

Week Eleven: Objections to Jackson

1. The Objection From Linguistic Ignorance

One of the benefits of the 2D framework we looked at last week was that it explained how we could understand a sentence without knowing which proposition it expressed. And we could do this even if we give an account of understanding which is closely tied to the possible worlds semantics we use to analyse propositions. Really this can be done very easily, without appeal to any high-flying Kripkean cases. In “Analytic Metaphysics” Jackson discusses a very simple case of it. I can understand an utterance of “I have a beard” without knowing which proposition it expresses. I know how the proposition is generated from context plus meaning, if X is the speaker then the sentence expresses the proposition X has a beard. And that is enough for understanding. But if I don’t know who said the sentence, so I don’t know who X is, I don’t know which proposition is expressed by that utterance.

This case may seem quite different to the cases concerning water, heat, gold and the metre rod which we discussed last time. On Jackson’s framework the cases are quite alike. The right analysis of ‘gold’ shows that it contains a hidden indexical element, the word ‘actually’. This is an indexical, like ‘I’, because it refers to the world of utterance. One may understand it because you understand the rule behind the indexical, without knowing what it refers to. (Of course there is a sense in which you know the referent of a use of ‘actually’, it refers to this world. In just the same sense you know the referent of any use of ‘I’, it refers to the user. This clearly isn’t the relevant sense of knowing the referent. We will return to this point when discussing Yablo’s objections.) This explains how we can understand the sentence ‘Gold has atomic number 79’ without knowing it is necessary; we know the function from context to proposition expressed, and thus we know that in some contexts it expresses the necessary proposition, but we don’t know that we are in such a context.

One kind of linguistic ignorance is this indexical kind. It is unsurprising that we don’t know the referent of every indexical we understand. But it isn’t obvious that all ignorance is like this, particularly if we adopt Jackson’s approach to analysis. So consider the following passage:

Our account sees conceptual analysis as an empirical matter in the following sense. It is an empirical fact that we use a certain term for the kinds of situations and particular that we do in fact use it for, and the conclusions we come to on the subject are fallible... We also noted that conceptual analysis in our sense is of a kind with [other] empirical investigations. The question we now face, accordingly, is: In what sense is conceptual analysis concerned with the *a priori*? (47)

The answer, a few pages later, is that conceptual analysis is concerned with A-intensions not C-intensions. To put the distinction in the matrix terms we’ve been using, imagine that at each point in the matrix we don’t just a truth-value for a sentence, we put a set of things which have some property. The C-intension of a term is

roughly the function from worlds to the set which occurs immediately below them on the top row. That is, the function from worlds to things fitting that term in that world. The A-intension of a term is roughly the function from worlds to the set which immediately below on the diagonal. It is the set of things which fit the term if that world is the actual world.

As we've seen, this is often a very useful distinction to draw. But how can it possibly be relevant here? After all, the whiff of empirical investigation is just as strong when analysing *square* as when analysing *water*. If this is a little non-constructive, suggesting something has gone wrong without saying just what, it is because I really don't understand the answer that Jackson gives to this problem. And I certainly don't understand what the difference between A-intensions and C-intensions has to do with the problem.

The only answer I can see involves a slightly different move. A sentence is *a priori* iff understanding the sentence is sufficient to know that it is true. This is different from being analytic because (a) we can use out knowledge of how indexical reference is fixed and (b) we can use mathematical and conceptual reasoning to work out that the sentence is true. So if empirical investigation is required solely to get an understanding of the sentence, this is compatible with the sentence being *a priori*. So sometimes *a priori* research does require opinion polls!

2. Yablo's Objections

The Yablo paper and the Block and Stalnaker paper both make a fairly distracting blunder in their presentations of Jackson's position. Both hold that he is still an opponent of physicalism. Now while he once was a fairly prominent opponent of physicalism, the two papers setting out the Mary argument are among the most famous anti-physicalist papers of the 1980's, he has fairly publicly retreated from this position. (And I am fairly confident he would have made this retreat known in any comments he made on work about his current position.) So in both cases there is some sloppiness in the scholarship. In neither case does it distract from the quality of the arguments, but it means we have to be careful to distinguish those arguments which are directed against real Jackson from those arguments which are directed against straw Jackson.

2.1. Textbook Kripkeanism

Yablo starts off with a nice presentation of what he calls Textbook Kripkeanism. As has been remarked in a few places, this is neither Kripke's view, nor a view frequently found in the textbooks (at least before **Jackson**), so the name may be a misnomer. But it seems to be an interesting position, and one which Jackson may fairly be said to hold.

