Lewisian Meaning without Naturalness

Wolfgang Schwarz <wo@umsu.de>

Draft, January 4, 2006

Abstract. It is widely assumed that objective naturalness plays a central role in Lewis's theory meaning. I argue that this is wrong: naturalness does figure in Lewis's theory, but its role is limited and it could be dropped without great damage.

1 The magnetic conception of meaning

How come "tiger" means *tiger* and "it's raining" that it's raining? Some say it's because of how we use these words in communication; some say it's because competent speakers associate certain descriptions, or 'theories', with them; others say causal relations link the words to their meanings; and some argue that the nature of the meanings themselves plays a role: some things, on this account, are by their very nature more eligible to be meanings than others – they are "reference magnets" [Lewis 2004: 5, Fn.2]. The property *tiger*, for example, is more natural and therefore more eligible to be meant than the property *animal that looks like a tiger from a distance*. The true meanings of our words then are the ones that strike the best compromise between objective naturalness and whatever else enters into the determination of meaning.

The magnetic conception is often credited to David Lewis (see, *inter* many *alia*, [Sider 2001: xxif.], [Weatherson 2003], [Stalnaker 2004b]). And indeed, Lewis writes in "New Work for a Theory of Universals":

> 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. [1983b: 47].

And in "Putnam's Paradox":

> *Ceteris paribus*, an eligible interpretation is one that maximises the eligibility of referents overall. [...] overall eligibility of reference is a matter of degree, making total theory come true is a matter of degree, the two desiderata trade off. The correct, 'intended' interpretations are the ones that strike the best balance. [1984: 65f.]

I will argue that these quotes do not represent Lewis's own view: eligibility, or objective naturalness, plays only a very limited role in Lewis's theory of meaning, and it wouldn't do much damage if it were dropped entirely.

To show this, I will first recapitulate Lewis's account of language and mental content. Then I will explain why he appears to defend a rather different account in the passages just quoted. I will close by discussing, and rejecting, two rather different motivations for the magnetic conception.

2 Lewis on language

Lewis's theory of language, as set out in [1969], [1975], [1979b], [1980a], [1986a: 40–50], [1992], goes roughly as follows.

Language is a conventional means of communication. The conventions of language mostly regulate which sentences may be uttered under which conditions: "it's raining" when it's raining, "I'm being attacked by a tiger" when the speaker is attacked by a tiger, and so on. A convention, in general, is a regularity of behaviour accompanied by certain intentions and expectations. If it is a convention to utter "it's raining" only when it's raining, this means that it is common knowledge within the community that everyone tries to utter "it's raining" only when it's raining. This explains why we can use these words to communicate that it's raining: since the addressee knows that people in general only utter "it's raining" when it's raining, they will regard our utterance as evidence for rain.

Part of this story is the same for all languages. There is always a convention to utter certain words under certain conditions. *Which* words under *which* conditions varies from language to language. The task of a *grammar* (as Lewis uses the term) is to systematise this assignment of 'truth conditions' – possible situations in which a sentence may be uttered in accordance with the convention – to sentences for a particular language: "[t]o plug into its socket in an account of the use of a language, a semantically interpreted grammar has to specify which speakers at which times at which worlds are in a position to utter which sentences truly" [1986a: 41].

Notice what is *not* part of that job description. First, a grammar of English need not model the cognitive mechanisms underlying our linguistic competence. At the current stage of neuroscience, Lewis argues, it would be idle to speculate about these matters. Moreover, it is at least conceivable that different speakers employ different internal grammars; a systematic account of our linguistic conventions should therefore abstract away from individual implementations (see [1969: 183, 199f.], [1975: 178], [1980a: 24], [1992: 151, F.6]).

Second, the semantic values employed by a grammar are not supposed to qualify as *referents* or *meanings* in any traditional sense of these terms. "The object", Lewis explains, "is not that we should find entities capable of deserving names from the established jargon of semantics" [1986a: 41]. That's why he generally avoids speaking of "reference", "meaning" and even "intension", and prefers the neutral "semantic value" (see [1974b], [1980a: §4], [1986a: 41f.]).

