

What Good are Counterexamples?

The following kind of scenario is familiar throughout analytic philosophy. A bold philosopher proposes that all *F*s are *G*s. Another philosopher proposes a particular case that is, intuitively, an *F* but not a *G*. If intuition is right, then the bold philosopher is mistaken. Alternatively, if the bold philosopher is right, then intuition is mistaken, and we have learned something from philosophy. Can this alternative ever be realised, and if so, is there a way to tell when it is? In this paper, I will argue that the answer to the first question is *yes*, and that recognising the right answer to the second question should lead to a change in some of our philosophical practices.

The problem is pressing because there is no agreement across the sub-disciplines of philosophy about what to do when theory and intuition clash. In epistemology, particularly in the theory of knowledge, and in parts of metaphysics, particularly in the theory of causation, it is almost universally assumed that intuition trumps theory. Shope's *The Analysis of Knowledge* contains literally dozens of cases where an interesting account of knowledge was jettisoned because it clashed with intuition about a particular case. In the literature on knowledge and lotteries it is not as widely assumed that intuitions about cases are inevitably correct, but this still seems to be the working hypothesis.¹ And recent work of causation by a variety of authors, with a wide variety of opinions, generally takes the same line: if a theory disagrees with intuition about a case, the theory is wrong.² In this area exceptions to the rule are a little more frequent, particularly on the issues of whether causation is transitive and whether omissions can be causes, but in most cases the intuitions are taken to override the theories. Matters are quite different in ethics. It is certainly not a good thing for utilitarian theories that we very often feel that the action that maximises utility is *not* the right thing to do. But the existence of such cases is rarely taken to be obviously and immediately fatal for utilitarian theories in the way that, say, Gettier cases are taken to be obviously and immediately fatal for theories of knowledge that proclaim those cases to be cases of knowledge. Either there is some important difference here between the anti-utilitarian cases and the Gettier cases, a difference that justifies our differing reactions, or someone is making a mistake. I claim that it is (usually) the epistemologists and the metaphysicians who are wrong. In more cases than we usually imagine, a good philosophical theory can teach us that our intuitions are mistaken. Indeed, I think it is possible (although perhaps not likely) that the justified true belief (hereafter, JTB) theory of knowledge is so plausible that we should hold onto it in preference to keeping our intuition that Gettier cases are not cases of knowledge.

¹ See, for example, DeRose (1996) and Nelkin (2000).

² See, for example, Menzies (1996), or any of the papers in the special *Journal of Philosophy* issue on causation, April 2000.

My main interests here are methodological, not epistemological. Until the last section I will be arguing for the JTB theory of knowledge, but my main interest is in showing that one particular argument against the JTB theory, the one that turns on the fact that it issues in some rather unintuitive pronouncements about Gettier cases, is not in itself decisive. Still, the epistemological issues are important, which is one reason I chose to focus on the JTB theory, and at the end I will discuss how the methodological conclusions drawn here may impact on them in an unexpected way.

1. Intuitions

Let us say that a **counterexample** to the theory that all *F*s are *G*s is a possible situation such that most people have an intuition that some particular thing in the story is an *F* but not a *G*. The kinds of intuition I have in mind are what George Bealer (1998) calls intellectual “seemings”. Bealer distinguishes intellectual seemings, such as the intuition that Hume’s Principle is true, or that punishing a person for a crime they did not commit is unjust, from physical seemings, such as the ‘intuition’ that objects fall if released, or perhaps that the sun rotates around the earth. We shall be primarily concerned here with intellectual seemings, and indeed I shall only call these intuitions in what follows.

As Bealer notes, whether something seems to be true can be independent of whether we believe it to be true. Bealer himself notes that Frege’s Axiom V seems to be true, though we know it is false. It does not seem to be the case, in the relevant sense, that $643 \times 721 = 463603$. Unless one is rather good at mental arithmetic, there is nothing that 643×721 seems to be; it is out of the reach of intuition. These are not the only ways that seemings and belief can come apart. One can judge that something seems to be the case while neither believing nor disbelieving it. This is a sensible attitude to take towards the view that one cannot *know* that a particular ticket will lose in a fair lottery. This is despite the fact that it certainly *seems* one cannot know this. If one’s intuitions are running rampant, one may even have an intuition about something that one believes to be strictly indeterminate. For example, some people may have the intuition that the continuum hypothesis is true, even though they believe on reflection that it is indeterminate whether it is true.

The distinction between intuitions and belief is important because it helps reduce the violence that revisionary philosophical views do to our pre-existing positions. When I say that Gettier cases may be cases of knowledge, I am not denying that there is a strong intuition that they are not cases of knowledge. I am not denying that a Gettier case does not *seem* to be a case of knowledge. The same thing occurs in ethics. Utilitarians rarely deny that it seems that punishing innocents is the wrong thing to do. They urge that in certain, rare, cases this might be

one of those things that seems to be true despite being false. The case that knowledge is justified true belief is meant to be made in full awareness of the fact that certain cases of justified true beliefs seem to not be cases of knowledge.

Actually, although we will not make much of it here, this last claim is not true as a general statement about all people. Stephen Stich and Shaun Nichols have reported (2001) that the intuition that Gettier cases are not cases of knowledge is not universally shared. It is not entirely clear what the philosophical relevance of these discoveries is. It *might* show that we who have Gettier intuitions speak a different language from those who do not. It *might* show (though as Stich and Nichols point out it is rather hard to see how) that philosophers know a lot more about knowledge than other folk. I think it is rather unlikely that this is true, but we shall bracket such concerns for now, and continue on the assumption that all parties have the Gettier intuitions. Since I shall want to argue that knowledge may still be justified belief in any case, I am hardly tilting the playing field in my direction by making this assumption.

Given that intuitions are what Bealer calls intellectual seemings, and given that the example of Axiom V shows that seemings can be mistaken, what evidence have we that they are not mistaken in the cases we consider here? Arguably, we have very little indeed. Robert Cummins (1998) argues that in general intuition should not be trusted as an evidential source because it cannot be calibrated. We wouldn't have trusted the evidence Galileo's telescope gave us about the moon without an independent reason for thinking his telescope reliable. Fortunately, this can be done; we can point the telescope at far away terrestrial mountains, and compare its findings with the findings of examining the mountains up close and personal. There is no comparable way of calibrating intuitions. Clearly we should be suspicious of any method that has been tested and found unreliable, but there are tricky questions about the appropriate level of trust in methods that have not been tested. Ernest Sosa (1998) argues in response to Cummins that this kind of reasoning leads to an untenable kind of scepticism. Sosa notes that one can make the same point about perception as Cummins makes about intuition: we have no independent way of calibrating perception as a whole. There is a distinction to be drawn here, since perception divides into natural kinds, visual perception, tactile perception, etc, and we can use each of these to calibrate the others. It is hard to see how intuitions can be so divided in ways that permit us to check some kinds of intuitions against the others. In any case, the situation is probably worse than Cummins suggests, since we know that several intuitions are just false. It is interesting to note the many ways in which intuition does, by broad agreement, go wrong.