I have been running so far with this description of the Kripke cases. What we didn't learn from Kripke was that something we once thought possible was really impossible. What we did learn was that what we once thought (or acted as if we thought) an appropriate description of a possibility is in fact not an appropriate description of any possibility. Yablo gives it a slightly different spin, though I think this is just a

superficial difference. When we thought that there was a possible world in which gold has atomic number 42, we were correctly perceiving the possibility of something, but it isn't quite what we thought. What was possible is that the sentence "Gold has atomic number 42" expressed a truth, what isn't possible is that what is expressed by "Gold has atomic number 42" is true.

Textbook Kripkeanism is the claim that this is the only way for intuitions about possibility to go astray. So whenever we think something is possible, but really it isn't, this is because of the distinction between the possibility that the sentence expresses a truth and the possibility that what the sentence expresses is true. As a consequence we get infallibilism about idealised conceptual intuitions, that is, about intuitions about what kinds of points are on the diagonal. And we get a quick test for telling us whether it is likely that this conceptual possibility might not be a metaphysical possibility. If we think *S* is possible, then it is conceptually possible. And if there are no rigid designators in *S*, then if it is conceptually possible it is metaphysically possible.

2.2. *Zombies and Physicalism*

It isn't too hard to see how Textbook Kripkeanism might be parlayed into an argument for dualism. First, it seems conceptually possible that there are zombies. By the infallibility postulate, what seems to be the case here is the case. (As Yablo puts it, there are no delusions, only hallucinations, and these are all of one kind.) But none of the terms in "There are zombies" is rigid, so it is metaphysically possible that there are zombies. But if this is possible then physicalism is false, so physicalism is false. So the error of scholarship which Yablo and Block and Stalnaker make is understandable, if regrettable.

The last step in this argument may be a little controversial. Yablo writes as if the physicalist claim is that all worlds are physical. This seems patently absurd. Is there really no possible world at all which contains Casper the friendly ghost? Are beliefs about monkeys going to heaven necessarily false? It is better to say, as Jackson, Lewis, Chalmers *et al* do that physicalism is a doctrine about *this* world, though of course we need to state it as an inter-world supervenience thesis to capture it precisely. The hard question then is, which worlds are in the range of the supervenience claim.

To make this a little more precise, consider the following kind of world. It is physically just like this world, but it contains some extra stuff, call it ectoplasm. This ectoplasm causes our physical duplicates in that world to be devoid of conscious experience, though they are also caused to talk incessantly about their conscious experiences, particularly in philosophy of mind classes or various dubious Internet chat rooms. Such a world seems possible, to me at least! We have here a physical duplicate of the world in which people don't have conscious experiences. If it is possible, is this a problem for physicalism?

Here's an argument for *no*. Let's assume that this world contains no ectoplasm. Since we are trying to construct a *reductio* against physicalism, this seems like a safe assumption. And Chalmers in fact thinks it is true – he is a property dualist, not a substance dualist, whatever that means. (Did you know the spelling program on Microsoft Word doesn't recognise 'dualist'? What an out-of-date idea it must be!) Extra premise:

the structure of the parts of modal space which contain ectoplasm is irrelevant to whether non-ectoplasmic worlds are physicalist. Just as the existence of Casper in some other world is compatible with physicalism, whatever he does in some other worlds is also compatible with physicalism, assuming we are ghost-free. So the existence of an ectoplasmic world like the one I just described is compatible with physicalism. Hence the existence of some zombie worlds is compatible with physicalism.

Here's an argument for *yes*. If we buy a broadly counterfactual analysis of causation, and we accept that the above story describes a world, we should believe that it is not physical stuff alone which causes our conscious experiences. Rather, it is the physical stuff plus the absence of ectoplasmic blockers which cause our conscious experiences. And a world where ectoplasm plays such a crucial role in the causal story of the world cannot be a world in which physicalism is true. I like the first argument, John Hawthorne likes something like the second, though undoubtedly I haven't presented his case fairly.

2.3. *Understanding and Knowing Which*

Some of Yablo's objections around page 6 and 7 can be difficult to follow. And I'm not sure he makes the right choice at every turn in the argument. But there may be an objection in here. One worry is that we may know which worlds are *p*-worlds, and it be true that all worlds are *p*-worlds, yet we don't know that all worlds are *p*-worlds. For example, we may know that *p* is true in all and only worlds in which water is H₂O, so in one sense we know which worlds are *p*-worlds. But if we don't know that all worlds are worlds in which water is H₂O, we don't know that all worlds are *p*-worlds.