Consider reference. No doubt "London" refers to London. But we have no guarantee that a systematic grammar of English will assign the city London as semantic value to the word "London". We don't even have a guarantee that it will assign any value at all to "London" – a Russellian grammar would not. And if it does assign something, that value might not be London, but a set of sets (as in [Montague 1973], see [Lewis 1970a: §VII]), or a function from various 'index coordinates' to things (as in [Lewis 1970a]), or whatever. "Semantic values may be anything, so long as their job gets done", says Lewis [1980a: 26]. Their only job is to be part of a theory that systematically delivers the correct truth conditions for the language – the truth conditions that fit the linguistic conventions. (On subsentential expressions, Lewisian conventions remain silent: there is no convention to mean *London* by "London"; there are only conventions to utter entire sentences under certain conditions.)

Now what if, as seems likely, rather different assignments of semantic values to subsentential expressions deliver exactly the same truth-conditions for all sentences? Which of them is the real grammar of English?

This question would make sense if "the real grammar" meant "the implicit grammar underlying the linguistic competence of all speakers", or "the grammar that assigns to all English words their intuitive referents". In Lewis's project however, the question does not make sense. We may favour a grammar that is "simple, natural, reasonable, easy and good" [1969: 198], but not because we somehow know that The Real Grammar is simple, natural, reasonable, easy and good (see also [1975: 176f.]).

At this stage, we could go magnetic and just *stipulate* that by "the real grammar" we mean the one with the most natural semantic values. I doubt that this would suffice to rule out all alternative grammars. But more importantly, there is little point in this stipulation. Why not stipulate instead that the real grammar is the one whose semantic values are the second-most natural, or the least natural?

Lewis doesn't worry about the indeterminacy of subsentential semantic values. However, he does worry about a potential indeterminacy of truth-conditions for entire sentences. Consider some *very* long and complicated sentence, impossible for us to process and understand: we do not have the right intentions and expectations to establish conventional truth conditions for this sentence. So if the interpretation of all sentences is determined by our conventional intentions and expectations, very long and complicated sentences appear to remain uninterpreted.

This is the problem Lewis discusses in "Meaning without Use" [1992]. To solve it, he appeals to the relative naturalness of grammatical rules: the truth-conditions for very long and complicated sentences are determined by the most natural extrapolation from the truth-conditions of actually used sentences. "Use determines some meanings, those meanings determine the rules, and the rules determine the rest of the meanings" [1992: 149].

This is not magnetism. Lewis doesn't talk about subsentential expressions at all in "Meaning without Use"; in particular, he doesn't say that a correct grammar should assign natural meanings to individual words. Moreover, objective naturalness only enters the picture to settle the truth conditions of extremely long and complicated sentences (or almost only, see [1992: 149f., fn.4]); the meanings of all other sentences – of all sentences

that have any chance of being actually uttered – is directly determined by our linguistic conventions.

So apart from the interpretation of sentences that nobody ever uses, objective naturalness plays no direct role in Lewis's theory of language. It does however play a role in his theory of intentionality. We'll have to look at this next, for if naturalness enters into the interpretation of mental states, it indirectly – via the intentions and expectation that ground linguistic conventions – also enters into the interpretation of language.

3 Lewis on intentionality

Lewis has always been a functionalist about mental content: whether a mental state is a belief that it's raining, a desire for mushroom soup or an intention to utter "it's raining" only when it's raining is determined by the state's causal role, by whether it has the typical causes and effects attributed to that kind of state by folk psychology.

With respect to intentional states, folk psychology, according to Lewis, looks "a lot like Bayesian decision theory", [1979a: 148f.] (see also [1974a: 113f.], [1983b: 49–51], [1986a: 36–38], [1994: 320f.]). That is, it says that people typically behave in a way that would serve their desires according to their beliefs: if somebody wants a mushroom soup and believes she can get one by placing an order, she will typically place an order.

But the principles of decision theory do not suffice to determine the interpretation of mental states. Mary's ordering of a mushroom soup could equally well be explained by assuming that she wants a saucer of mud and believes she will get one by ordering a mushroom soup (see [1986a: 107f.]).

