Intuitions and Conceptual Analysis: Week Five

1. Where We're Up To

An important tradition in metaphysics takes its job to be finding a limited number of ingredients with which we can tell the complete story of the world (or some subject matter). Physicalism, for example, claims that the list of ingredients sufficient to tell the complete story about the very small, or about the non-sentient, is sufficient to tell the complete story about all of the world. Some people take the moral of this kind of metaphysics to be eliminativist; that we can tell the complete story of the world without meanings, or inflations, shows that meaning and inflation do not exist. Most people are not so blasé about rejecting commonsense opinions. Inflations, wars, rivers and beliefs all exist, but there is nothing but atoms in the void, so we must find a way of showing that the arrangement of atoms in the void makes true the stories about inflations and so on.

Jackson notes that one simple way for this to happen is if the story about the atoms entails the story about inflation. This should not be controversial; the sufficiency of this criteria is not seriously questioned. (If the story about atoms is true, and it really does entail a story about inflation, then the story about inflation is true. And one part of the story about inflation is that inflation exists.) What is surprising is that Jackson claims this is a necessary condition. Should the story told in some simple language, say the austerely physical, be complete, then the *only* true sentences are those entailed by that story. Hence inflation (or whatever) only exists if its existence is entailed by the arrangement of atoms in the void.

So the job of the metaphysician who takes the threat of eliminitavism about the social, or the psychological, or the everyday, seriously, is to show that such entailments hold. And doing that job requires conceptual analysis, as we'll now see.

2. Why do Conceptual Analysis?

The Big Picture

Imagine an argument between an eliminitavist about some class of things, say baseballs, and a realist. For simplicity, assume that the disputants agree about the full story of the world told in terms of atoms in the void. But one, the realist, thinks that the full story about atoms in the void is compatible with the existence of baseballs, indeed in some sense makes it true that there are baseballs, while the eliminitavist thinks otherwise. In other words, one asserts, but the other denies, that the sentence "Baseballs exist" is true. How is the dispute to be resolved?

Well, to resolve the dispute we first have to be clear what we are disputing. If the realist just stipulates that by 'baseballs exist' she just means 'there are atoms arranged baseball-like', then the sentence will be true, in her idiolect. If the eliminitavist stipulates that by 'baseballs exist' she means that any complete story of the world must explicitly refer to baseballs, then she will have ensured that the sentence is true. And if someone else stipulates that by 'baseballs exist' she means 'Bill Clinton is President', the sentence will be true, but now it is quite clear why we

shouldn't care. If we just rely on stipulative definitions of sentences under dispute, we will not be resolving the question we originally thought we were discussing.

Since the resolution of the dispute must go through a clarification, or resolution, of the meaning of the disputed sentences, and since this resolution cannot be stipulative, it must be by working out what we always meant by the sentence. In other words, by working out the folk meaning of the sentence. And to do this requires conceptual analysis. So here's the argument in three easy steps.

(a) We need an argument against eliminitavism;
(b) This argument must turn on a clarification of the ordinary usage of terms under dispute;
(c) To clarify ordinary usage, we must do some conceptual analysis.

From now, we turn to the details of the argument.

Having a Reason for not being Eliminitavist

Jackson's first argument turns on the claim that we need something to say to the eliminitavist; to the person who thinks the completeness of the austerely physical story shows that prices and beliefs and chairs don't really exist. It's not immediately clear whether this argument can work. In particular, it isn't clear that we have to give an argument against eliminitavists which isn't question-begging. Compare Lewis's 'argument' against those who believe there are true contradictions.

> The reason we should reject this proposal is simple. No truth does have, and no truth could have, a true negation. Nothing is, and nothing could be literally true and false. This we know for certain, and *a priori*, and without any exception for especially perplexing subject matters. The radical case for relevance should be dismissed just because the hypothesis is requires us to entertain is inconsistent.
>
> That may seem dogmatic. And it is: I am affirming the very thesis that Routley and Priest have called into question and – contrary to the rules of debate – I decline to defend it. Further, I concede that it is indefensible against their challenge. They have called so much into question that I have no foothold on undisputed ground. So much the worse for the demand that philosophers always must be ready to defend their theses under the rules of debate (Lewis, "Logic for Equivocators", pg 101)

The thought here is that we can insist it is a Moorean fact that baseballs and inflations exist, and just refuse to countenance any arguments to the contrary. More conservatively, we might reason as follows:

(i) Baseballs exist; (Moorean fact)
(ii) If baseballs exist then the existence of baseballs is entailed by the arrangement of atoms in the void (From Jackson gives in chapter 1)

(iii) Therefore, the arrangement of atoms in the void entails the existence of baseballs

We don't have to come up with the specific reduction of baseballs to atoms in the void, we just need to know, in the most general terms, that such a reduction exists. (And, we might say in a sneaky voice, this is all we ever get from existing analyses anyway.) It would be nice to have an argument for (i), and analysis offers such an argument, but is it really likely that the premises will be more plausible than the conclusion?