Many people are prone to many kinds of systematic **logical** mistakes. Most famously, the error rates on the Wason Selection Task are disturbingly large. Although this test directly measures beliefs rather than intuitions, it

seems very likely that many of the false beliefs are generated by mistaken intuitions. As has been shown in a variety of experiments, the most famous of which were conducted by Kahneman and Tversky, most people are quite incompetent at **probabilistic** reasoning. In the worst cases, subjects held that a conjunction was more probable than one of its conjuncts. Again, this only directly implicates subjects' beliefs, but it is very likely that the false beliefs are grounded in false intuitions. (The examples in this paragraph are discussed in detail in Stich 1988 and 1990.)

As noted above, most philosophers would agree that many, if not most, people have mistaken **moral** intuitions. We need not agree with those consequentialists who think that vast swathes of our moral views are in error to think that (a) people make systematic moral mistakes and (b) some of these mistakes can be traced to mistaken intuitions. To take the most dramatic example, for thousands of years it seemed to many people that slavery was morally acceptable. On a more mundane level, many of us find that our intuitive judgements about a variety of cases cannot be all acceptable, for it is impossible to find a plausible theory that covers them all.³ Whenever we make a judgement inconsistent with such an intuition, we are agreeing that some of our original intuitions were mistaken.

From a rather different direction, there are many mistaken **conceptual** intuitions, with the error traceable to the way Gricean considerations are internalised in the process of learning a language. Having learned that it would be improper to use *t* to describe a particular case, we can develop the intuition that this case is not an *F*, where *F* is the property denoted by *t*. For example, if one is careless, one can find oneself sharing the intuition expressed by Ryle in *The Concept of Mind* that morally neutral actions, like scratching one's head, are neither voluntary nor involuntary. The source of this intuition is the simple fact that it would be odd to describe an action as voluntary or involuntary unless there was some reason to do so, with the most likely such reason being that the action was in some way morally suspect. The fact that the intuition has a natural explanation does not stop it being plainly false. We can get errors in conceptual intuitions from another source. At one stage it was thought that whales are fish, that the Mars is a star, the sun isn't. These are beliefs, not intuitions, but there are clearly related intuitions. Anyone who had these beliefs would have had the intuition that in a situation like *this* (here demonstrating the world) the object in the Mars position was a star, and the objects in the whale position were fish. The empirical errors in the person's belief will correlate to conceptual errors in their intuition. To note further that the kind of error being made here is conceptual not empirical, and hence the kind of error that occurs in intuition,

³ The myriad examples in Unger (1996) are rather useful for reminding us just how unreliable our moral intuitions are, and how necessary it is to employ reflection and considered judgement in regimenting such intuitions.

note that we need not have learned anything new about whales, the sun or Mars to come to our modern beliefs. (In fact we did, but that's a different matter.) Rather, we need only have learned something about the vast bulk of the objects that are fish, or stars, to realise that these objects had been wrongly categorised. The factor we had thought to be the most salient similarity to the cases grouped under the term, being a heavenly body visible in the night sky for 'star', living in water for 'fish', turned out not to be the most important similarity between most things grouped under that term. So there is an important sense in which saying whales are fish, or that the sun is not a star, may reveal a conceptual (rather than an empirical) error.

There seems to be a link between these two kinds of conceptual error. The reason we say that the Rylean intuitions, or more generally the intuitions of what Grice (1989: Ch. 1) called the Type-A philosophers, are mistaken is that the rival, Gricean, theory attaches to each word a relatively natural property. There is no natural property that actions satisfy when, and only when, we ordinarily describe them as voluntary. There is a natural property that covers all these cases, and other more mundane actions like scratching one's head, and that is the property we now think is denoted by 'voluntary'. This notion of naturalness, and the associated drive for systematicity in our philosophical and semantic theories, will play an important role in what follows.

2. *Correcting Mistakes*

The following would be a bad defence of the JTB theory against counterexamples. We can tell that all counterexamples to the JTB theory are based on mistaken intuitions, because the JTB theory is true, so all counterexamples to it are false. Unless we have some support for the crucial premise that the JTB theory is true, this argument is rather weak. And that support should be enough to not only make the theory *prima facie* plausible, but so convincing that we are prepared to trust it rather than our judgements about Gettier cases.

In short, the true theory of knowledge is the one that does best at (a) accounting for as many as possible of our intuitions about knowledge while (b) remaining systematic. A 'theory' that simply lists our intuitions is no theory at all, so condition (b) is vital. And it is condition (b), when fully expressed, that will do most of the work in justifying the preservation of the JTB theory in the face of the counterexamples.

The idea that our theory should be systematic is accepted across a wide range of philosophical disciplines. This idea seems to be behind the following plausible claims by Michael Smith: "Not only is it a platitude that rightness is a property that we can discover to be instantiated by engaging in rational argument, it is also a platitude that such arguments have a characteristic coherentist form." (1994: 40) The second so-called platitude just points out that it is a standard way of arguing in ethics to say, you think we should do *X* in circumstances *C*₁,

circumstances C_2 are just like C_1 , so we should do X in C_1 . The first points out that not only is this standard, it can yield surprising ethical knowledge. But this is only plausible if it is more important that final ethics is systematic than that first ethics, the ethical view delivered by intuition, is correct. In other words, it is only plausible if ethical intuitions are classified as mistaken to the extent that they conflict with the most systematic plausible theory. So, for example, it would be good news for utilitarianism if there was no plausible rival with any reasonable degree of systematicity.

This idea also seems to do important work in logic. If we just listed intuitions about entailment, we would have a theory on which disjunctive syllogism (A and $\sim A \vee B$ entail B) is valid, while *ex falso quodlibet* (A and $\sim A$ entail B) is not. Such a theory is unsystematic because no concept of entailment that satisfies these two intuitions will satisfy a generalised transitivity requirement: that if C and D entail E , and F entails D then C and F entail E . (This last step assumes that $\sim A$ entails $\sim A \vee B$, but that is rarely denied.) Now one can claim that a theory of entailment that gives up this kind of transitivity can still be systematic enough, and Neil Tennant (1992) does exactly this, but it is clear that we have a serious cost of the theory here, and many people think avoiding this cost is more important than preserving all intuitions.

In more detail, there are four criteria by which we can judge a philosophical theory. First, counterexamples to a theory count against it. While a theory can be reformist, it cannot be revolutionary. A theory that disagreed with virtually all intuitions about possible cases is, for that reason, false. The theory: *X knows that p iff X exists and p is true* is systematic, but hardly plausible. As a corollary, while intuitions about any particular possible case can be mistaken, not too many of them could be. Counterexamples are problematic for a theory, the fewer reforms needed the better, it's just not that they are not fatal. Importantly, not all counterexamples are as damaging to a theory as others. Intuitions come in various degrees of strength, and theories that violate weaker intuitions are not as badly off as those that violate stronger intuitions. Many people accept that the more obscure or fantastic a counterexample is, the less damaging it is to a theory. This seems to be behind the occasional claim that certain cases are "spoils to the victor" – the idea is that the case is so obscure or fantastic that we should let theory rather than intuition be our guide. Finally, if we can explain why we have the mistaken intuition, that counts for a lot in reducing the damage the counterexample does. Grice did not just assert that the theory on which an ordinary head scratch was voluntary was more systematic than the theory of voluntariness Ryle proposed, he provided an explanation of why it might seem that his theory was wrong in certain cases.