There is much possibility for confusion here, so let's run through some less philosophically loaded examples. If the detective knows which suspect killed the butler, and the suspect who killed the butler is Scarlet, then she had better know that the suspect who killed the butler is Scarlet. On the other hand, if she knows which suspect killed the butler, and the suspect who killed the butler also killed the gardener, then she needn't know that the suspect who killed the butler killed the gardener. Sometimes we can substitute co-referring expressions into 'knows which' contexts, and sometimes we can't, and it isn't clear that Jackson always respects this distinction.

The story Yablo tells is all told in 'Kripkean' terms, so we might think that there is some kind of confusion between A-possibility and C-possibility at the core of his objection. But perhaps we can replace references to possible worlds with references to points in Jackson's matrix without loss of any argumentative force. So, for example, we can know that *p* is true in all worlds in which water is H₂O and not know that *p* is true in all worlds because we don't know that water is H₂O in all worlds. This is the kind of thing the matrix account can explain; we don't know which world is the actual world so we don't know which proposition that sentence expresses. There are cases which seem to be more problematic for Jackson. Say we know *p* is true in all worlds in which either there are infinitely many twin primes or gravity is an attractive force. Well, in one sense we know which worlds *p* is true in, we just said so, but in another sense we don't. Notice we don't even know whether *p* is true in all points of the matrix or not. If there are infinitely many twin primes it definitely

is. I presume the *gravity* concept is much messier, is it necessary that gravity is attractive? is it *a priori*? I don't know the answers to these questions, so I don't know whether p is true at all points, even though in a sense I know which worlds it is true in.

This looks like a problem Jackson can resolve by stipulation, and I think that is what Jackson should do. Here's the sense in which understanding requires knowing which worlds a proposition is true in. We know which worlds p is true in, in the relevant sense, iff for any pair of worlds x, y we can say whether p is true in x given that y is actual. We must, as it were, know the matrix. We can do this without knowing which worlds, in Yablo's sense, the proposition is true in because we don't know which world is actual.

But even this looks like it might be problematic. Let's go back to the arguments Jackson was looking at. This is meant to highlight the distinction between the *a posteriori* valid argument on the left and the *a priori* valid argument on the right. (The bracketed lines are not part of the argument.)

(1a) H₂O covers most of the earth

(1b) The complete physical story is S

(2a) (H₂O plays the water role)

(2b) (State X plays the belief role)

(3a) Water covers most of the earth

(3b) Brian believes that snow is white

The distinction between the arguments is meant to be that on the left, (1a) does not entail (2a) in any way, and certainly not *a priori*, so the argument is not *a priori* valid. But on the right, physicalists should think that (1b) does entail (2b), so the argument is *a priori* valid. This all looks dubious, and I think this is one of the points that Yablo is making. All the physicalist is committed to is that (1b) entails (2b), not that (1b) entails it *a priori*. If the inference from (1b) to (2b) is necessary *a priori*, then the inference from (1b) to (3b) is *a priori* valid, I suppose. But why should we think that?

The argument Jackson gives on page 83 is rather brief on this point. In fact it isn't clear that Jackson has spotted this. Yablo attempts to offer some kind of defence, but it is a bit ambiguous. The thought is that the only way for an argument to be *a posteriori* valid is for it to be based on contextual sensitivities. And since neither the full physical description nor the term 'the belief role' is context specific, this can't be an *a posteriori* valid argument. I think most of the other points Yablo makes on page 7 can be dealt with if we understand 'knowing which' in the way suggested above, but one methodological point remains. If one was antecedently tempted by the thought that the inference from (1b) to (3b) was *a posteriori* valid for non-Kripkean reasons, Jackson's argument will be completely ineffective.

2.4. *The Quick Argument for A Priori Entailment*

While on this part of Jackson's book, we should take a quick look at the little argument at the end of chapter three for *a priori* entailment of the psychological by the physical. The argument doesn't depend on any of the technicalities we have been looking at so far. The argument looks something like this:

-
-
- (1) The physical description of an organism with 1 cell *a priori* entails its complete description.
 - (2) If the physical description of an organism with n cells *a priori* entails its complete description, the physical description of an organism with $n+1$ cells *a priori* entails its complete description.
 - (3) So the physical description of any organism *a priori* entails its complete description.

Quick pop quiz: How many things can you see which are wrong with *this* argument?