Whose Concepts Are We Talking About?

Even assuming the serious metaphysician must show **how** the story told in primitive vocabulary makes the story told in everyday vocabulary true (as opposed to resting content with showing that it **does**) it isn't clear why we have to do conceptual analysis. Why think that showing how a neurophysiological story makes a psychological story true requires conceptual analysis, rather than ordinary scientific deduction?

The argument here is a little quick. If we stipulate that a belief that snow is white is just a firing of the 427th neuron at a rate of 49Hz, then a bit of scientific investigation could reveal that some people really believe that snow is white. This would clearly be a useless move in a debate with eliminitavists about belief. For that debate was, a little, interesting, but now it has turned into a "easy exercise in deduction from stipulative definitions together with accepted facts." Clearly we must have meant something else by 'belief' when this debate was going on.

What could that be? Jackson says it is the ordinary, or folk, meaning. Now of course it wasn't obvious that the only choices to start with were the folk definition, or the 427th neuron definition. So Jackson's argument can't be an easy disjunctive syllogism. What Jackson must be suggesting is that any proposed meaning other than the folk meaning will be vulnerable to this type of objection. I suppose this is right, but it isn't clear that it is.

More positively, the argument might run like this. We were unselfconsciously having debates about the existence of free will, or the existence of beliefs, without worrying over which definition of 'free will' or 'belief' we were using. So we must have been using these words in something like their ordinary sense. Hence to discover the meanings of these words which are relevant to these debates, we have to go and find their ordinary senses.

Identifying the Ordinary Conception I: The Use of Possible Cases

"But how should we identify our ordinary conception? The only possible answer, I think, is by appeal to what seems to us most obvious and central about free action, determinism, belief, or whatever, as revealed by our intuitions about possible cases." (**Jackson**, 31)

To the extent this is meant as anything more than a platitude, it is clearly mistaken. We'll be looking more into this issue next week, so I won't say much here. But just note quickly that another way to identify the ordinary conception is by asking ordinary folk what their conception is. Or put another way, by asking the folk whether they think certain general principles are true of a concept.

Let's look at a specific case, one which Jackson discusses, to put this more in focus. One way to find out someone's concept of a socialist is to ask them whether they think particular people are socialists. So we might ask Is Warren Beatty a socialist? Is Richard Gephardt a socialist? Is Bill Clinton a socialist? Is Pat Buchanan a socialist?

and so on. Another way, is to ask them directly for general properties of socialists. So we can ask questions like Is anyone who believes the subway should be publically owned a socialist? Is anyone who doesn't support the flat tax a socialist? Is anyone who believes there should be a welfare system a socialist? and so on.

The same thing works in philosophical cases. To find someone's concept of the good, we can ask questions about particular acts, like Is abortion wrong? Is stealing from the rich wrong? Is not giving to charity wrong? and so on. Or we can ask them about particular ethical principles, like Can two actions with identical consequences differ in their ethical status? Does the belief that some action is good provide motivation for carrying out that action? Does the fact that an agent is a member of our species rather than some other species affect that agent's ethical standing? and so on. The distinction I'm drawing here isn't particularly sharp (on what side, for example, does the question Do we have ethical obligations to the non-sentient? fall), but it seems workable in practice.

Finally, just to float a case I say a bit about in "What Good are Counterexamples?", which we'll be looking over next week, compare what Gettier does which the concepts of knowledge and justification in his famous counterexamples to the JTB (justified true belief) theory of knowledge. To show that the cases are not cases of knowledge, he simply appeals to our intuition about the cases, the intuition that they are not cases of knowledge. To show that the cases are cases of justified belief, he can't make a similar appeal to intuition, because no one actually has the intuition that the cases are cases of justified belief. Rather, he has to appeal to a general principle, which intuition says does hold, that justification is preserved when making valid inferences, to show the cases are cases of justified belief. Why, one may wonder, don't we use *modus tollens* here rather than *modus ponens*, and conclude that the intuitive principle is wrong, the cases are not cases of justified belief, so there is no threat to the JTB theory? The answer will have to be that we don't want to give up these theoretical beliefs about some concepts, even when they lead to conflicts with beliefs about particular cases.