Fortunately, folk psychology goes beyond decision theory. Another part of it concerns not the output of mental states – the behaviour they typically cause –, but their input. "Folk psychology says that beliefs change constantly under the impact of perceptual evidence: we keep picking up new beliefs, mostly true, about our perceptual surroundings" [1994: 320]. People with working eye sight confronting a herd of elephants will typically come to believe that there are animals nearby (see also [1980c: 274], [1983a: 380], [1983b: 50], [1986a: 106], [1994: 299f.], [1997: §2]).

Do those input and output conditions together suffice to rule out all deviant interpretations? Lewis [1983b: 50–52] argues that they do not: our perceptual evidence always leaves open that we inhabit all kinds of counterinductive worlds. Maybe Mary wanted not an ordinary mushroom soup, but a mushroom soup that turns into a kangaroo when unobserved, and now believes to have received one. Sure, she *says* she wanted an ordinary mushroom soup, but maybe she believes "ordinary mushroom soup" means *mushroom soup that turns into a kangaroo when unobserved*? With sufficient ingenuity, such an interpretation can probably be extended to fit all Mary's sensory input and behaviour (and the input and behaviour of her counterparts and other people in the same state, upon which the interpretation of the state also depends for Lewis).

Thus Lewis puts forward another constraint to rule out bizarre counterinductive interpretations: when the input and output of a state is compatible with different interpretations, the most natural ones take precedence (see [1974a: 112f.], [1983b: 52–54],

[1986a: 38f., 107], [1994: 320]).

This constraint is actually a bunch of constraints, only some of which are related to objective naturalness. Others are principles of charity and humanity saying that other people have roughly the same basic desires and expectations as we have: all things equal, people do not have a basic desire for a saucer of mud, to use Lewis's favourite example. But a saucer of mud is hardly less natural in any objective sense than a saucer of mushroom soup. Likewise, not all bizarre distributions of beliefs – say, about the history of mankind – necessarily involve very unnatural properties.

The precise status of the naturalness and charity principles in Lewis's theory is not entirely clear. For one, as Lewis was agnostic about the Language of Thought, interpreting a mental state did not mean, for him, assigning (natural or unnatural) properties to mentalese predicates; interpretations always assign content holistically to entire brain states. More importantly, a reductive theory of intentionality should not make reference to the content of our own states. We should therefore replace the principle that people share our own basic beliefs and desires by its individual instances: by the principle that people do not have a basic desire for a saucer of mud, etc.

We could do the same with the constraints involving naturalness: it is part of folk psychology that people in general do not believe that soup turns into a kangaroo when unobserved. According to Lewis, those principles are derived from a single general principle saying that content is to be objectively natural. But it isn't obvious that we really need that general principle. (Lewis thought it was ultimately up to fundamental physics to discover the natural properties. Hence if folk psychology said that fundamental physics is the ultimate authority on the content of our beliefs and desires, this would be a point in favour of the general principle. But I doubt that folk psychology says that.)

One might also wonder whether the more harmless charity and humanity principles don't already suffice to rule out most deviant interpretations left over from the input and output constraints. A desire for a mushroom soup that turns into a kangaroo when unobserved is certainly even more unusual than a desire for mud. Note, incidentally, that many theories of intentionality try to do with input constraints *alone*. Thus I don't believe that dropping the objective naturalness constraint from Lewis's account would result in any serious amount of indeterminacy.

But suppose I'm wrong here. Suppose without the naturalness constraint, the content of our mental states is quite indeterminate, so that it is indeterminate whether I expect you to utter "I'd like a mushroom soup" in situations where you'd like a mushroom soup or in situations where you'd like a mushroom soup that turns into a kangaroo when unobserved. Then the truth conditions of "I'd like a mushroom soup" might well also remain indeterminate between these two interpretations. If so – and I suspect this is what Lewis believed – it is the greater objective naturalness of *mushroom soup* over *mushroom soup that turns into a kangaroo when unobserved* that makes "mushroom soup" mean the former and not the latter.

But that is not because there is a naturalness constraint on the interpretation of our words. The constraint is only on the holistic interpretation of mental states, including the states that ground our linguistic conventions, which indirectly affects the truth conditions of our sentences and thereby, even more indirectly, also the assignment of semantic values

to individual words.

