to the principle of charity we have used a lot already, that means that $\mathscr{L}_4$ is not the language most people speak.

The central point here is that when we are ruling out particularly deviant interpretations of some speakers, we have to make heavy use of the requirement that the interpretation of their shared language rationalises what they do. In part that means it must rationalise why they utter the strings that they do in fact utter. And when we're considering this, we should remember the role of naturalness in a theory of rationality. But it also means that it must rationalise why people respond to various strings with non-linguistic actions, such as walking to the fridge, or evacuating Lower Manhattan. Naturalness has less of a role to play here, but the Lewisian theory still gets the right answers provided we apply it carefully. Since the Lewisian theory gets the right answers, and the **U&N Theory** gets the wrong answers, it follows that the **U&N Theory** isn't Lewis's theory, and so orthodoxy is wrong.

7. What is the Use of a Predicate?

We concluded the last section with an argument that Lewis isn't vulnerable to the claim that his theory assigns complicated mathematical claims as the meanings of ordinary English sentences. That interpretation, we argued, is inconsistent with the way those sentences are used. In particular, it is inconsistent with the way that *hearers* use sentences to guide their actions.

So far so good, we might think. But notice how much has been packed into the notion of use to get us this far. In identifying the use O'Leary makes of "The beer is in the fridge", we have to say a lot about O'Leary's beliefs and desires. And in identifying the use Daniels makes of it, we *primarily* talk about the sentence's effects on Daniels's beliefs and desires. That is, just saying how the sentence is used requires saying a lot about mental states of speakers. And that will often require appealing to constitutive rationality; we say that Daniels's beliefs about the fridge changed because we need to rationalise his fridge-directed behaviour.

And this should all make us suspicious about the prospects for identifying meaning (in a Lewisian theory) with use plus naturalness. The argument above that naturalness mattered to meaning relied on the idea that naturalness matters because it affects which states are rational, and hence which states are actualised. A belief that all emeralds are grue is unnatural, so it is hard to hold. And since it is hard to hold, it is hard to think one is conforming to a convention of truthfulness in a language if one utters sentences that mean, in that language, that all emeralds are grue. That's why it is wrong, *ceteris paribus,* to interpret people as speaking about grueness.

But now consider what happened when we were talking about Daniels and O'Leary. Even to say how they were using the sentence "The beer is in the fridge", we had to say what they believed before and after the sentence was uttered. In other words, their mental states were constitutive of the way the sentence was used.

Now add in the extra premise, argued for above, that naturalness matters to Lewis's theory of linguistic content because, and only because, it matters to his theory of mental content. (And it only matters to mental content because it matters to the principle of charity that Lewis uses.) If mental states, and their changes, are part of how the sentences are used, it will be rather misleading to say that meaning is determined by use plus naturalness. A better thing to say is that meaning is determined by use, and that some key parts of use, i.e., mental states of speakers and hearers, are determined in part by naturalness.

So I'm sceptical of the **U&N Theory**. We can put the argument of the last few paragraphs as a dilemma. There are richer and thinner ways of identifying the use to which a sentence is put. A thin way might, for instance, just focus on the observable state of the part of the physical world in which the sentence is uttered. A rich way might include include, inter alia, the use that is made of the sentence in the management of belief and the generation of rational action. If we adopt the thin way of thinking about use, then adding naturalness won't be enough to say what makes it the case that O'Leary and Daniels are speaking $\mathscr{L}_1$ rather than $\mathscr{L}_4$. If we adopt the rich way of thinking about use, then the role that naturalness plays in the theory of meaning has been incorporated into the metaphysics of use. Neither way makes the **U&N Theory** true while assigning naturalness an independent role. This dilemma isn't just an argument that we shouldn't attribute the **U&N Theory** to Lewis; it is an argument against anyone adopting that theory.

8. From Theory to Applied Semantics

So far we've argued that Lewis's semantic theory did not look a lot like the orthodox interpretation. It's true that he thought the way a sentence was used was of primary importance in determining its meaning. And it's true that he thought naturalness mattered to meaning. But that wasn't because naturalness came in to resolve the indeterminacy left in a use-based theory of meaning. Rather, it was because naturalness was in a part of the theory of mental content, and specifying the mental states of speakers and hearers is part of specifying how the sentence is used.

But note that these considerations apply primarily to investigations at a very high level of generality, such as when we're trying to solve the problems described in "Radical Interpretation". They don't apply to investigations into applied semantics. Let's say we are trying to figure out what O'Leary and Daniels mean by "green". And assume that we are taking for granted that they are speaking a language which is, in most respects, like English. This is hardly unusual in ordinary work in applied semantics. If we are writing a paper on the semantics of colour terms, a paper like, say, "Naming the Colours", we don't concern ourselves with the possibility that every sentence in the language refers to some complicated mathematical claim or other.

Now given those assumptions, we can identify a moderately thin notion of use. We know that O'Leary uses "green" to describe things that are, by appearance, both green and grue. We also know that when O'Leary makes such a description, Daniels expects the object will be both green and grue. So focus on a notion of use such that the *use* of a predicate just is a function of which objects speakers will typically apply the predicate to, and which properties hearers take those objects to have once they hear the predication. If we wanted to be more precise, we could call this notion of 'use' simply *predication*. When we are doing applied semantics, especially when we are trying to figure out the meaning of predicates, we typically know which objects a speaker is disposed to predicate a predicate of, and that's the salient feature of use. (This is why I said the most accurate heuristic would be meaning is

predication plus naturalness; predication is the bit of use we care about in this context.)