In sum, the only content this quote from Jackson could have is that we should identify someone's concept of an *F* by looking to see which cases they classify as *F*s, *rather* than looking at their theoretical commitments concerning *F*ness. This is at best very controversial, and really seems to be a mistake. It turns out that Jackson isn't strongly committed to this reading of the quote, so maybe it is only meant as a platitude.

Something to think about for next week then, is the question of what we should do when our general and specific intuitions clash. As a rough generalisation, we seem to take the general intuitions to have more force when discussing normative concepts, and the specific intuitions to have more force when discussing descriptive concepts. But it's a long way from this *is* to an *ought*.

Identifying the Ordinary Conception II: Divergent Idiolects

Jackson thinks that it is possible we could have similar, but different, concepts. In particular, he suggests that for some people, it is correct to assert that the Gettier cases are cases of knowledge. He thinks that according to *their* concept, the agents in Gettier cases do get knowledge, albeit by luck. If there is a folk concept of knowledge, it is an amalgam of the various concepts different individuals have.

I wonder if this is a plausible dissolution of the dissent about possible cases in the literature on knowledge. As we discussed two weeks ago, it doesn't in general seem right that dissent over possible cases implies lack of a

shared concept. Conservatives and liberals typically differ over what is right and wrong, but this doesn't (always) mean that they have different moral concepts. (And, linking this all together, we can see this in part because they have similar general beliefs about ethics. In particular, they agree about the motivational role of ethical beliefs, which is why ethical debates have such urgency.) So if it is wrong for moral concepts, how could it be right for epistemic concepts?

One possibility is that we don't have any interesting theoretical commitments when it comes to knowledge, so there is nothing we could share if we disagree about the possible cases. Another possibility is that if a disagreement between two people about the description of a possible case has no impact on their non-verbal dispositions, then the dispute should be classified as verbal. And if a dispute should be classified as verbal, this presumably means that the participants to the dispute lack some shared concept. The problem is that (a) distinguishing the verbal intuitions is tricky and (b) even when we do it, this theory seems like it might be false. Say that I think the Battle of Hastings happened in 1055, and you think, correctly, that it happened in 1066. This is a non-verbal disagreement (isn't it?) but it is hard to see what non-verbal dispositions of mine would be altered should I come to share your view.[1]

I can't see what the paragraph about young children contributes to the discussion. My most charitable reading of it is that Jackson is trying to spread the burden of guilt. If all we are doing is what these highly reputable scientists are doing, how can it be that bad?

Finally, (though this is a little out of order) the paragraph about opinion polls on pages 36-7 seems very odd. We will do more on this in the next few weeks, but here's one issue to think about. Conceptual analysis has been advertised, historically, as a way of finding out *a priori* truths. Now it seems that an essential part of doing conceptual analysis is conducting opinion polls. As anyone who's ever taken a poll would know, they are the quintessentially *a posteriori* way of finding out about the world. So how can we recover from this morass the *a priori* knowability of the conceptual truths?

Really finally, do people agree with Jackson about the results of the unscientific poll on the Gettier cases? How many people found their students willing to say the Gettier cases are clearly not cases of knowledge? How about the number of people whose students thought that *ex false quodlibet* is a clearly valid inference rule?

The Subject of Analysis: Words or Concepts?

The passage at the bottom of page 33 where Jackson discusses this question is a bit murky. First, it isn't clear precisely what question he is trying to answer. What is it to study "the words *per se*" rather than the cases they cover? Would it be an investigation into their phonics? Surely what it is to investigate words (or terms) such as 'free action', 'knowledge', 'justice' and so on is to look at what possible cases those words can be properly predicated of.

[1] Perhaps I would be now be disposed to look at a book were I to want to find out about the Battle of Hastings and I believed that book had information on everything that happened in 1066. But how do we distinguish this from cases where we have a real verbal dispute, but coming around to your way of predicating '*t*' of *F*s would instil in me the 'disposition' to look at books concerning everything described by '*t*' should I want to learn about *F*s.