Secondly, the analyses must not have too many *theoretical* consequences which are unacceptable. Consider Kahneman and Tversky's account of how agents actually make decisions, prospect theory, as an analysis of 'good

decision'. (Disclaimer: This is not how Kahneman and Tversky intend it.) So the analysis of 'good decision' is 'decision authorised by prospect theory'. It is a consequence of prospect theory that which decision is "best" depends on which outcome is considered to be the neutral point. In practice this is determined by contextual factors. Redescribing a story to make different points neutral, which can be done by changing the context, licences different decisions. I take it this would be unacceptable in an analysis of 'good decision', even though it means the theory gives intuitively correct results in *more* possible cases than its Bayesian rivals⁴. In general, we want our normative theories to eliminate arbitrariness as much as possible, and this is usually taken to be more important than agreeing with our pre-theoretic intuitions about particular cases. Unger uses a similar argument in *Living High and Letting Die* to argue against the reliance on intuitions about particular cases in ethics. We have differing ethical intuitions towards particular cases that differ only in the conspicuousness of the suffering caused (or not prevented), we know that conspicuousness is not a morally salient difference, so we should stop trusting the particular intuitions. (Presumably this is part of the reason that we find Tennant's theory of entailment so incredible, *prima facie*. It is not just that violating transitivity seems unsystematic, it is that we have a theoretical intuition that transitivity should be maintained.)

Thirdly, the concept so analysed should be theoretically significant, and should be analysed in other theoretically significant terms. This is why we now analyse 'fish' in such a way that whales aren't fish, and 'star' in such a way that the sun is a star. This is not just an empirical fact about our language. Adopting such a constraint on categories is a precondition of building a serious classificatory scheme, so it is a constraint on languages, which are classificatory schemes *par excellence*. Even if I'm wrong about this, the fact that we do reform our language with the advance of science to make our predicates refer to theoretically more significant properties shows that we have a commitment to this restriction.

Finally, the analysis must be simple. This is an important part of why we don't accept Ryle's analysis of 'voluntary'. His analysis can explain all the intuitive data, even without recourse to Gricean implicature, and arguably it doesn't do *much worse* than the Gricean explanation on the second and third tests. But Grice's theory can explain away the intuitions that it violates, and importantly it does so merely with the aid of theories of pragmatics that should be accepted for independent reasons, and it is simpler, so it trumps Ryle's theory.

My main claim is that even once we have accepted that the JTB theory seems to say the wrong thing about Gettier cases, we should still keep an open mind to the question of whether it is true. The right theory of knowledge, the one that attributes the correct meaning to the word 'knows', will do best on balance at these four

⁴ A point very similar to this is made in Horowitz (1998).

tests. Granted that the JTB theory does badly on test one, it seems to do better than its rivals on tests two, three and four, and this may be enough to make it correct.

3. *Naturalness in a theory of meaning*

Let's say I have convinced you that it would be better to use 'knows' in such a way that we all now assent to "She knows" whenever the subject of that pronoun truly, justifiably, believes. You may have been convinced that only by doing this will our term pick out a natural relation, and there is evident utility in having our words pick out relations that carve nature at something like its joints. Only in that way, you may concede, will our language be a decent classificatory scheme of the kind described above, and it is a very good thing to have one's language be a decent classificatory scheme. I have implicitly claimed above that if you concede this you should agree that I will have thereby corrected a *mistake* in your usage. But, an objector may argue, it is much more plausible to say that in doing so I simply changed the meaning of 'knows' and its cognates in your idiolect. The meaning of your words is constituted by your responses to cases like Gettier cases, so when I convince you to change your response, I change the meaning of your words.

This objection relies on a faulty theory of meaning, one that equates meaning with use in a way which is quite implausible. If this objection were right, it would imply infallibilism about knowledge ascriptions. Still, the objection does point to a rather important point. There is an implicit folk theory of the meaning of 'knows', one according to which it does not denote justified true belief. I claim this folk theory is mistaken. It is odd to say that we can all be mistaken about the meanings of our words; it is odd to say that we can't make errors in word usage. I think the latter is the greater oddity, largely because I have a theory which explains how we can all make mistakes about meanings in our own language.

How can we make such mistakes? The short answer is that meanings ain't in the head. The long answer turns on the kind of tests on analyses I discussed in section two. The meaning of a predicate is a property in the sense described by Lewis (1983)⁵: a set, or class, or plurality of possibilia. (That is, in general the meaning of a predicate is its intension.⁶) The interesting question is determining which property it is. In assigning a property to a predicate, there are two criteria we would like to follow. The first is that it validates as many as possible of our

⁵ The theory of meaning outlined here is deeply indebted to Lewis (1983, 1984, 1992).

⁶ There are tricky questions concerning cointensional predicates, but these have fairly familiar solutions, which I accept. For ease of expression here I will ignore the distinction between properties and relations – presumably 'knows' denotes a relation, that is a set of ordered pairs.

pre-theoretic beliefs. The second is that it is, in some sense, simple and theoretically important. How to make sense of this notion of simplicity is a rather complex matter. Lewis canvasses the idea that there is a primitive ‘naturalness’ of properties which measures simplicity and theoretical significance⁷, and I will adopt this idea. Space restrictions prevent me going into greater detail concerning ‘naturalness’, but if something more definite is wanted, for the record I mean by it here just what Lewis means by it in the works previously cited.⁸

So, recapitulating what I said in section two, for any predicate t and property F , we want F meet two requirements before we say it is the meaning of t . We want this meaning assignment to validate many of our pre-theoretic intuitions (this is what we test for in tests one and two) and we want F to be reasonably natural (this is what we test for in tests three and four). In hard cases, these requirements pull in opposite directions; *the* meaning of t is the property which on balance does best. Saying ‘knows’ means ‘justifiably truly believes’ does not do particularly well on the first requirement. Gettier isolated a large class of cases where it goes wrong. But it does very well on the second, as it analyses knowledge in terms of a short list of simple and significant features. I claim that all its rivals don’t do considerably better on the first, and arguably do much worse on the second. (There are considerations pulling either way here, as I note in section seven, but it is *prima facie* plausible that it does very well on the second, which is all that we consider for now.) That the JTB theory is the best trade-off is still a live possibility, even considering Gettier cases.

This little argument will be perfectly useless this theory of meaning (owing in all its essential features to Lewis) is roughly right. There are several reasons for believing it. First, it can account for the possibility of mistaken intuitions, while still denying the possibility that intuitions about meaning can be systematically and radically mistaken. This alone is a nice consequence, and not one which is shared by every theory of meaning on the market. Secondly, as was shown in sections one and two, it seems to make the right kinds of predictions about when meaning will diverge from intuitions about meaning.

Thirdly, it can account for the fact that some, but not all, disagreements about the acceptability of assertions are disputes about matters of fact, not matters of meaning. This example is from Cummins: “If a child, asked to use ‘fair’ in a sentence, says, “It isn’t fair for girls to get as much as boys,” we should suspect the child’s politics, not his language” (1998: 120). This seems right; but if the child had said “It is fair that dreams are purple”, we would suspect his language. Perhaps by ‘fair’ he means ‘nonsensical’ or something similar. A theory of meaning needs to account for this divergence, and for the fact that it is a vague matter when we say the problem is with the

⁷ ‘Measures’ may be inappropriate here. Plausibly a property is simple because it is natural.