However (*if* objective naturalness plays an important role in the determination of mental content, which I doubt), the result does look a bit like the magnetic conception. This partly explains, I think, why at three places in his writings, Lewis appears to endorse that conception: in "Putnam's Paradox" [1984], pp.45–49 of "New Work for a Theory of Universals" [1983b], and in footnote 6 of "Many but Almost One" [1993]. The other part of the explanation is that Lewis didn't actually believe what he was writing there.

4 Global Descriptivism and "Putnam's Paradox"

To follow their linguistic conventions, the members of a community have to know, at least implicitly, under what conditions their sentences may be uttered. These 'truth conditions' are not the possible worlds at which the sentence is true – at least not on the now standard usage of that phrase. For one, more often than not, the conditions are not conditions on entire worlds, but on utterance contexts. That is why perhaps you can truly say "it's raining" while I cannot, even though we inhabit the same world (if it's raining at your place but not at mine). Secondly, the conditions I'm talking about do not always reflect the sentence's modal status: it is not common knowledge among competent speakers of English that "Tony Blair is actually the current prime minister", or "water is H_2O", may be uttered no matter what, even though these sentences are necessarily true (at least on some reading). We know that our world – or rather, our present context – must be a certain way in order for those sentences to be true, and that it is an empirical question whether it really is that way. We know, for example, that "water is H_2O" may be truly uttered only if the watery stuff of our surroundings is H_2O.

This is how descriptivism fits into Lewis's theory of language. It is not an alternative to the convention-based semantics, but a part of it.

Notice that in this kind of descriptivism, what people associate with linguistic expressions are *conditions*, not *further expressions* (compare [Jackson 1998: 201–204]). That we happen to speak a fairly rich language in which we can express the truth-conditions of many sentences by means of different sentences is, linguistically, a mere coincidence. It is a coincidence that however enables us to wonder how things described in one fragment of our language relate to things described in another fragment – how, say, moral or phenomenal truths relate to facts we can express in physical vocabulary. This explains why in Lewis's writings, descriptivism mainly comes up in metaphysics (like [1966], [1989], [1994]), not in the philosophy of language. (Even "How to define theoretical terms" [1970b] is in the "Ontology" section of Lewis's *Philosophical Papers I*, not in the "Language" section.)

In "Putnam's Paradox" [1984], Lewis discusses a more radical account on which all our words indeed get their meaning by the *linguistic descriptions*, or theories, in which they figure – including the words that make up those theories. The idea is that the correct interpretation of our words is the one that makes our total theory come true. This Lewis calls *Global Descriptivism*.

Global Descriptivism has seemed attractive to some, perhaps because it promises to cash out the idea that all words ultimately owe their meaning to some kind of definition. But on closer scrutiny, it is a really bad theory. To mention just a few of its most obvious problems: it blatantly ignores just about everything one might reasonably assume to enter into the determination of meaning – the conventional (and non-conventional) use of a language, the mental states of speakers and hearers, causal connections between words and things, and so on. Second, it doesn't even begin to work unless certain words – the 'logical operators' – are exempted from the descriptivist account; presumably their meaning is settled by divine intervention. Then the theory makes expressive redundancy a precondition of meaningful languages, which is absurd: consider a fragment of English that only contains "it's raining", "it's snowing", and "the sun is shining". This is a possible language that could be used in a community to convey information about the weather; but none of its terms are definable by means of any others (and certainly not any interpretation that assigns something true to one of the sentences and something arbitrary to the others is acceptable). Fourth, even for very rich languages, Global Descriptivism leaves the interpretation of all sentences radically underdetermined: as long as our total theory is consistent, we can always find an interpretation in, say, arithmetic that makes all its claims true.

This fourth problem is what Lewis discusses in "Putnam's Paradox". To avoid it, he suggests that one might use objective naturalness of referents to constrain interpretations. Thus we arrive at the magnetic account I quoted in the beginning:

> overall eligibility of reference is a matter of degree, making total theory come true is a matter of degree, the two desiderata trade off. The correct, 'intended' interpretations are the ones that strike the best balance. [1984: 65f.]