If there is a distinction to be drawn here (and I'm not sure there is) it might be nice were it to turn out that we are investigating words, not concepts. I know why we trust folk intuitions about the meanings of words. Meanings are conventional, so collectively we are somewhat infallible about meaning. (Not entirely infallible, as we'll soon see, but we're guaranteed to be close.) But it isn't clear that the nature of concepts is conventional. (Or is it?) So it isn't clear why we mightn't be completely mistaken about the nature of concepts. In which case the question of whether we should place more trust in general intuitions or intuitions about possible cases becomes fairly moot. We would then have to conduct a full-scale investigation into the question of whether intuitions are reliable *at all*.

Something that people might like to think about, because I don't have much to say about it at this stage. If concepts are distinct from meanings of conceptual terms, then (a) what are concepts? and (b) what are meanings?

Individuation of Concepts

This might be familiar to a lot of people, but it might be worth stating quickly. On page 34 Jackson mentions that even though what is important about concepts is which cases they cover, concepts which (necessarily) cover the same cases may be distinct in some important sense. This isn't just a Kripkean point about necessary and *a priori* truths being distinct. It may be both necessary and *a priori* that concepts cover the same cases, but nevertheless the concepts are distinct.

The example, which should be familiar from high-school geometry, is that it can be proven quickly that in Euclidean space a triangle is equilateral (all sides are the same length) if and only if it is equiangular (all angles are the same size).[2] Despite this, many people seem to think that the concepts *being an equilateral triangle* and *being an equiangular triangle* are distinct concepts. Some evidence for this is that one could grasp one concept without (apparently) grasping the other concept. Some more evidence is that we could (it seems) have a machine which tests for being an equilateral triangle, and this would be a different kind of machine to one which tests for being an equiangular triangle. And, in a related point, it is an interesting proof that all equilateral triangles are equiangular triangles, and it wouldn't be clear why this is interesting if the two concepts are identical.

None of this seems particularly convincing to me, but if it does seem convincing to you, it might be reassuring to know there is a technical way of accommodating such intuitions. Divide the concepts into the simple and the complex concepts. The rough rule is that if a concept is referred to by a single word, or by a non-compositional phrase, it is simple, if the phrase picking out the concept can be understood compositionally, the concept is complex. So *being an equilateral triangle* is complex because we understand it by understanding the concepts *being a triangle* and *being equilateral*. Now, simple concepts are individuated coarsely; if it is necessary *a priori* that *F* and *G* cover the same cases, *F* and *G* are identical. But complex concepts are individuated by their

[2] From memory this also holds in most popular non-Euclidean geometries; but I'm not certain about this. I'm sure it would be *possible* to construct a geometry in which it was false, provided there were odd asymmetries between the properties of different parts of the space. In any case, anyone who is concerned about the rigour of the example can stipulate that we are talking about *Euclidean* triangles at all times.

method of construction from simple concepts. Since *being an equilateral triangle* and *being an equiangular triangle* are made of different parts, they are different. (To prove *being equilateral* and *being equiangular* are distinct, note that rectangles are equiangular without being equilateral, rhombuses are equiangular without being equilateral.)

This is a rather pretty solution. It is also rather old, tracing back to Carnap's *Meaning and Necessity*, which I think came out before 1950. And like most old ideas, it has its critics. I'll just mention one, because we are getting a little sidetracked. In a recent paper, Jerrold Katz has argued that this criteria of individuation slices concepts too finely. (I don't have the reference, but can *possibly* find it if anyone is interested.) He has the intuition that while the concepts *being an equilateral triangle* and *being an equiangular triangle* are distinct, the concepts *being a square* and *being an equilateral equiangular quadrilateral* are identical, presumably because the second phrase could be used as a definition of the first. (Katz is more interested in synonymy of phrases than identity of concepts, but I think the translation is harmless.) I simply don't have the intuition that there is a natural sense in which the first pair of concepts are distinct, but the second pair are identical, so I don't know what to say here. But it would be wrong to think that this question was entirely settled.

3. The Role and Nature of Mistakes

How do the folk make mistakes?

Here's two platitudes about conceptual analysis. First, the meaning of a word is defined by the cases to which the folk apply it. Second, sometimes the folk make mistakes in their application, and not just because they are mistaken about the nature of the case in question. The folk make conceptual mistakes, and not just factual mistakes. These two platitudes look to be in tension; what are we to do?

