Intuitions and Conceptual Analysis: Week Six

The Nature of Counterexamples (Mostly)

1. *Wide Content and Conceptual Analysis*

Lycan argues that the conceptual analysis approach to philosophy of mind is bound to fail. The problem is that mental terms, like 'belief', may have their meaning, in some important sense, determined by the way the world is, not by the way intuitions about the world are. To give an example, we all now think that 'water' refers to H_2O, not to the watery stuff. But this isn't something that could have been revealed by conceptual analysis. Conceptual analysis can only tell us what the folk believe. In the case of 'water', the folk beliefs were either incomplete or false, so they can't be of any help.

The problem with Lycan's story is that it just isn't clear which of the folk beliefs about 'water', or about water, are meant to be false. In fact, we can give a quick little argument that all of these beliefs will turn out to be true.

The folk beliefs about the actual properties of water are all true. Water is the stuff which falls from the sky, fills the oceans, is clear, drinkable and so on. And indeed these beliefs couldn't be false, in some suitably strong sense. If we read in the New York Science Times tomorrow that it has all been a big mistake, that the stuff which fills the oceans *etc.* is not H_2O but rather XYZ, then we will conclude that water is XYZ. Whatever turns out to be the stuff with those properties, it will be water, thus underwriting the infallibility. Indeed, some philosophers argue that water isn't H_2O, just because it doesn't have quite the right properties. The problem cases arise when we compare two bodies, say the Dead Sea and me, such that one contains a higher percentage H_2O, but only the other is described as water. I'm not sure this argument works, but the point that the folk are right about the actual properties of water seems beyond doubt.

So are the folk wrong about the essential properties of water. Again probably not. There is a small problem that the folk don't always use the appropriate methodology to interpret their own modal beliefs, but once they know how to work out modal beliefs, they do agree with Putnam that water is essentially H_2O. And, as Jackson points out, this has to be the case. If the folk overwhelmingly did not share Putnam's account of the case, this would show that Putnam was wrong (as Lewis argues in *Reduction of Mind*.) Putnam (and Kripke) are making appeals to modal intuition, not showing that intuition is mistaken.

So it seems that the folk don't have false beliefs about water. They do, or at least did, have slightly incomplete beliefs about water. Before knowing what the internal structure of water is, they didn't know which thing was water in other possible worlds. But this doesn't show that we can't use conceptual analysis here. Remember that for Jackson's project, the important thing is to say how to give the story about water, and beliefs, given the entire

physical story. Now once we know this, we will know which property fills the water role, so this kind of incompleteness isn't going to be that relevant.

On this reading, the comments Jackson make about multiple realisability are a little redundant. These are meant to show that even if Lycan can make his case about water, the analogy he is trying to push breaks down. Since Lycan can't make his case about water, this is a bit of overkill. Anyway, the point of these comments is to stress that we don't have the intuition that beliefs in other possible world will have to have the same microphysical structure as beliefs in this world. Indeed, we have the contrary intuition that beliefs can have all sorts of different microphysical structures, as long as those structures play a certain functional role. And, just to rub in the redundancy, Jackson reminds us that if we didn't have these functionalist intuitions, this would just show that conceptual analysis refutes functionalism, not that advances in cognitive science refute functionalism.

2. *Two Types of Twin Earths*

See Jackson's comments on page 39

3. *Causal Descriptivism*

See Jackson's comments in footnote 16, page 40

4. *The Modest Role of Conceptual Analysis*

This is a very confusing passage. What, precisely, is meant by saying that conceptual analysis is only allowed to play a modest role? That is, what is wrong with the use Geach makes of conceptual analysis in the argument that he gives. I really have very little to say about this, other than to pose a question. If the conceptual analysis Geach provides is true, what is wrong with using it in an argument? (Could it be that really things don't change?!) If the analysis he provides is not true, why is it acceptable as a bit of conceptual analysis in any role at all, modest or not? Any suggestions?