⁸ For more recent applications of naturalness in Lewis’s work, see Langton and Lewis (1998, 2001) and Lewis (2001).

child's language, and when with his politics. In short, saying which disputes are disputes about facts (or values or whatever), and which about meanings, is a compulsory question for a theory of meaning.

The balance theory of meaning I am promoting can do this, as the following demonstration shows. This theory of meaning is determinedly individualistic. Every person has an idiolect determined by her dispositions to apply terms; a shared language is a collection of closely-enough overlapping idiolects. So the child's idiolect might differ from ours, especially if he uses 'fair' to mean 'nonsensical'. But if the idiolect differs in just how a few sentences are used, it is likely that the meaning postulate which does best at capturing his dispositions to use according to our *two* criteria, is the same as the meaning postulate which does best at capturing our dispositions to use. The reason is that highly natural properties are pretty thin on the ground; one's dispositions to use a term have to change quite a lot before they get into the orbit of a distinct natural property. So despite the fact that I allow for nothing more than overlapping idiolects, in practice the overlap is much closer to being exact than on most 'overlapping idiolect' theories.

With this, I can now distinguish which disputes are disputes about facts, and which are disputes about meaning. Given that there is a dispute, the parties must have different dispositions to use some important term. In some disputes, the same meaning postulate does best on balance at capturing the dispositions of each party. I say that here the parties mean the same thing by their words, and the dispute is a dispute about facts. In others, the difference will be so great that different meaning postulates do best at capturing the dispositions of the competing parties. In these cases, I say the dispute is a dispute about meaning.

Now, I can explain the intuition that the JTB theorist means something different to the rest of us by 'knows'. That is, I can explain this intuition away. It seems a fair assumption that the reasonably natural properties will be evenly distributed throughout the space of possible linguistic dispositions. If this is right, then any change of usage beyond a certain magnitude will, on my theory, count as a change of meaning. And it is plausible to suppose the change I am urging to our usage, affirming rather than denying sentences like, "Smith knows Jones owns a Ford" is beyond that certain magnitude. But the assumption of even distribution of the reasonably natural properties is false. That, I claim, is what the failure of the 'analysis of knowledge' merry-go-round to stop shows us. There are just no reasonably natural properties in the neighbourhood of our disposition to use 'knows'. If this is right, then even some quite significant changes to usage will not be changes in meaning, because they will not change which is the closest reasonably natural property to our usage pattern. The assumption that the reasonably natural properties are reasonably evenly distributed is plausible, but false. Hence the hunch that I am trying to change the meaning of 'knows' is plausible, but false.

The hypothesis that when we alter intuitions because of a theory we always change meanings, on the other hand, is not even plausible. When the ancients said “Whales are fish”, or “The sun is not a star”, they simply said false sentences. That is, they said that whales are fish, and believed that the sun is not a star. This seems platitudinous, but the ‘use-change implies meaning-change’ hypothesis would deny it.

It has sometimes been suggested to me that conceptual intuitions should be given greater privilege than other intuitions; that I am wrong to generalise from the massive fallibility of logical, ethical or semantic intuitions to the massive fallibility of conceptual intuitions. Since I am on much firmer ground when talking about these non-conceptual cases, if such an attack were justified it would severely weaken my argument. Given what has been said so far we should be able to see what is wrong with this suggestion. Consider a group of people who systematically assent to “If A then B implies if B then A .” On this view these people are expressing a mistaken logical intuition, but a correct conceptual intuition. So their concept of ‘implication’ doesn’t pick out implication, or at the very least doesn’t pick out our concept of ‘implication’. Now if we are in that group, this summary becomes incoherent, so this position immediately implies that we can’t be mistaken about our logical intuitions. Further, we are no longer able to say that when these people say “If A then B implies if B then A ,” they are saying something false, because given the reference of ‘implies’ in their idiolect, this sentence expresses a true proposition. This is odd, but odder is to come. Assuming again we are in this group, it turns out to be vitally important in debates concerning philosophical logic to decide whether we are engaging in logical analysis or conceptual analysis. It might turn out a correct piece of conceptual analysis of ‘implication’ picks out a different relation to the correct implication relation we derive from purely logical considerations. If logical intuitions are less reliable than conceptual intuitions, as proposed, and assent to sentences like “If A then B implies if B then A ” reveals simultaneously a logical and a conceptual intuition, this untenable conclusion seems forced. I conclude that conceptual intuitions are continuous with other intuitions, and should be treated in a similar way.

4. Keeping Conceptual Analysis

The following would be a bad way to respond to the worry that the JTB theory amounts to a change in the meaning of the word ‘knows’. For the worry to have any bite, facts about the meaning of ‘knows’ will have to be explicable in terms of facts about the use of ‘knows’. But facts about use can only tell us about the beliefs of this community about knowledge, not what knowledge really is. Since different communities adopt different standards for knowledge, we should only trust ours over theirs if (a) we have special evidence that our is correct or (b) we are so xenophobic that we trust ours simply because it is ours. “Many of us care very much whether or cognitive

processes lead to beliefs that are true, or give us power over nature, or lead to happiness. But only those with a deep and free-floating conservatism in matters epistemic will care whether their cognitive processes are sanctioned by the evaluative standards that happen to be woven into our language” (Stich 1988: 109). “The intuitions and tacit knowledge of the man or woman in the street are quite irrelevant. The theory seeks to say what [knowledge] really is, not what folk [epistemology] takes it to be” (Stich 1992: 252)⁹. Facts about use can only give us the latter, so they are not what are relevant to my inquiry.

Stich takes this to be a general reason for abandoning conceptual analysis. Now while I think, and have argued above, that conceptual analysis need not slavishly follow intuition, I do not think that we should abandon it altogether. Stich’s worry seems to be conceptual analysis can only tell us about our words, not about our world. But is this kind of worry coherent? Can we say what will be found when we get to this real knowledge about the world? Will we be saying, “This belief of Smith’s shouldn’t be called knowledge, but really it is”? We need to attend to facts about the meaning of ‘knows’ in order to define the target of our search. If not, we have no way to avoid incoherencies like this one.

To put the same point another way, when someone claims to find this deep truth about knowledge, why should anyone else care? She will say, “Smith really knows that Jones owns a Ford, but I don’t mean what everyone else means by ‘knows’.” Why is this any more interesting than saying, “Smith really is a grapefruit, but I don’t mean what everyone else means by ‘grapefruit’”? If she doesn’t use words in the way that we do, we can ignore what she says about our common word usage. Or at least we can ignore it until she (or one of her colleagues) provides us with a translation manual. But to produce a translation manual, or to use words the way we do, she needs to attend to facts about our meanings. Again, incoherence threatens if she doesn’t attend to these facts but claims nevertheless to be participating in a debate with us. These points are all to be found in Chapter 2 of Jackson (1998).

An underlying assumption of the first reply is that there is a hard division between facts about meaning and facts about the world at large; that a principle like: *No ‘is’ from a ‘means’* holds. This principle is, however, mistaken. All instances of the following argument pattern, where *t* ranges over tokenings of referring terms, are valid.