One way out would be to deny the second platitude. Some mistaken applications of concepts are clearly empirical mistakes. When the folk applied the concept *flat* to the earth, it wasn't that they didn't understand the concept *flat*; their mistake was geographical, not conceptual. Maybe all apparent conceptual mistakes are like this. Well, this can't be right, because sometimes the folk appear to make mistakes about artificial possible cases. In those cases, we tell them precisely what the facts are about the case, so we rule out this kind of error.

Still, someone might bite the bullet here and insist that the answers people give in these cases just can't be wrong. The difficulty for such a bullet-biter will be in explaining disagreements. People differ as to what the rational decision in Newcomb Problem cases is (does everyone know this problem?), or the morally right action in runaway trolley examples (does everyone know these?). Possibly we could say what Jackson (and Stich) say and argue that this is evidence that we have different concepts. But as Slote points out, this gets the phenomenology of the disputes completely wrong. We think there is a serious dispute here, and a good theory of concepts should explain this.

In other words, I take it to be beyond question that sometimes in disputes like those concerning trolleys or Newcomblike cases, we (a) have a common concept and (b) all understand the facts of the case. So someone (if not everyone) must be mistakenly applying the concept to a situation which is perfectly understood. This is what I am calling a conceptual error, as opposed to empirical errors like thinking the earth is flat.

The hard question, then, is how is the possibility of conceptual error consistent with the platitude that meanings are conventional. One way is to say that those in error have misunderstood the convention. But this would mean that we could settle Newcomb's problem and the runaway trolley problem by opinion poll. And this just seems mistaken. I know for a fact that the vast majority of people who have heard of Newcomb's problem are one-boxers. And I still think they are wrong. And this doesn't look like I'm embracing a contradiction, even once we've accepted the conventionality of meaning. Put another way, it seems in these conceptual disputes, the minority *might* be right some of the time, so the error cannot be simply an error as to the nature of the convention.

Simple Answer: The way mistakes always get made

Jackson gives two answers to this question, one rather superficial, the other subtle and interesting. The superficial answer is that the folk make mistakes for the reason mistakes always get made. People are inattentive, they are sloppy, they say the first thing that comes into their heads, and so on. The intuitions which are constitutive of meaning are those which are produced under ideal circumstances, so the paradox goes away.

This can't be the complete answer. Are one-boxers about Newcomb's problem really just confused, or inattentive, or whatever. We need to say more, and more is said.

And at least one of the specifics Jackson mentions seems wrong. He says we can discount intuitions that *a* is an *F* if they are produced because someone thinks (correctly) that *a* is a *G* and has the intuition that all *G*s are *F*s. In other words, we can discount general intuitions in favour of specific ones. But this just seems mistaken. If I think switching the trolley is right, because I think it will maximise expected utility, and I think maximising expected utility is right, why should this intuition be discarded. As Jack Smart says, sometimes we should test our specific intuitions against our general ones, and not our general ones against our specific ones.

Subtle Answer: The Theory of Important Criteria

This answer is really nice. The broad outlines are in Slote's 1966 paper, and there's some further development in the first two papers in **Lewis**. Call any old set of possible cases a *property*, meaning this in a loose sense. Now every concept picks out a property, though as we saw above it is possible that some properties are picked out by more than one concept. The paradox of mistakes can be easily stated using properties: which property is picked out by a conceptual term is determined by our intuitions about the cases to which that term applies, but sometimes our intuitions about whether a case is an element of this property are mistaken. And the paradox can be equally easily solved: only some properties are eligible to be the referent of conceptual terms.

To show that this solves the paradox, we have to show how it makes each of the conflicting platitudes true. First, to show that it allows for mistakes. Let *S* be the set of (possible) cases I think is covered by some term *t*. If *S* is not eligible to be the referent of *t*, then at least one of my intuitions must be wrong. Whatever is the referent of *t*, it will not allow that all my intuitions about the applicability of *t* are right. Second, to show that it makes meaning conventional. Well, just stipulate that the referent of *t* is the 'closest' eligible property to *S*. Had I had really different intuitions about the applicability of *t*, so the set of possible cases to which I thought it applied was *S´*, the nearest eligible referent may well be different.