I don't think magnetism can solve the fourth problem of Global Descriptivism, let alone the other problems. It certainly won't help for relatively poor languages where 'total theory' might be something like {"it's snowing"}. And it is far from obvious that an arithmetical interpretation has to assign particularly unnatural referents to our words.

So why does Lewis defend magnetising Global Descriptivism? Partly, as I explained in the last section, because he thought the magnetised version contains a grain of truth: in Lewis's own account, objective naturalness also plays a certain role in the determination of meaning, albeit a much more indirect and limited role. (Arithmetical interpretations are certainly ruled out long before we wield the naturalness constraint.) Moreover, it isn't unusual for Lewis to defend a theory he didn't endorse. Lewis was always interested in working out and strengthening alternatives to his own accounts. (Recall that he once published a paper under the pseudonym of his cat, [LeCatt 1982], to attack one of his own theories.)

Anyway, Lewis certainly did not endorse Global Descriptivism, either with or without the naturalness constraint: Global Descriptivism flatly contradicts almost everything he wrote elsewhere on language, both before and after 1984 (like [1986a: 40–50] and [1992]). At most, we could assume that for a very brief period around 1984, Lewis gave up the brilliant account of language he had developed in great detail since the mid sixties

and replaced it by an utterly ridiculous alternative, without anywhere pointing out this change of mind and only to return to his old view shortly afterwards. This is incredible.

What's more, if we look at the beginning of "Putnam's paradox", we find a couple of caveats. One is that Lewis will assume for the sake of Putnam's argument that a simple model-theoretic semantics that assigns referents – things and properties – to our words will do for natural language. As we saw above, Lewis did not himself believe that. Here is the other caveat:

> 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. [1984: 57f.]

He adds a footnote: "For a discussion of the 'relocated' problem and its solution, see the final section of my 'New Work for a Theory of Universals' " [1984: 58].

Looking up the final section of "New Work for a Theory of Universals", we find – after an abridged version of "Putnam's Paradox" on pp.45–49 and again the note that the problem rests on a misguided view of language – exactly the story about intentionality I've told in the previous section: that mental content is determined by folk-psychological causal roles, that however the folk psychological input and output conditions do not suffice to rule out all devious interpretations, wherefore principles of naturalness and charity and humanity are needed as further constraints. This, then, is the properly relocated problem and its solution.

Robert Stalnaker [2004b] reads the remark just quoted as indicating that Lewis endorsed magnetised Global Descriptivism as his theory of mental content. This is even more incredible. Again, it contradicts Lewis's numerous writings on intentionality both before and after 1984. It also contradicts what we find when we follow the footnote attached to the remark. In fact, Lewis's theory of intentionality is so far removed from magnetised Global Descriptivism that it isn't even possible to *state* that account in a Lewisian theory of mind: without a Language of Thought, what are the sentences and theories we're supposed to ramsify, what are the terms and predicates we're supposed to assign fairly natural referents?[1]

1 Stalnaker notes some of these problems. How could he nevertheless believe that "Putnam's Paradox", of all places, contains Lewis's final theory of mental content? (Why not consult, say, the section "Content" in "Reduction of Mind" [1994]?)

The answer, I think, is that Stalnaker conflates two notions of *narrow content*. Lewis always maintained that mental content is primarily narrow in the sense that people generally share their beliefs and desires with their functional duplicates (belonging to the same species in the same world), no matter whether those duplicates live on earth, on twin earth, or in a vat (see e.g. [1979a: 142f.], [1994: 312–324]). The primary content of our water perceptions, for instance, is not the presence of H_2O before our eyes, but the presence of *a transparent liquid*. (That the liquid is H_2O is something we have to find out by chemical investigations, not by mere looking.) Stalnaker, by contrast, maintains that the content of our water perceptions is the presence of H_2O, and apparently he believes that the only way to avoid this conclusion is to determine mental content without any recourse to external causes and effects (see in particular [Stalnaker 1993], [Stalnaker 2004a]). This would indeed only leave

5 Motives for magnetism

When Lewis invokes naturalness constraints, it is to resolve indeterminacies of interpretations. He calls it a "Moorean fact" that our language has "a fairly determinate interpretation" [1983b: 47].