⁹ The paper from which this quote is drawn is about the content of mental states, so originally it had ‘mental representation’ for ‘knowledge’ and ‘psychology’ for ‘epistemology’. But I take it that (a) this isn’t an unfair representation of Stich’s views and (b) even if it is, it is an admirably clear statement of the way many people feel about the use of intuitions about possible cases, and worth considering for that reason alone.

P1: t refers unequivocally to α .

P2: t refers unequivocally to β .

C: $\alpha = \beta$

For example, from the premise that 'POTUS' refers unequivocally to the President of the United States, and the premise that 'POTUS' refers unequivocally to Bush, we can validly infer that Bush is President of the United States. Since P1 and P2 are facts about meaning, and C is a fact about the world, any principle like *No 'is' from a 'means'* must be mistaken. So this worry about how much we can learn from conceptual analysis, from considerations of meaning, is mistaken.

I call this inference pattern the R-inference. That the R-inference is valid doesn't just show Stich's critique rests on the false assumption *No 'is' from a 'means'*. It can be used to provide a direct response to his critique. The problem is meant to be that conceptual analysis, the method of counterexamples, can at best provide us with claims like: 'knows' refers to the relation *justifiably truly believes*. We want to know facts about knowledge, not about the term 'knows', so the conceptual analyst seems to have been looking in the wrong place. But it is a platitude that 'knows' refers to the relation *knows*. I call such platitudes, that ' t refers to t ', instances of the R-schema¹⁰. We can use the R-schema together with the R-inference to get the kind of conclusion our opponents are looking for.

P1: 'Knowledge' refers unequivocally to the relation *justifiably truly believes*.

P2: 'Knowledge' refers unequivocally to the relation *knows*.

C: The relation *knows* is the relation *justifiably truly believes*.

More colloquially, the conclusion says that knowledge is justified true belief. Everyone agrees (I take it) that conceptual analysis could, in principle, give us knowledge of facts of the form of P1. So the opponents of conceptual analysis must either deny P2, or deny that C follows from P1 and P2. In other words, for any such argument they must deny that the R-schema is true, or that the R-inference is valid. I hope the reader will agree that neither option looks promising.

¹⁰ Horwich (1999: 115-130) discusses similar schema, noting that instances involving words in foreign languages, or indexical expressions, will not be platitudinous. He also notes a way to remove the presumption that there is such a thing as knowledge, by stating the schema as $\forall x$ ('knowledge' refers to x iff knowledge = x). For ease of expression I will stick with the simpler formulation in the text.

5. *Against the Psychologists*

Someone excessively impressed by various results in the psychological study of concepts may make the following objection to the theory of meaning here proffered. “Why think that we should prefer short lists of necessary and sufficient conditions? This seems like another one of those cases where philosophers take their aesthetic preferences to be truth-indicative, much like the ‘taste for desert landscapes’ argument. Besides, haven’t psychologists like Eleanor Rosch shown that our concepts don’t have simple necessary and sufficient conditions? If that’s right, your argument falls down in several different places.”

Strictly speaking, my preference is not just for short lists of necessary and sufficient conditions. But it is, for reasons set out more fully in the next section, for short theories that fit the meaning of some term into a network of other properties. And my argument would fall down if there was no reason to prefer such short theories. And, of course, short lists of necessary and sufficient conditions are paradigmatically short theories. One reason I prefer the JTB analysis to its modern rivals is its brevity. Some of the reasons for preferring short lists are brought out by considering the objections to this approach developed by psychologists. I’ll just focus on one of the experiments performed by Rosch and Mervis, the points I make can be generalised.

Rosch and Mervis (1975) claim that “subjects rate superordinate semantic categories as having few, if any, attributes common to all members.” (p. 20) (A superordinate semantic category is one, like ‘fruit’, which has other categories, like ‘apple’, ‘pear’ and ‘banana’, as sub-categories.) Here’s the experiment they ran to show this. For each of six superordinate categories (‘furniture’, ‘fruit’, ‘weapon’, ‘vegetable’, ‘vehicle’ and ‘clothing’) they selected twenty category members. So for ‘fruit’ the members ranged from ‘orange’ and ‘apple’ to ‘tomato’ and ‘olive’. They then asked a range of subjects to list the attributes they associated with some of these 120 category members. Each subject was presented with six members, one from each category, and for each member had a minute and a half to write down its salient attributes.

[F]ew attributes were given that were true of all twenty members of the category – for four of the categories there was only one such item; for two of the categories, none. Furthermore, the single attribute that did apply to all members, in three cases was true of many items besides those within that superordinate (for example, “you eat it” for fruit). (Rosch and Mervis 1975: 23)

They go on to conclude that the superordinate is not defined by necessary and sufficient conditions, but by a ‘family resemblance’ between members. This particular experiment was taken to confirm that the number of

attributes a member has with other members of the category is correlated with a previously defined measure of prototypicality.¹¹ They claim that the intuition, commonly held amongst philosophers, that there must be some attribute in common to all the members, is explicable by the fact that the highly prototypical members of the category all do share quite a few attributes in common, ranging from 3 attributes in common to the highly prototypical vegetables, to 36 for the highly prototypical vehicles.

One occasionally hears people deride the assumption that there are necessary and sufficient conditions for the application of a term, as if this was the most preposterous piece of philosophy possible. Really, this assumption is no more than the assumption that dictionaries can be written, and without any reason to think otherwise, seems perfectly harmless. Perhaps, though, the Rosch and Mervis experiments provide a reason to think otherwise, a reason for thinking that the conditions of applicability for terms like 'fruit', 'weapon', and perhaps 'knowledge' are Wittgensteinian family resemblance conditions, rather than short lists of necessary and sufficient conditions, the kinds of conditions that fill traditional dictionaries.

When we look closely, we see that the experiments do not show this at all. One could try and knock any such argument away by claiming the proposal is incoherent. The psychologists claim that there are no necessary and sufficient conditions for being a weapon, but something is a weapon iff it bears a suitable resemblance to paradigmatic weapons. In one sense, bearing a suitable resemblance to a paradigmatic weapon is a condition, so it looks like we just have a very short list of necessary and sufficient conditions, a list of length one. Jackson (1998: 61) makes a similar point in response to Stich's invocation of Rosch's experiments. This feels like it's cheating, so I'll move onto other objections. I'll explain below just why it feels like cheating.

Philosophers aren't particularly interested in terms like 'weapon', so these experiments only have *philosophical* interest if the results can be shown to generalise to terms philosophers care about. In other words, if can be shown that terms like 'property', 'justice', 'cause' and particularly 'knows' are cluster concepts, or family resemblance terms. But there is a good reason to think this is false. As William Ramsey (1998) notes, if *F* refers to a cluster concept, then for any proposed list of necessary and sufficient properties for *F*-hood, it should be easy to find an individual which is an *F* but which lacks some of these properties. To generate such an example, just find an individual which lacks one of the proposed properties, but which has several other properties from the cluster. It should be harder to find an individual which has the properties without being an *F*. If the proposed analysis is even close to being right, then having these conditions will entail having enough of the cluster of properties that are

¹¹ In previous work they had done some nice experiments aimed at getting a grip on our intuition that apples are more prototypical exemplars of fruit than olives are.

constitutive of *F*-hood to be an *F*. Note, for example, that all of the counterexamples Wittgenstein (1953) lists to purported analyses of 'game' are cases where something is, intuitively, a game but which does not satisfy the analysis. If game is really a cluster concept, this is how things should be. But it is not how things are with knowledge; virtually all counterexamples, from Gettier on, are cases which are intuitively not cases of knowledge, but which satisfy the proposed analysis. This is good evidence that even if some terms in English refer to cluster concepts, 'knows' is not one of them.