3. Yablo on Conceivability

The arguments in section 13 of Yablo's paper seem very mysterious. He wants to reject any kind of infallibility of conceptual facilities, particularly the kind he associates with textbook Kripkeanism. Yablo seems to need a distinction which I don't think exists.

Question: Given what you now know, can you conceive of water not containing hydrogen? Well, there is a good sense in which I suppose the answer is *no*, so conceivability in some sense is constrained by empirical knowledge. But it isn't clear in what sense.

Question: Given what you now know, can you conceive of Bill Clinton not being the winner of the 1992 Presidential election? Given what I *know*, this is impossible. So this can't be the kind of conceivability that Yablo needs. And I'm not sure there is any other kind around.

Anyway, that's not the most fun part of the argument. Let's grant him a concept I don't think exists, just to be polite. The argument against weak conceptualism still seems interestingly dubious. Here's roughly how it goes.

- (1) Which worlds we properly think are possible depends on what evidence we have.
- (2) So the genuinely possible worlds will be those that a reasonable person finds possible given all the evidence in the world.
- (3) But there are different reasonable responses to any evidence whatsoever (this is where the name-calling about Carnap enters the argument).
- (4) So there is no fact of the matter about which worlds are really possible.

I suppose (4) is the *reductio*. But the argument has a lot of hurdles to overcome before it ever gets there. There seem to be three possible objections to (3). First, we could reject the epistemological pluralism which underlies it. Secondly, we could reject epistemological pluralism about this type of *conclusion*. Thirdly, we could reject epistemological pluralism about this type of *evidence*. I'll deal with these in order.

Let 'evidence' be shorthand for whatever is relevant to the justification of belief. So if you are an externalist about justification, if for example you think being justified has something to do with reliability, 'evidence' will include facts about the world and not just about the thinker. This seems highly dubious to me, but that's for another kind of seminar. So if my evidence is e the full statement of e includes everything relevant to the justificatory status of my beliefs: my surface irritations over time, my inferential processes, the

reliability of said mechanisms, and so on. Epistemological pluralism is the view that two people with the same evidence, in this sense, can have incompatible beliefs while both are reasonable. Yablo relies on pluralism as the motivation for (3).

There are more and less extreme kinds of pluralism. (These names are my stipulations, so don't take them too seriously.) Moderate pluralists say that X and Y with identical evidence may be such that X believes p and Y is agnostic. Extreme pluralists say that X and Y with identical evidence may be such that X believes p and Y believes $\neg p$. That's the fun version, so let's work with it. The argument for it just goes by looking at possible, and for that matter actual, cases. A fancy example depends on incompatible theories about the foundations of physical science. I prefer more simple-minded cases. Let e be the evidence we all have about the JFK assassination. I think the following three states are reasonable for someone with evidence e

- Believing the lone gunman theory
- Believing there was a conspiracy
- Being agnostic between these positions

(Did you see there was more evidence of a conspiracy in the weekend papers?!) If this is right then epistemological pluralism is true.

Two related objections. First, it seems that to reasonably believe p we should have sufficient reason to rule out $\neg p$. But *ex hypothesi* we here don't have sufficient reason to rule out the conspiracy theory, so we should not believe the lone gunman theory. Similarly we don't have sufficient reason to rule out the lone gunman theory, so we should not believe in the conspiracy. So in such a circumstance, agnosticism is the only reasonable position.

Secondly, if pluralism is right, something like this should be assertible: " p but it would be reasonable for me to believe $\neg p$ ". This seems very odd, possibly for just the reasons listed under the first objection. So pluralism isn't obviously coherent. Note that the same considerations apply to *belief to degree x* as they do to absolute belief, so maybe Carnap was right all along!

Even granting Yablo pluralism, which I think is true but not trivially so because of these objections, two problems remain. First, just because we are pluralists it doesn't follow automatically that we should be pluralists about beliefs about possibility. Yablo has something like an argument here. The kinds of belief which constrain beliefs about possibility are very ordinary; they are just beliefs about material constitution and the like. So if we were monists about beliefs about possibility, we'd have to be monists about beliefs about the constitution of natural kind terms, and that seems implausible. Maybe that response will fly.

It still isn't clear that the argument goes through, because it still isn't clear whether we should be pluralists when X and Y know all there is to know about the world. It is logically coherent to say that when X and Y have evidence e , and e is less than a full world description, X and Y may reasonably react differently, but when e is a full world description there is only one rational response. In fact that seems right to me. And if that's true, I don't see where Yablo's argument could be headed.