That seems right: "it's raining", for example, is definitely about the weather, and not about numbers. But we don't need the naturalness constraint to rule out arithmetical interpretations in Lewis's theory of language. Above I have argued that it is an open question just how much indeterminacy would ensue if we completely dropped Lewis's naturalness constraints. Another open question, I think, is whether that degree of indeterminacy is really objectionable.

Consider the long and complicated sentences. How bad is it to say that their truth conditions are radically indeterminate? How bad is that if we *define* a sentence's "truth conditions" as the class of possible situations in which the sentence may be uttered in accordance with the relevant linguistic conventions? Not too bad, I believe, for there really are no interesting conventions about those sentences.

Suppose we're offered two grammars of English that disagree only about the meaning of sentences too complicated for any human being to parse or understand. Is it really obvious that at least one of these grammars is false? We would certainly reject a grammar that contains extra clauses for such sentences – saying, as it were, that they all mean that God is great. But would we reject those clauses because we know them to be *false* or just because they are useless complications of a theory that already accounts for all the phenomena – and because saving the phenomena is all we want from a grammar? (Remember that our project is not to uncover hidden mechanisms of the brain.)

We should not be lured into rejecting linguistic indeterminacy because of disquotation principles. It is strongly counterintuitive to suggest that "London" does not refer to London or that "it's raining" does not mean that it is raining, or that it is indeterminate between *it's raining* and *it's raining and there have been black dogs*. Of course it is not. But suppose "it's raining" in English is really indeterminate between two interpretations A and B. (By an interpretation I mean an assignment of possible situations, not a translation or an assignment of a meta-linguistic sentence.) Then this also affects our use of "it's raining" in the metalanguage, and it will still be true that "it's raining" is determinately true iff it's raining. So whether or not the meaning of our words and sentences is indeterminate, that "London" refers to London and "it's raining" means that it's raining is not in dispute.[2]

Lewis also does not manage to secure a perfectly determinate interpretation of our thoughts and language. For instance, he doesn't provide a detailed ranking of his var-

room for something like phenomenalism or a mentalese version of Global Descriptivism. But Lewis's mental content is not narrow in this other sense in which narrow content is determined independent of external causes and effects. On the contrary, such causal relations are the main cornerstones of Lewis's folk psychological analysis.

2 If the conventions of English leave the interpretation of "it's raining" indeterminate between A and B, then the conventions of French presumably leave the interpretation of "il pleut" equally indeterminate. So we don't even get indeterminacy of translation: "il pleut" definitely means the same as "it's raining".

ious constraints on the interpretation of mental states. And he explicitly denies that our beliefs can be assigned precise numerical degrees ([1986a: 30]); which entails that *what* we believe is also to some extent indeterminate, for believing something means assigning *sufficiently high* probability to it. Moreover, it is very likely that our linguistic conventions do not completely regulate what should be said in contexts nobody expects to occur – what we should say when it turns out that we are all swampmen or brains in a vat.

All this should not worry us. It is a Moorean fact that our language has a fairly determinate interpretation; but it is a philosophical prejudice that our language has a *perfectly* determinate interpretation.

Other philosophers have put forward a quite different reason for magnetism, one that has little to do with intuitions about determinacy and more with intuitions about fish.

Those philosophers are likely to tell the following story. Scientists once discovered that whales are not really fish and therefore that people had always wrongly called whales "fish". For the term "fish" was always used to pick out a certain *natural kind*, a structurally and functionally homogeneous class of things, one including herrings and carps but excluding whales. In general, to introduce a natural kind term, it suffices to point at one or two exemplars; the term will then automatically come to pick out the most natural class involving those exemplars.

I myself do not believe in this story. But even if it were true, it would not support magnetism. There is an alternative, and superior, explanation of what is going on (in the story); it is that homogeneity is simply part of the criteria by which terms like "fish" are defined: competent speakers know that for something to count as a fish, it must not only look like a fish but also resemble other fish in biological constitution and evolutionary history; "fish" (in the story) *means* "member of the most natural class containing this and that individual", or something like that (compare [Jackson 2003: 95]).