Secondly, Rosch and Mervis's conclusions about the nature of the superordinate categories makes some rather mundane facts quite inexplicable. In this experiment the subjects weren't told which category each member was in, but for other categories they were. Imagine, as seems plausible, one of the subjects objected to putting the member in that category. Many people, even undergraduates, don't regard olives and tomatoes as fruit. ("Fruit on pasta? How absurd!") When the student asks why is this thing called a fruit, other speakers can provide a response. It is not a brute fact of language that tomatoes are fruit. It is not just by magic that we happened to come to a shared meaning for fruit that includes tomatoes, and that if faced with a new kind of object, we would generally agree about whether it is a fruit. It is because we know how to answer such questions. This answer to the *Why is it called 'fruit'?* question had better be a sufficient condition for fruitness. If not, the subject is entitled to ask why having that property makes it a fruit. And unless there are very many possible distinct answers to this question, which seems very improbable, there will be a short list of necessary and sufficient conditions for being a fruit. But for this example, at least, 'fruit' was relatively arbitrary, so there will be a short list of necessary and sufficient conditions for being an *F*, for pretty much any *F*.

Thirdly, returning to 'fruit', we can see that Rosch and Mervis's experiments could not possibly show that many superordinate predicates in English are cluster concepts. For they would, if successful, show that 'fruit' is a cluster concept, and it quite plainly is not. So by *modus tollens*, there is something wrong with their methodology. Some of the other categories they investigate, particularly 'weapon' and 'furniture' *might* be relatively cluster-ish, in a sense to be explained soon, but not 'fruit'. As the OED says, a fruit is "the edible product of a tree, shrub or other plant, consisting of the seed and its envelope." If nothing like this is right, then we couldn't explain to the sceptical why we call tomatoes, olives and so on fruit.

So the conclusion that philosophically significant terms are likely to be cluster concepts is mistaken. To close, I note one way the cluster concept view could at least be coherent. Many predicates do have necessary and sufficient conditions for their applicability, just as traditional conceptual analysis assumed. In other words, they have analyses. However, any analysis must be in words, and sometimes the words needed will refer to quite

recherche properties. The properties in the analysans may, that is, be significantly less natural than the analysandum.

In some contexts, we only consider properties that are above a certain level of naturalness. If I claim two things say my carpet and the Battle of Agincourt, have nothing in common, I will not feel threatened by an objector who points out that they share some gruesome, gerrymandered property, like being elements of {my carpet, the Battle of Agincourt}. Say that the best analysis of *F*-hood requires us to use predicates denoting properties which are below the contextually defined border between the 'natural enough' and 'too gruesome to use'. Then there will be a sense in which there is no analysis of *F* into necessary and sufficient conditions; just the sense in which my carpet and the Battle of Avignon have nothing in common. Jackson's argument feels like a cheat because he just shows that there will be necessary and sufficient conditions for any concept provided we are allowed to use gruesome properties, but he makes it sound like this proviso is unnecessary. If Rosch and Mervis's experiments show anything at all, it is that this is true of some common terms in some everyday-ish contexts. In particular, if we restrict our attention to the predicates that might occur to us within ninety seconds (which plausibly correlates well with some level of naturalness), very few terms have analyses. Thus far, Rosch and Mervis are correct. They go wrong by projecting truths of a particular context to all contexts.

6. In defence of analysis

In the previous section I argued that various empirical arguments gave us no reason to doubt that 'knows' will have a short analysis. In this section we look at various philosophical arguments to this conclusion. One might easily imagine the following objection to what has been claimed so far. At best, the above reasoning shows that if 'knows' has a short analysis, then the JTB analysis is correct, notwithstanding the intuitions provoked by Gettier cases. But there is little reason to think English terms have analyses, as evidenced by the failure of philosophers to analyse even one interesting term, and particular reasons to think that 'knows' does not have an analysis. These reasons are set out by Williamson (2000: Ch. 3), who argues, by appeal to intuitions about a particular kind of case, that there can be no analysis of 'knows' into independent clauses, one of which describes an internal state of the agent and the other of which describes an external state of the agent. This does not *necessarily* refute the JTB analysis, since the concepts of justification and belief in use may be neither internal nor external in Williamson's sense. And if we are going to revise intuitions about the Gettier cases, we may wish to revise intuitions about Williamson's cases as well, though here it is probably safest to *not* do this, because it is unclear just what philosophical benefit is derived from this revision. In response to these arguments I will make two moves: one defensive and one offensive. The

defensive move is to distinguish the assumptions made here about the structure of the meaning of 'knows', and show how these assumptions do not have some of the dreadful consequences suggested by various authors. The offensive move, with which we begin, is to point out the rather unattractive consequences of *not* making these assumptions about the structure of the meaning of 'knows'.

In terms of the concept of naturalness used above, the relation denoted by 'knows' might fall into one of three broad camps:

- (a) It might be rather unnatural;
- (b) It might be fairly natural in virtue of its relation to other, more natural, properties; or
- (c) It might be a primitive natural property, one that does not derive its naturalness from anything else.

My preferred position is (b). I think that the word 'knows', like every other denoting term in English, denotes something fairly natural. And I don't think there are any primitively natural properties or relations in the vicinity of the denotation of this word, so it must derive its naturalness from its relation to other properties or relations. If this is so, we can recover some of the structure of its meaning by elucidating those relationships. If it is correct, that is exactly what I think the JTB theory does. This is not to say that justification, truth or belief are themselves primitively natural properties, but rather that we can make some progress towards recovering the source of the naturalness of knowledge via its decomposition into justification, truth and belief. But before investigating the costs of (b), let us look at the costs of (a) and (c).

I think we can dispense with (c) rather quickly. It would be surprising, to say the least, if knowledge was a primitive relation. That X knows that p can hardly be one of the foundational facts that make up the universe. If X knows that p , this fact obtains in virtue of the obtaining of other facts. We may not be able to tell exactly what these facts are in general, but we have fairly strong opinions about whether they obtain or not in a particular case. This is why we are prepared to say whether or not a character knows something in a story, perhaps a philosophical story, without being told exactly that. We see the facts in virtue of which the character does, or does not, know this. This does not *conclusively* show that knowledge is not a primitively natural property. Electrical charge presumably is a primitively natural property, yet sometimes we can figure out the charge of an object by the behaviour of other objects. For example, if we know it is repulsed by several different negatively charged things, it is probably negatively charged. But in these cases it is clear our inference is from some facts to other facts that are inductively implied, not to facts that are constituted by the facts we know. (Only a rather unreformed positivist would say that

charge is *constituted* by repulsive behaviour.) And it does not at all feel that in philosophical examples we are inductively (or abductively) inferring whether the character knows that *p*.