So we have a two-step theory of meaning for conceptual terms. First, find the possible cases to to which the relevant agent(s) think(s) the term applies. (The relevant agent will be an individual if you have an individualistic theory of meaning, a community if you have a communitarian theory of meaning.) This picks out some property *S*, in the sense we've been using the term. Then find the nearest important, or natural, property to *S*, and that's the meaning. The first step guarantees the conventionality of meaning, the second step the possibility of mistakes.

One important fact to note is that at the second step, many different sets of intuitive cases covered get mapped onto the same referent. Even if one person thinks that *t* applies to cases in S_1, and another thinks it applies to cases in S_2, where $S_1 \neq S_2$, it may be that the most natural property near each of S_1 and S_2 is S_3. So each person will be referring to the same property by *t*. This makes it the case, as we wanted, that the dispute is not a verbal dispute. In an important sense, the two disputants mean the same thing by *t*, but they have a substantive disagreement about the cases in question.

Important Properties

So far I've said very little about which properties are important, or natural. Since this idea plays such an important part in the solution, this is something which should be fixed. Unfortunately, I have no idea how to fix it, and what is said in the literature is of little help. So I just want to discuss one question which I suppose we should answer before we can do much more. Is the importance, or naturalness as Lewis puts it, of different properties *a priori* or *a posteriori*? And if it's a different question, is it a matter of necessity or contingency?

Lewis seems to think that it's *a priori* which properties are natural, though this comes out much more clearly in the discussion in *On the Plurality of Worlds*, chapter 1, than in the papers in **Lewis**. One reason for this (I think) is that naturalness is needed to solve the grue/bleen problem. How can it be rational to infer from *All emeralds have been green* to *All emeralds are green*, but not from *All emeralds have been grue* to *All emeralds are grue*? Well, perhaps the rules of induction only apply to reasonably natural properties, like green, and not to unnatural properties, like grue. But the solution to the problem of induction must be *a priori*. (Since we need to solve the problem of induction to discover **anything** *a posteriori*, the solution to it had better not be *a posteriori*.) Hence the naturalness of properties must be *a priori*. Also, naturalness is used to solve the problem of when two objects, in different possible worlds, are duplicates. If naturalness differed between worlds, we would have the awkward possibility that *a* is a duplicate of *b* while *b* is not a duplicate of *a*.

Slote, on the other hand, seems to think that naturalness is *a posteriori*. Think of the use of naturalness to solve the problem of whether whales are fish. Even if the folk once classified whales as fish (which I suppose they did) they were wrong because the most important property shared by most of the things which the folk classify as fish is a property which whales lack. This all seems plausible enough, but is it really *a priori* that having gills is a more important property than living in water? I know why it is more important in this world, because the neatest biological theory of this world makes more use of the property of having gills than the property of living in water. But I don't know why it is more important in all worlds. Similarly, when Lewis wants to give examples of perfectly natural properties, he lists properties like having unit positive charge. This is obviously very important for this-worldly physics, but mightn't it be kind of gruesome in other worlds, with different kinds of laws?

I wish I had a solution to this problem, but I don't. As Lewis points out in "New Work" (Chapter 1 in **Lewis**), the need for a distinction into the more and less natural properties is essential to so much of modern philosophy that it really is indispensable. But its indispensability is about all we know about the distinction.

Digression on Network Theories of Meaning

In "Metaphysics by Possible Cases", Jackson gives a different, and I think inferior, solution to the problem of how to reconcile the conventionality of meaning with the possibility of error. (Or, equivalently, the problem of how we can have different conceptual intuitions without meaning different things by our words.) The solution, roughly, is that the meaning of a word is identified by its place in a network of inferential relations. On every theory, to do a conceptual analysis of *t*, we first assemble something from the folk beliefs about *t*. On this picture, what we assemble is "the folk theory thought of as a network of principles teasing out the connections between concepts, which would typically include the circumstances in which the various concepts are instantiated, and what characteristically follows from the instantiation of the concepts." So "if it is free action we are seeking to analyse, [the network] will be the mass of connections between free action, moral responsibility, causal explanations of various kinds, the justifiability of punishment, personal identity and so on, along with a catalogue of those cases most obviously judged to be of free action."