5. *Recap*

Serious metaphysics is meant to be discriminatory and complete. Being discriminatory means that when listing the ingredients of the world, we don't list all that there is. Being complete means that in some sense, the full story of the world supervenes on the full story told in terms of those things that we list. Since we cash out both completeness and entailment in terms of inter-world supervenience relations, it turns out that a commitment to completeness implies a

commitment to certain entailment theses. In particular, it implies a commitment to there existing an entailment from the story told in terms of the privileged items to the complete story.

But how are we meant to find out if such an entailment exists, or better yet, the details of the entailment. Since the entailment won't be a nice syntactic entailment, like from *A* and *B* to *A*, we must look at the meanings of the terms. (The entailment will be more like that from "This is red" to "This is coloured".) So we must know the meanings of the terms to know which entailments hold. But we can't just stipulate the meanings; this would lead to clearly absurd results. We must use the meanings of the words in their ordinary usage. And that requires doing some conceptual analysis. It requires, that is, us to find which cases the folk regard as *F*s, or whatever, which principles they hold to be true of *F*s, and so on.

Given that this is a interesting and important project, let's look at the details of how it should be done.

6. *Smart on Counterexamples*

This paper from Smart is a contribution to a long exchange on certain cases where it appears the action which will maximise utility is to execute an innocent man. Smart agrees that intuitively, this is the wrong thing to do. So intuitively, it is not always right to perform actions which will maximise utility.

The generality of this kind of example should not be underestimated. It is not as if we are only trading intuitions in the moral case, while in debates about personal identity, or causation, or knowledge, we are reporting observational data. In each case we are reporting intuitions, just in one case they are intuitions about personal identity, or causation, or knowledge, and in the other case they are intuitions about morality. So if moral intuitions can go wrong, it seems possible in principle that non-moral intuitions could also go wrong.

Smart's response is to say that, in the moral case at least, theory is more important than intuitions about specific cases. As Smart puts it, the utilitarian thinks we should test our intuitions against our theory, not our theory against our intuitions. This is a quite misleading thing to say. Why do we accept the theory at all? Well, because we have a *theoretical* intuition that it is right. So what Smart should be saying is that we test our specific intuitions against our theoretical intuitions, and not *vice versa*.

One further point to note is that the disanalogy Smart tries to draw between science and ethics isn't really right. As anyone who's done any science knows, the first thing one thinks when a piece of data conflicting with a generally established principle comes in is, What went wrong? We never throw out a theory on one piece of aberrant data; we assume the data must be wrong if it conflicts with the general theory. In practice, the methods of science are much like what Smart wants the methods of ethics to be.

7. Unger on Counterexamples

Unger is, if this is possible, more extreme than Smart about the importance of theory in ethics. On page 2 of the chapter I distributed, he implies that 'moral common sense' only includes propositions about general moral principles, not instructions on what to do in specific cases.

The bulk of Unger's case in this chapter is built by describing our intuitive reactions to a mass of examples, and arguing that these cannot possibly be consistent. So what looks like the same set of choices, but with a different background of presentation, provoke different reactions. Or at least so he claims. It would be nice to have some scientific studies of the cases he describes. (Not studies of actual runaway trolleys, but studies of people's reaction to the runaway trolley stories!)

It is worthwhile, I think, to spend some time judging one's own reaction to the cases that Unger presents. Of course we are not the 'fresh folk' whose intuitions are most valued, but that's a small loss. For example, I didn't think the conduct in the Foot example seemed all that bad; but the intuition that it is bad is something that Unger takes himself to be explaining away. (We will spend a fair bit of time next week on how to explain away aberrant counterexamples.)

As always in these ethical cases, there is a theoretical intuition which is being relied upon here. In Unger's case the principle even has a name: The Independence of Irrelevant Alternatives, or IIA. The principle says that if A is the best choice out of A, B and some other options, then it is the best choice out of A and B. 'Best' here is deliberately ambiguous, leading to different principles. If we interpret it as 'morally best', we get the moral IIA, which is what Unger relies on. If we interpret it as 'prudentially best', we get the popular decision-theoretic IIA.