The more interesting question is whether (a) might be correct. This is, perhaps surprisingly, consistent with the theory of meaning advanced above. I held, following Lewis, that the meaning of a denoting term is the most natural object, property or relation that satisfies most of our usage dispositions. It is possible that the winner of this contest will itself be quite unnatural. This is what happens all the time with vague terms, and indeed it is what causes, or perhaps constitutes, their vagueness. None of the properties (or relations) that we may pick out by 'blue' is much more natural than several other properties (or relations) that would do roughly as well at capturing our usage dispositions, were they the denotation of 'blue'.¹² And indeed none of these properties (or relations) are particularly natural; they are all rather arbitrary divisions of the spectrum. The situation is possibly worse when we consider what Theodore Sider (2001) calls maximal properties. A property *F* is maximal iff things that massively overlap an *F* are not themselves an *F*. So *being a coin* is maximal, since large parts of a coin, or large parts of a coin fused with some nearby atoms outside the coin, are not themselves coins. Sider adopts the following useful notation: something is an *F** iff it is suitable to be an *F* in every respect save that it may massively overlap an *F*. So a coin* is a piece of metal (or suitable substance) that is (roughly) coin-shaped and is (more or less) the deliberate outcome of a process designed to produce legal tender. Assuming that any collection of atoms has a fusion, in the vicinity of any coin there will be literally trillions of coin*s. At most one of these will be a coin, since coins do not, in general, overlap. That is, the property *being a coin* must pick out exactly one of these coin*s. Since the selection will be ultimately arbitrary, this property is not very natural. There are just no natural properties in the area, so the denotation of 'coin' is just not natural.

These kind of considerations show that option (a) is a live possibility. But they do not show that it actually obtains. And there are several contrasts between 'knows', on the one hand, and 'blue' and 'coin' on the other, which suggest that it does not obtain. First, we do not take our word 'knows' to be as indeterminate as 'blue' or 'coin', despite the existence of some rather strong grounds for indeterminacy in it. Secondly, we take apparent disputes between different users of the word 'knows' to be genuine disputes, ones in which at most one side is correct, which we do not necessarily do with 'blue' and 'coin'. Finally, we are prepared to use the relation denoted

¹² I include the parenthetical comments here so as not to prejudge the question of whether colours are properties or relations. It seems unlikely to me that colours are relations, either the viewers or environments, but it is not worth quibbling over this here.

by 'knows' in inductive arguments in ways that seem a little suspect with genuinely unnatural relations, as arguably evidenced by our attitudes towards 'coin' and 'blue'. Let's look at these in more detail.

If we insisted that the meaning of 'knows' must validate *all* of our dispositions to use the term, we would find that the word has no meaning. If we just look at intuitions, we will find that our intuitions about 'knows' are inconsistent with some simple known facts. (Beliefs, being regimented by reflection, *might* not be inconsistent, depending on how systematic the regimentation has been.) For example, the following all seem true to many people.

- (1) Knowledge supervenes on evidence: if two people (not necessarily in the same possible world) have the same evidence, they know the same things.
- (2) We know many things about the external world.
- (3) We have the same evidence as some people who are the victims of massive deception, and who have few true beliefs about their external world.
- (4) Whatever is known is true.

These are inconsistent, so they cannot all be true. We could take any three of these as an argument for the negation of the fourth, though probably the argument from (1) (2) and (3) to the negation of (4) is less persuasive than the other three such arguments. I don't want to adjudicate here which such argument is sound. All I want to claim here is that there is a fact of the matter about which of these arguments is sound, and hence about which of these four claims is false. If two people are disagreeing about which of these is false, at most one of them is right, and the other is wrong. If 'knows' denoted a rather unnatural relation, there would be little reason to believe these things to be true. Perhaps by more carefully consulting intuitions we could determine that one of them is false by seeing that it had the weakest intuitive pull. If we couldn't do this, it would follow that in general there was no fact of the matter about which is false, and if someone wanted to use 'know' in their idiolect so that one particular one of these is false, there would be no way we could argue that they were wrong. It is quite implausible that this is what should happen in such a situation. It is more plausible that the dispute should be decided by figuring out which group of three can be satisfied by a fairly natural relation. This, recall, is just how we resolve disputes in many other

areas of philosophy, from logic to ethics. If there is no natural relation eligible to be the meaning of 'knows', then probably this dispute has no resolution, just like the dispute about what 'mass' means in Newtonian mechanics.¹³

The above case generalises quite widely. If one speaker says that a Gettier case is a case of knowledge and another denies this (as Stich assures us actually happens if we cast our linguistic net wide enough) we normally assume that one of them is making a mistake. But if 'knows' denotes something quite unnatural, then probably each is saying something true in her own idiolect. Each party may make other mistaken claims, that for example what they say is also true in the language of all their compatriots, but in just making these claims about knowledge they would not be making a mistake. Perhaps there really is no fact of the matter here about who is right, but thinking so would be a major change to our common way of viewing matters, and hence would be a rather costly consequence of accepting option (a). Note here the contrast with 'blue' and 'coin'. If one person adopts an idiosyncratic usage of 'blue' and 'coin', one on which there are determinate facts about matters where, we say, there are none, the most natural thing to say is that they are using the terms differently to us. If they insist that it is part of their intention in using the terms to speak the same way as their fellows we may (but only may) revise this judgement. But in general there is much more inclination to say that a dispute over whether, say, a patch is blue is merely verbal than to say this about a dispute over whether X knows that p .

Finally, if knowledge was a completely unnatural relation, we would no more expect it to play a role in inductive or analogical arguments than does grue, but it seems it can play such a role. One might worry here that blueness also plays a role in inductive arguments, as in: The sky has been blue the last n days, so probably it will be blue tomorrow. If blueness is not natural, this might show that unnatural properties can play a role in inductive arguments. But what is really happening here is that there is, implicitly, an inductive argument based on a much narrow colour spectrum, and hence a much more natural property. To see this, note that we would be just as surprised tomorrow if the sky was navy blue, or perhaps of the dominant blue in Picasso's blue period paintings, as if it were not blue at all.

So there are substantial costs to (a) and (c). Are there similar costs to (b)? If we take (b) to mean that there is a decomposition of the meaning of 'knows' into conditions, expressible in English, which we can tell *a priori* are individually necessary and jointly sufficient for knowledge, and such that it is also *a priori* that they represent natural properties, then (b) would be wildly implausible. To take just one part of this, Williamson (2000) notes it is clear that there are some languages in which such conditions cannot be expressed, so perhaps English is such a language

¹³ Note that in that dispute the rivals are quite natural properties, but seem to be matched in their naturalness. In the dispute envisaged here, the rivals are quite unnatural, but still seem to be matched. For more on 'mass', see Field (1973).

too. And if this argument for 'knows' works it presumably works for other terms, like 'pain', but it is hard to find such an *a priori* decomposition of 'pain' into more natural properties. Really, all (b) requires is that there be some connection, perhaps only discoverable *a posteriori*, perhaps not even humanly comprehensible, between knowledge and other more primitively natural properties. These properties need not be denoted by any terms of English, or any other known language.

Most importantly, this connection need not be a decomposition. If knowledge is the most general factive mental state, as Williamson proposes, and being factive and being a mental state are natural properties, then condition (b) will be thereby satisfied. If knowledge is the norm of assertion, as Williamson also proposes, then that could do as the means by which knowledge is linked into the network of natural properties. This last assumes that *being an assertion* is a natural property, and more dangerously that norms as natural, but these are relatively plausible assumptions in general. In neither case do we have a factorisation, in any sense, of knowledge into constituent properties, but we do have, as (b) requires, a means by which knowledge is linked into the network of natural properties. It is quite plausible that for every term which, unlike 'blue' and 'coin' are not excessively vague and do not denote maximal properties, something like (b) is correct. Given the clarifications made here to (b), this is consistent with most positions normally taken to be anti-reductionist about those terms, or their denotata.