4. Block and Stalnaker

Again it is **SO** irritating to see Jackson labelled as an anti-physicalist. Block and Stalnaker cite, in the paper, a personal correspondence with Jackson, and I strongly suspect that Jackson would have pointed out in that correspondence that he no longer believed in physicalism. So this is perhaps worse than sloppy scholarship.

4.1. *The Analysis of Life*

Despite that, this seems to be a very interesting paper, and one that is already receiving much attention around these parts. Let's start with the discussion of 'life'. The underlying argument appears to be as follows:

- (1) We can't give an *a priori* reduction of facts about what is living to microphysics, and it looks unlikely that there is such a reduction to be found.
- (2) Given the current state of biology, it would be absurd to suppose that the existence of life poses a threat to physicalism.
- (3) So, physicalism does not require *a priori* reductions.

The argument for (1) seems to rely a lot on the moving van example. But it seems there at least Jackson does have things to say. Maybe 'alive' is a rigid designator of a rather unusual kind. Anything which is of a type with something which had most of the paradigm properties in the old days is alive! That seems like it might be extensionally correct, provided we gerrymander types properly.

More seriously, Jackson can pose the following dilemma. Either we can say something about why the van is not alive, or we cannot. If we can say something, we can put it into the analysis. If we can't, then the proper course of action is to give up the claim that the van is not alive. It isn't decent to insist that something really isn't an *F* if it is like things you acknowledge to be *F*s in every salient way. It really would be very surprising if we could disprove the common-sense functionalist theory of life by appeal to common-sense.

4.2. *Ghost Water*

Block and Stalnaker suggest that the entry by entailment thesis may be in trouble because of examples like their ghost water example. Imagine a possible world where there are two kinds of watery substances, one of which is physical, and the other which is non-physical. Now the microphysical story will not entail that the physical one is the only watery substance, because it will not entail that the ghost water does not exist. For that entailment to go through, we must tell the physical story plus a *that's all* clause.

Jackson does this, so it is hard to see what the point of this section really is. It might prove something, but it isn't exactly what Block and Stalnaker want. Maybe it proves that there is no uniqueness condition in the platitudes about 'water'. After all, there is a possible world in which there is physical water and ghost water, and in that world, both things are water. So this is even less of a problem for Jackson than we may have thought.

4.3. *Two-Dimensions Again*

How might we find that S is true in w considered as actual? Well, here's one way, we could go to w and ask them. Of course w is too far away for that really to be practical, but even if it were tenable, there are philosophical objections to the exercise. The sentence "water is watery" is meant to be true in every world considered as actual, but if we go to w , we may find that they use that sentence to mean that grass is blue, and as it turns out in that world, grass is not blue. So we must ask some different type of question about the world.

Already we are stretching up against the limits of the two-dimensional framework which we are using. The whole point of inserting a second axis was that the proposition expressed by certain words may vary depending on which world is actual. What better way could there be for words to change meaning other than by changing the language in which they are embedded. So on the most natural interpretation of the framework, we must conclude that the sentence "Water is watery" is true at all points on the diagonal.

This is the interpretation Block and Stalnaker urge in section 10 of their paper, and since Stalnaker invented the matrices, I suppose he should be listened to with some respect. But not too much respect. A formalism, after all, is just a formalism, and we should be free to interpret it any way that we like. This is the core of Chalmers's response to Block and Stalnaker's argument.

At first glance, that looks like a plausible move, but there remains an issue as to whether there is *any* other viable interpretations of the formalism. One serious worry is that to get the kind of interpretation Jackson and Chalmers need, we need to use some of the concepts they want to use the formalism to explain. So here's one kind of response. We don't just ask people in w whether they endorse "Water is watery", we ask them whether they endorse the sentence in their language which means the same thing as "Water is watery" in our language. But now we need a concept of 'same meaning' as *input* into the 2D apparatus. And I thought that the framework was meant to explain this concept of meaning. After all, meaning here isn't the 'possible worlds' sense of meaning, 'water' doesn't mean H_2O , but what other pre-theoretic sense do we have?

5. For Next Week

We will really be going to the material on moral realism. We will be focussing on the use of Ramsey sentences in the first part of chapter 6, with the intent of doing something which looks more familiar to an ethics tradition the following week. So the readings will be chapters 5 and 6 of Jackson, and the following Lewis papers, "How to Define Theoretical Terms" and "Psychophysical and Theoretical Identifications". The Lewis papers are literally on philosophy of mind, which links to what we have been discussing, but as Jackson points out the tools he develops can be very widely applied.