This identification of use wouldn't make any sense if we were engaged in theorising at a much more abstract level. If we are doing radical interpretation, then we have to take non-semantic inputs, and solve simultaneously for the values of the subject term and the predicate term in a (simple) sentence. But when we are just doing applied semantics, and working just on the meaning of a term like "green" in a well-functioning language, we can presuppose facts about the denotation of the subject term in sentences like *S is green*, and presuppose facts about what is the subject and what is the predicate in that sentence, and then we can look at which properties hearers come to associate with that very object on hearing that sentence.

Now that we have a notion of use that's distinct from naturalness, we can ask whether it is plausible that predicate meaning is use (in that sense) plus naturalness. And, quite plausibly, the answer is yes. The arguments in (Sider 2001a) and (Weatherson 2003) in favour of this theory look like, at the very least, good arguments that the theory does the right job in resolving Kripkensteinian problems. The theory is immune to objections based on radical reinterpretations of the language, as in (Williams 2007), because those will be inconsistent with the use so defined. And the theory fits nicely into Lewis's broader theory of meaning, i.e., his metasemantics, which is in turn well motivated. So I think there are good reasons to hold that when we're doing applied semantics, the **U&N Theory** delivers the right verdicts, and delivers them for Lewisian reasons. That's the heart of what's true about the **U&N Theory**, even if it isn't a fully general theory of meaning.

Brian Weatherson
Professor of Philosophy
University of Michigan
brian@weatherson.org

Notes

[1] Holton (2003) is more nuanced, but does tell a similar story in the context of discussing Lewis's account of (potential) semantic indeterminacy. Weatherson (2010) follows Holton in this respect.

[2] As some further evidence for how orthodox the 'orthodox' interpretation is, note that Williams (2007) is a prize winning essay published with two commentaries in the Philosophical Review. That paper takes the orthodox interpretation as its starting point, and neither of the commentaries (Bays (2007) and Hawthorne (2007)) criticise this starting point.

[3] Schwarz (2006) develops his criticism of orthodoxy in more detail, and in English, but it is as yet unpublished.

[4] These points are stressed by Wolfgang Schwarz (2006, 2009). He also notes that in "Putnam's Paradox" Lewis explicitly sets these parts of his theory aside so he can discuss Putnam's arguments on grounds most favourable to Putnam. As Schwarz says, this should make us suspicious of the central role "Putnam's Paradox" plays in defences of the orthodox interpretation. We will return to this point in the section on textual evidence for and against orthodoxy.

A referee notes, correctly, that the phrase 'in the first instance' is doing a lot of work here. That's right; we'll return in much more detail below to Lewisian theories of word meaning, and what role naturalness plays in them.

[5] The view I'm attributing to Lewis is endorsed by one prominent supporter of the orthodox interpretation, namely Ted Sider. See his (2012, 35ff).

[6] If you think sentences with a thousand words are too easy to understand for the argument of this paragraph, make the threshold higher; as long as the threshold is finite, it won't affect the argument.

References

T. Bays. The Problem with Charlie: Some Remarks on Putnam, Lewis and Williams. *Philosophical Review* 116:401–425, 2007.

J. Hawthorne. Craziness and Metasemantics. *Philosophical Review* 116:427–440, 2007.

R. Holton. David Lewis's Philosophy of Language. *Mind and Language* 18:286-295, 2003.

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

D. Lewis. Radical Interpretation. *Synthese* 27:331–344, 1974.

D. Lewis. Languages and Language. In *Minnesota Studies in the Philosophy of Science*, 7:3–35. Minneapolis: University of Minnesota Press, 1975.

D. Lewis. Attitudes De Dicto and De Se. *Philosophical Review* 88: 513–543, 1979.

D. Lewis. Mad Pain and Martian Pain. In Ned Block, editor, *Readings in the Philosophy of Psychology*, pages 216-232. Cambridge: Harvard University Press, 1980.

D. Lewis. New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61: 343–377, 1983.

D. Lewis. Putnam's Paradox. *Australasian Journal of Philosophy* 62: 221-236, 1984.

D. Lewis. *On the Plurality of Worlds*. Oxford: Blackwell Publishers, 1986.

D. Lewis. Meaning without Use: Reply to Hawthorne. *Australasian Journal of Philosophy* 70: 106-110, 1992.

D. Lewis.. Reduction of Mind. In Samuel Guttenplan, editor, *A Companion to the Philosophy of Mind*, pages 412–431. Oxford: Blackwell, 1994. Reprinted in Lewis 1999. References to reprint.

D. Lewis. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, 1999.

W. Schwarz. *Lewisian Meaning without Naturalness*. Unpublished manuscript, 2006.

W. Schwarz. *David Lewis: Metaphysik und Analyse*. Paderborn: Mentis-Verlag, 2009.

T. Sider. Criteria of Personal Identity and the Limits of Conceptual Analysis. *Philosophical Perspectives* 15: 189–209, 2001a.

T. Sider. *Four-Dimensionalism*. Oxford: Oxford University Press, 2001b.

T. Sider. *Writing the Book of the World*. Oxford: Oxford University Press, 2012.

R. Stalnaker. Lewis on Intentionality. *Australasian Journal of Philosophy* 82: 199–212, 2004.

B. Weatherson. What Good Are Counterexamples?. *Philosophical Studies* 115: 1-31, 2003.

B. Weatherson. Vagueness as Indeterminacy. In Richard Dietz and Sebastiano Moruzzi, editors, *Cuts and Clouds: Vaguenesss, its Nature and its Logic*, pages 77–90. Oxford: Oxford University Press, 2010.

J. R. G. Williams. Eligibility and Inscrutability. *Philosophical Review* 116: 361-399, 2007.