7. Naturalness and the JTB theory

I have argued here that the following argument against the JTB theory is unsound.

P1. The JTB theory says that Gettier cases are cases of knowledge.

P2. Intuition says that Gettier cases are not cases of knowledge.

P3. Intuition is trustworthy in these cases.

C. The JTB theory is false.

The objection has been that P3 is false in those cases where following intuition slavishly would mean concluding that some common term denoted a rather unnatural property while accepting deviations from intuition would allow us to hold that it denoted a rather natural property. Peter Klein (in conversation) has suggested that there is a more sophisticated argument against the JTB theory that we can draw out of the Gettier cases. Since this argument is a good illustration of the way counterexamples should be used in philosophy, I'll close with it.

Klein's idea, in effect, is that we can use Gettier cases to argue that *being a justified true belief* is not a natural property, and hence that P3 is after all true. Remember that P3 only fails when following intuition too closely would lead too far away from naturalness. If *being a justified true belief* is not a natural property to start with, there is no great danger of this happening. What the Gettier cases show us, goes the argument, is that there are two ways to be a justified true belief. The first way is where the belief is justified in some sense because it is true. The second way is where it is quite coincidental that the belief is both justified and true. These two ways of being a justified true belief may be natural enough, but the property *being a justified true belief* is just the disjunction of these two not especially related properties.

I think this is, at least, a *prima facie* compelling argument. There are, at least, three important points to note about it. First, this kind of reasoning does not obviously generalise. Few of the examples described in Shope (1983) could be used to show that some target theory in fact made knowledge into a disjunctive kind. The second point is that accepting this argument is perfectly consistent with accepting everything I said above against the (widespread) uncritical use of appeal to intuition. Indeed, if what I said above is broadly correct then this is just the kind of reasoning we should be attempting to use when looking at fascinating counterexamples. Thirdly, if the argument works it shows something much more interesting than just that the JTB theory is false. It shows that naturalness is not always transferred to a conjunctive property by its conjuncts.

I assume here that *being a justified belief* and *being a true belief* are themselves natural properties, and *being a justified true belief* is the conjunction of these. The only point here that seems possibly contentious is that *being a true belief* is not natural. On some forms of minimalism about truth this may be false, but those forms seem quite implausibly strong. Remember that saying *being a true belief* is natural does not imply that has an analysis – truth might be a primitively natural component of this property. And remember also that naturalness is intensional rather than hyperintensional. If all true beliefs correspond with reality in a suitable way, and *corresponding with reality in that way* is a natural property, then so is *being a true belief*, even if truth of belief cannot be explained in terms of correspondence.

This is a surprising result, because the way naturalness was originally set up by Lewis suggested that it would be transferred to a conjunctive property by its conjuncts. Lewis gave three accounts of naturalness. The first is that properties are perfectly natural in virtue of being co-intensive with a genuine universal. The third is that properties are natural in virtue of the mutual resemblance of their members, where resemblance is taken to be a primitive. On either account, it seems that whenever *being F* is natural, and so is *being G*, then *being F and G* will be

natural.¹⁴ The second account, if it can be called that, is that naturalness is just primitive. If the Gettier cases really do show that *being a justified true belief* is not natural, then they will have shown that we have to fall back on just this account of naturalness.

References

- Armstrong, D. M. (1978) *Universals and Scientific Realism*. Cambridge: Cambridge University Press.
- Bealer, George (1998) "Intuition and the Autonomy of Philosophy" in DePaul and Ramsey (1998), pp 201-40.
- Cummins, Robert (1998) "Reflection on Reflective Equilibrium" in DePaul and Ramsey (1998), pp 113-28.
- DePaul, Michael and William Ramsey (1998) *Rethinking Intuition*. Lanham: Rowman & Littlefield.
- DeRose, Keith (1996) "Knowledge, Assertion and Lotteries" *Philosophical Review* 74: 568-79.
- Field, Hartry (1973) "Theory Change and the Indeterminacy of Reference" *Journal of Philosophy* 70: 462-81.
- Grice, H. P. (1989) *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Horowitz, Tamara (1998) "Philosophical Intuitions and Psychological Theory" *Ethics* 108: 367-85.
- Horwich, Paul (1999) *Meaning*. Oxford: Oxford University Press.
- Jackson, Frank (1998) *From Metaphysics to Ethics*. Oxford: Clarendon Press.
- Langton, Rae and David Lewis (1998) "Defining 'Intrinsic'" *Philosophy and Phenomenological Research* 58: 333-345.
- Langton, Rae and David Lewis (2001) "Marshall and Parsons on 'Intrinsic'" *Philosophy and Phenomenological Research* 63: 353-355
- Lewis, David (1983) "New Work for a Theory of Universals" *Australasian Journal of Philosophy* 61: 343-77.
- Lewis, David (1984) "Putnam's Paradox" *Australasian Journal of Philosophy* 62: 221-36.
- Lewis, David (1992) "Meaning Without Use" *Australasian Journal of Philosophy* 70: 106-110.
- Lewis, David (2001) "Redefining 'Intrinsic'" *Philosophy and Phenomenological Research* 63: 381-398.
- Menzies, Peter (1996) "Probabilistic Causation and the Pre-emption Problem" *Mind* 105: 85-117.
- Nelkin, Dana (2000) "The Lottery Paradox, Knowledge, and Rationality" *Philosophical Review* 109: 373-409.
- Ramsey, William (1998) "Prototypes and Conceptual Analysis" in DePaul and Ramsey (1998), pp 161-77.
- Rosch, Eleanor and Carolyn Mervis (1975) "Family Resemblances: Studies in the Internal Structure of Categories" *Cognitive Science* 8: 382-439.
- Ryle, Gilbert (1950) *The Concept of Mind*. New York: Barnes and Noble.

¹⁴ I follow Armstrong (1978) here in assuming that there are conjunctive universals.

- Shope, Robert (1983) *The Analysis of Knowledge* Princeton: Princeton University Press.
- Sider, Theodore (2001) "Maximality and Intrinsic Properties" *Philosophy and Phenomenological Research* 63: 357-364
- Smith, Michael (1994) *The Moral Problem*. London: Blackwell.
- Sosa, Ernest (1998) "Minimal Intuition" in DePaul and Ramsey (1998), pp. 257-69.
- Stich, Stephen (1988) "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity" *Synthese* 74: 391-413.
- Stich, Stephen (1992) "What is a Theory of Mental Representation?" *Mind* 101: 243-63.
- Stich, Stephen and Jonathan Weinburg (2001) "Jackson's Empirical Assumptions" *Philosophy and Phenomenological Research* 62: 637-643.
- Tennant, Neil (1992) *Autologic*. Edinburgh: Edinburgh University Press.
- Unger, Peter (1996) *Living High and Letting Die*. Oxford: Oxford University Press.
- Wittgenstein, Ludwig (1953) *Philosophical Investigations*. London: Macmillan.