This picture of meaning has some obvious affinities with Quine's metaphor of the 'web of belief', but as a theory of meaning it was most extensively promoted in Ned Block's "Advertisement for a Semantics for Psychology," from 1986. Jackson likes it because it explains how people can understand a concept without having opinions (or having false opinions) on the philosophical disputes about that problem. His theory (at least in this paper) is that to understand a concept you have to know its place in the network, and this is knowable *a priori*. In general, however, it is *a posteriori*, which concept fills that place in the network in the actual world. So to take a famous example, to grasp the concept of water, you have to know the *a priori* facts that water is typically clear and drinkable, it fills the oceans and rivers and so on. To know what stuff really is water, you have to know some *a posteriori* facts, that the stuff which has these properties is H_2O. Jackson thinks that to grasp the concept of, say, identity through time, you have to know some platitudes connecting inter-temporal identity with moral responsibility and other concepts. That is, you have to know the function role the identity through time property plays. To know which property plays that role, you have to know some *a posteriori* facts, facts which Jackson think support a four-dimensionalist account.

So much for the benefits. This theory also has some problems. (I learned most of these from Fodor's book *Concepts*, but I suppose they are also mentioned other places in the literature.) One big problem is identifying when people share the same concept. If I believe that all snakes are poisonous, but you don't, this doesn't mean we don't have a common concept of snakes. It might just be that I have a false belief about snakes, but you don't. Nevertheless, *snake* will occupy a different spot in my inferential network to yours, because I will infer from *That's a snake* to *That's poisonous*, and you won't. So we have to identify which inferences are meaning determining, and which are not. (It was fashionable, I think especially during the 1970s, to claim to believe that whenever we learned

something new about a subject, the meaning of words referring to that subject changed, at least for us. But we all know how bad fashions were in the 1970s, and this one is sadly typical.)

One possibility is to say that only the analytic, or *a priori* inferences are meaning determining. (Fodor has the quaint idea that the is no analytic/synthetic distinction, so this defence is bound to fail.) My false belief is *a posteriori*; perhaps if I believed it to be *a priori* that snakes are poisonous, it would be plausible to say I meant something different by 'snake' to you. And I think this is Jackson's position in "Metaphysics by Possible Cases." But it doesn't seem that plausible. It seems that the dispute between compatibilists and liberationists about free will is not just a debate about the meaning of words; they both mean the same thing by 'free will', but differ as to how the world would have to be for people to have free will. As Jackson puts it, they share inferences about the role of free will in discussions of moral responsibility, personal identity and so on. But liberationists will hold, and hold that it is *a priori*, that we can infer from *Jill acted freely* to *Jill could have done otherwise*. And compatibilists will hold that this is not *a priori*. (They can, and should, allow that the sentence *Any free actions are actions such that the agent could have done otherwise* is *a posteriori*; but they must deny that the liberationist inference is *a priori* valid.) So it seems by Jackson's lights that there is no common concept here.

Jackson may flat out deny some of the premises I have used. He might insist that the liberationist need not think the inference from *Jill acted freely* to *Jill could have done otherwise* is *a priori* valid. I haven't seen any particularly *a posteriori* arguments for liberationism, but maybe some exist. And he has a plausible argument that debates over personal identity should be settled by *a posteriori* means. If we have Cartesian souls, clearly an *a posteriori* matter, then all the debates between psychological and physical identity theorists are moot, identity goes by preservation of soul. But in worlds where there are no souls, there can still be personal identity through time. So the correct theory of personal identity must be different in different worlds, and hence must be *a posteriori*. Jackson thinks the same might be true of the right theory of laws of nature, and of non-personal persistence through time, and perhaps it is also true for free will. The schema, remember, is that what is *a priori* is the functional role the concept plays, what is *a posteriori* is what actually plays that functional role.

Still, this doesn't seem convincing for all cases. Go back to the trolley cases, and presume for the sake of the argument that consequentialists think it is right to switch the car, and Kantians think that it is wrong. And, we are presuming, this doesn't show that they mean different things by 'right' and 'wrong'. Settling ethical disputes is not that easy. Let *L* be the long story setting up the case. The consequentialist will endorse the following inference: *You are in situation L* so *You should switch the car*, and the Kantian will not endorse it. Now, could that inference be *a posteriori*? What is it about the world which makes it right to switch the cars when in situation *L* in this world, but not in situation *L* in other worlds? It just doesn't seem that there could be anything. (My argument is a little weak at this point; intuitions that nothing could play a certain role are fairly defeasible, they can be defeated by a quick demonstration of something playing that role.) So it seems here we have meaning identity despite differences over the *a priori* inferences. Hence the network theory of meaning (and of analysis) must be wrong.