Since the moral IIA has such strong intuitive support, that seems enough reason to incorporate it into idealised folk morality. If the folk didn't think IIA was right, then the intuitions Unger adduces would just be odd, not outright inconsistent. The oddest intuition, at least to my mind, is the intuition that in the four-option Skater case, the right thing to do is send in the big guy, but in the two-option case, we should just leave the six people tied on the track to die. If you don't think IIA is right, then this is consistent, but odd.

Interestingly, the IIA is denied by some decision-theorists, though I have never quite understood the reasons why. There is a nice argument to show it must be accepted. Say you think that the best choice out of A, B and C is A, and the best choice out of A, B is B. Now assume that you have to make a two-step choice, first between A and B, and then between that one and C. If you choose A the first time, you will have deliberately chosen something that is sub-optimal in that choice. If you choose B, then despite the fact that this situation seems just like a three-way choice between A, B and C, in which A is the best choice, you won't have ended up with A. Some opponents of the IIA (like Isaac Levi) appear to favour the second option, and deny that these 'dynamic' choices, where you have to make a sequence of decisions, are just like 'static' choices, where you just have to make one choice. (The first option seems more plausible, but no one to my current knowledge has advocated it.) Anyway, the

moral of this story is just that you can hold onto any wacky principle you like, as long as you accept all its wacky consequences.

Digression on Group Decision-Making. More importantly, IIA is almost universally rejected as a principle for *collective* decision-making. This is because of a famous result due to the economist Kenneth Arrow. He showed that no decision procedure could obey the following four principles: (1) Non-dictatorship: the principle is not select a dictator and let her decide; (2) Pareto: If everyone prefers B to A, then A can't be the group choice; (3) Decisiveness: The procedure will lead to a decision, no matter what the individual preference-rankings are (This is misleadingly called 'Rationality' by Arrow); and (4) IIA: If the procedure would choose A out of {A, B}, it would choose A out of {A, B, C, …}. Since (1), (2) and (3) seem so appealing, (4) normally gets dropped. In most democracies, the election procedures violate (4), and only keep (3) by allowing coin-tosses at vital stages! So in some circumstances, IIA is allowed to fall away. I doubt this can be used to rescue the intuitive reactions to the Unger cases, but some might think it opens up an option. *End of Digression.*

Unger also has a principle which he thinks guides our moral choices, his principle of separability. People who are doing ethics might like to consider whether he is right about this. I take it that the philosophical moral he wishes to draw is that since the principle clearly carries no ethical weight, judgements which are made using this principle should also carry no ethical weight.

8. Shope on Counterexamples

In Shope's first chapter, he has a small discussion on the importance of intuitions about possible cases in theory of knowledge. Now in theory of knowledge there are many fewer bold souls who follow Smart and Unger in declaring that our intuitions about possible cases may be just wrong. Indeed in Shope's discussion, there is almost no consideration given to the possibility that our intuitions about the famous cases in theory of knowledge are so mistaken.

He does give some consideration to the idea that analysis might lead us to *replace* our concept of knowledge with a different concept, one more amenable to scientific inquiry. It seems this is his interpretation of some suggestions Lehrer made in the early 70's. This is a quite different idea to that floated by Smart and Unger. It has this much in common, both ideas recommend a change in usage by the folk. Lehrer apparently thinks that the folk concept of knowledge is sceptical; since this is a useless concept, we may as well replace it with a more useful one. (Why we can go on playing the counterexamples game now I have no idea.) Smart and Unger think that the folk are making a mistake, and should correct their usage to avoid this mistake.

Although both suggestions have this much in common, their differences are substantial. On Lehrer's suggestion, our intuitions about particular cases are not *false*, it is just that we are using sub-optimal concepts. On Smart and Unger's view, our intuitions about particular cases are false. It is not that they are suggesting we should

use more informative moral concepts; rather they are suggesting that we are misusing the moral concepts we actually have.

So despite the slightly detailed discussion of Lehrer, who suggests that intuitions about particular cases may be *irrelevant*, although true, Shope gives little consideration to the *really* radical position, that these intuitions are just false. One reason we might have for thinking that there is a distinction between the ethical and epistemic cases is that we don't have the same kind of principles about knowledge that we have about morality. There is nothing like the IIA principle to show that some of our intuitions have got to be wrong.