This explanation differs from magnetism in three important respects. First, it does not apply to *all* terms. There's no reason why every predicate should be defined as "member of the most natural class containing ...". Many predicates – "bachelor", "round", "transparent", "French", "natural", etc. – are clearly not: it doesn't make sense to suppose we might find out that someone is not really a bachelor by discovering that paradigmatic bachelors have certain hidden features he lacks.

Second, in the better explanation, "naturalness" is just an umbrella term for many different conditions. Some of our words may pick out *chemical kinds*, others *biological* or *physical* or *sociological* kinds. For every respect in which things can be similar to our designated paradigms, there is a possible 'kind' to be picked out.

Third, unlike magnetism, the alternative explanation does not introduce any secrets into semantics: if some kind of naturalness is part of the analysis of "fish", that is common knowledge in the linguistic community. On the magnetic conception, by contrast, naturalness is an *external* constraint: it may happen that I introduce a predicate, say "poiu" for *being either the number 7 or a member of the Bush administration* which then, unbeknownst to me and anyone else, suddenly means *being a member of the Bush administration* because that is a far more natural class. In a broadly Lewisian semantics, there is no place for such secrets: if the members of a community use a sentence S to

communicate that things are such-and-such – if they intend to utter the sentence when things are such-and-such and expect this from one another – then S cannot secretly mean something else.

6 Conclusion

I have argued that objective naturalness plays only a marginal role in Lewis's theories of linguistic and mental content. Lewis basically completed his theories (with [1969], [1970a], [1975], [1979b] and [1966], [1974a], [1979a], [1980c], [1980b]) *before* he began to believe in objective naturalness at around 1983, and he didn't make any great revisions thereafter. Naturalness is still not mentioned at all in Lewis's summary of his philosophy of language in [1986a: 40–50].

After 1983, Lewis does employ objective naturalness to resolve minor indeterminacies of mental content, and to assign meanings to extremely long and complicated sentences. However, it is unclear whether there is really much indeterminacy here to resolve, and how bad it would be if it were left unresolved.

Another possibility I haven't even discussed is to replace Lewis's objective naturalness constraint by some kind of subjective or language-relative constraint. Lewis is, I believe, too quick in dismissing this as circular (see [1983b: 54f.], [1984: 66f.], [1992: 151]). Consider the problem of long and complicated sentences: if we intuit that their meaning shall not be left indeterminate, why can't we say that their truth conditions are determined by the grammar that is simplest to express in English – in particular as this grammar will surely not contain any of the long and complicated sentences itself? Similarly, I don't think it's hopeless to use our subjective notion of naturalness in an analysis of mental content. I haven't discussed these possibilities because I wanted to stress that the problems they are meant to solve are not very serious in the first place: there is not much work for naturalness – however defined – to do in Lewis's theories of content.

On the other hand, naturalness does play an absolutely crucial role in the magnetised version of Global Descriptivism Lewis defends in "Putnam's Paradox". But magnetised or not, Global Descriptivism is a hopeless theory anyway, and there is no evidence that Lewis ever endorsed it.[3]

References

John Bacon, Keith Campbell und Lloyd Reinhardt (Hg.) [1993]: *Ontology, Causality and Mind: Essays in Honour of D.M. Armstrong*. Cambridge: Cambridge University Press

Frank Jackson [1998]: "Reference and Description Revisited". *Philosophical Perspectives*, 12: 201–218

— [2003]: "From H_2O to Water: The Relevance to A Priori Passage". In H. Lillehammer und G. Rodriguez-Pereira (Hg.) "Real Metaphysics," London: Routledge, 84–97

3 This paper has profited a lot from correspondence with Karl Schaefer, Brian Weatherson and Robert G. Williams. Thanks.

Bruce LeCatt [1982]: "Censored Vision". *Australasian Journal of Philosophy*, 60: 158–162

David Lewis [1966]: "An Argument for the Identity Theory". *Journal of Philosophy*, 63: 17–25. Mit Ergänzungen in David M. Rosenthal (Hg.), *Materialism and the Mind-Body Problem*, Engelwood Cliffs: Prentica-Hall, 1971, und in [Lewis 1983c]

— [1969]: *Convention: A Philosophical Study*. Cambridge (Mass.): Harvard University Press