Is this right? The way Jackson puts the Gettier cases makes it sound like we must have some theories. He says, "Sometimes it has seemed right to accuse [non-Gettierists] of confusion—they haven't properly understood the cases, or they haven't seen the key similarities to other cases where they accept that subjects do not know, or the key differences from cases they accept as cases of knowledge." (32) But saying that a similarity or a difference is key is a theory, hence if Jackson is right we must have some 'pre-theoretical' theories about knowledge.

Knowledge also seems to play a role in folk morality. If we know that a certain harm is about to befall someone, and we can do something about it, we acquire a responsibility to do something about it. I don't think the same holds if we merely truly believe that the harm is about to happen, but this isn't clear. The common law, an embodiment of (one version of) part of folk morality seems to have some references to knowledge, but more to justified belief. If this is right then there is a functional role that knowledge, whatever it is, has to play, so there is a theoretical restriction on theories of knowledge. Should it turn out that no theory can (a) capture our intuitions and (b) play this theoretical role, perhaps the intuitions are false.

9. Me on Counterexamples

It's a platitude that some intuitions are mistaken. As I said last week, intuitions sometimes clash, and in these cases we know at least one party is wrong. So it should be *possible* that intuitions about possible cases are mistaken. To get beyond platitudes, we should look at the kind of cases where intuition does, uncontroversially, go wrong, and see what factors they have in common. In the paper I listed six kinds of cases, perhaps everyone can suggest more. (The kinds of cases overlap, I hope harmlessly.)

Empirical

The folk used to think that the earth is flat, many still implicitly believe some kind of Aristotelian theory of motion, many still explicitly believe that humans are not descended from other primates, and especially not from pond slime. These intuitions are all wrong, though perhaps our distant ancestors are too primitive to be properly called 'pond slime'! These seem like quite different intuitions from the kind of intuitions we discuss in philosophy class, so we'll slide by these.

Logical

In the Wason Selection Task, most subjects do extraordinarily badly. In general, tasks associated with *modus tollens* are not performed at all well. And when it is pointed out to the folk that on the standard picture a contradiction entails everything, the folk often giggle. These problems are usually confined to the folk, but some other problems have made it into the academy. Some otherwise level-headed philosophers seem to think that clear instances of *modus ponens* are invalid! And almost no one in Australia believes disjunctive syllogism any more. But I promised to stick to uncontroversial matters.

Epistemic

These tend to overlap with the logical, but may be distinct. In probabilistic reasoning, it is surprising to find a single instance where the folk opinion is correct. (Though most folk correctly realise the Doomsday argument is bad.) In some fun cases the folk can think that *A* and *B* is more probable than *A*. The folk also hold onto their beliefs way, way too long. In *The Fragmentation of Reason*, Stich discusses a number of cases where the folk clearly don't know how to evaluate diagnostic information. (90% of people with disease *A* who took drug *B* were cured, the rest died; if you have disease *A* should you take drug *B*. Well, 90% looks like good odds!) The fact that modern statistics gives us a theory of rationality which is in conflict with these intuitions doesn't seem to be bad news for modern statistics.

Moral

Since we've spent so long on these already, I'll just skip over them now.

Semantic

The folk are notoriously bad at distinguishing false sentences from sentences that are true but misleading. (In my favourite moment of the whole impeachment debates, one of the House prosecutors denied that there was a distinction here to be drawn!) For instance, many folk think that the sentence *Superman came out of the phone booth and Clark Kent went in*, said of a typical dramatic moment in a Superman story, is false, because Clark Kent went into the phone booth and *then* Superman came out. As we'll discuss next week, this view is clearly crazy; the sentence is true but misleading. (There are famous examples of this concerning academic references which drive the point home well.) Such confusions inhibited debate about knowledge and belief well into the 1960's. Some people thought that whenever *A* knows that *p*, it would be improper to say *A believes that p*, hence *A* didn't believe that *p*. Or, in a snappy soundbite, knowledge excludes belief.

Conceptual

I think that there is an important distinction between these cases and the empirical cases mentioned earlier. The belief that the earth is flat and the belief that Mars is a star are both false, but they are false for different reasons. The people who thought that the earth is flat typically had a good working knowledge of what makes something flat, they just didn't know whether the earth had this property. On the other hand, the people who thought that Mars was a star didn't have a good working knowledge of what it was to be a star. Similarly people who thought that whales were fish, while not being totally acquainted with the nature of whales, could have held onto this belief while learning quite a lot about whales. It seems the bigger mistake these folk were making was about the nature of fish, or the nature of stars, rather than the properties of whales or of Mars. As I put it in the paper, we could have changed our opinion about whales, or Mars, without learning anything at all about the intrinsic nature of those entities.

Best Theories

In each case where intuition goes wrong, it clashes with the best theory. It is worth checking this for each of the examples I stated. I think that what it is for intuition to be mistaken is for the intuition to clash with best theory. Another way to illustrate my point is to look at cases where it isn't uncontroversial that intuitions are mistaken, where it isn't clear which intuition is right. Usually, disputants will try and promote their cause by showing that their intuitions can be systematised into a better theory than their opponents' intuitions. If their implicit methodology has any value, this is good news for my theory. (And if my theory is right, this is good news for those players!)

Now as stated, my claim is trivially true (modulo some anti-idealist worries about best theories being false.) What makes it (a little) substantive is that there is a little story I can tell about what it is for a theory to be best. I claim that the best theory is the theory which does best, on balance, on these four factors.

Fewest Counterexamples

Clearly it is important that a theory have as few counterexamples as possible. Even Unger wouldn't (or perhaps just shouldn't) claim that intuitions about particular cases have *no value whatsoever*. If I'm right, the counterexample game isn't useless, just overrated.

We can say a little more than this about the role of counterexamples. It isn't just add 'em up. (For one thing, counterexamples usually name the subjects, so any ce can be turned into infinitely many by altering the names. And as those who've toiled in these fields know, counting gets so problematic when the values go infinite.) Three factors seem to determine how important a particular ce is.

First, how **familiar** the case is. The more wacky a ce, the more apt the response that intuitions about these cases are not reliable.

Second, how **strong** our conviction about the case is. If we think that something is a ce, but we aren't really sure, we should just let best theory decide. (This is what Armstrong called the 'spoils to the victor' principle.)

Third, how **broad** the scope of the ce seems to make a difference. If a ce applies only to a very specific kind of case, it seems less important than if it applies to a wide variety of cases. (This happens all the time in the literature on causation, but the examples require quite a bit of background. Maybe we will look at them next week.)

What isn't relevant, I think, is whether the cases are **actual**, although this might impact on familiarity.

Fewest Odd Theoretical Consequences

As we saw in the moral case, there are theoretical constraints on analyses. I stress in the paper that all normative analyses are constrained by a rule requiring no arbitrariness. We'll look in more detail next week at the particular example I discuss from Horowitz. One point to note is that we have to distinguish apparent theoretical constraints from similar constraints which are actually true.

As noted above, it is interesting whether a theory of knowledge has *any* interesting theoretical consequences in this sense.

Theoretical Significance

The term, as analysed, should play an important role in our best theory. This, I think, is what explains our current views about glass not being a solid, whales not being fish, and the morning star not being a star. One possible objection to including this as a criteria on analyses is that the belief that the term, as analysed, be theoretically significant may be a theoretical belief. That is, where this criteria applies, it may be redundant.

Simplicity

We haven't seen the cases yet where this criteria applies, so it is a little hard to argue for. One thing to consider, for those who have spent some time on this puzzle, is whether an appeal to simplicity solves Quine's problem about the indeterminacy of translation. If it is a requirement on translations that they be simple, there is a reason to think 'gavagai' refers to rabbits, and not to undetached rabbit parts. Indeed, this move seems to be important for all cases where the pragmatics underdetermines the semantics. But that is a topic for next week.