— [1970a]: "General Semantics". *Synthese*, 22: 18–67. In [Lewis 1983c]

— [1970b]: "How to Define Theoretical Terms". *Journal of Philosophy*, 67: 427–446. In [Lewis 1983c]

— [1974a]: "Radical Interpretation". *Synthese*, 23: 331–344. In [Lewis 1983c]

— [1974b]: "'Tensions". In Milton K. Munitz und Peter K. Unger (Hg.) "Semantics and Philosophy," New York: New York University Press. In [Lewis 1983c]

— [1975]: "Languages and Language". In "Language, Mind, and Knowledge," 3–35. Und in [Lewis 1983c]

— [1979a]: "Attitudes *De Dicto* and *De Se*". *Philosophical Review*, 88: 513–543. In [Lewis 1983c]

— [1979b]: "Scorekeeping in a Language Game". *Journal of Philosophical Logic*, 8: 339–359. In [Lewis 1983c]

— [1980a]: "Index, Context, and Content". In S. Kanger und S. Öhmann (Hg.), *Philosophy and Grammar*, Dordrecht: Reidel, und in [Lewis 1998]

— [1980b]: "Mad Pain and Martian Pain". In Ned Block (Hg.), *Readings in the Philosophy of Psychology* Bd.1, Cambridge (Mass.): Harvard University Press, 216–222, und in [Lewis 1983c]

— [1980c]: "Veridical Hallucination and Prosthetic Vision". *Australasian Journal of Philosophy*, 58: 239–249. In [Lewis 1986b]

— [1983a]: "Individuation by Acquaintance and by Stipulation". *Philosophical Review*, 92: 3–32. In [Lewis 1999]

— [1983b]: "New Work for a Theory of Universals". *Australasian Journal of Philosophy*, 61: 343–377. In [Lewis 1999]

— [1983c]: *Philosophical Papers I*. New York, Oxford: Oxford University Press

— [1984]: "Putnam's Paradox". *Australasian Journal of Philosophy*, 61: 343–377. In [Lewis 1999]

— [1986a]: *On the Plurality of Worlds*. Malden (Mass.): Blackwell

— [1986b]: *Philosophical Papers II*. New York, Oxford: Oxford University Press

— [1989]: "Dispositional Theories of Value". *Proceedings of the Aristotelian Society*, Suppl. Vol. 63: 113-137. In [Lewis 2000]

— [1992]: "Meaning without Use: Reply to Hawthorne". *Australasian Journal of Philosophy*, 70: 106–110. In [Lewis 2000]

— [1993]: "Many, But Almost One". In [Bacon et al. 1993]: 23–38, und in [Lewis 1999]

— [1994]: "Reduction of Mind". In Samuel Guttenplan (Hg.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell, 412–431, und in [Lewis 1999]

— [1997]: "Naming the Colours". *Australasian Journal of Philosophy*, 75: 325–342. In [Lewis 1999]

— [1998]: *Papers in Philosophical Logic*. Cambridge: Cambridge University Press

— [1999]: *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press

— [2000]: *Papers in Ethics and Social Philosophy*. Cambridge: Cambridge University Press

— [2004]: "How Many Lives has Schrödinger's Cat? The Jack Smart Lecture, Canberra, 27 June 2001". *Australasian Journal of Philosophy*, 82: 3–22

Richard Montague [1973]: "The Proper Treatment of Quantification in Ordinary English". In J. Hintikka, J. Moravcsik und P. Suppes (Hg.) "Approaches to Natural Language," Dordrecht: Reidel, 221–242

Theodore Sider [2001]: *Four-Dimensionalism*. Oxford: Clarendon Press

Robert C. Stalnaker [1993]: "Twin Earth Revisited". *Proceedings of the Aristotelian Society*. In [Stalnaker 1999]

— [1999]: *Context and Content*. Oxford: Oxford University Press

— [2004a]: "Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics". *Philosophical Studies*, 118: 299–322

— [2004b]: "Lewis on Intentionality". *Australasian Journal of Philosophy*, 82: 199–212

Brian Weatherson [2003]: "What Good are Counterexamples?" *Philosophical Studies*, 115: 1–31