Week Ten: Two-Dimensional Modality

1. Two Objections

Our primary interest this week will be in two objections Jackson mentions which seem to threaten his program. Each of them is avoided by appeal to the two-dimensional framework we sketched last week. Before we go over that framework again, we will start by looking at the objections. For reasons that may become apparent shortly, we will look at them in reverse order. So first we'll look at this objection from Chapter 3, an objection which turns on the discovery of *a posteriori* necessities by Kripke and Putnam.

(1a)   Jackson's argument for the importance of conceptual analysis relies on the need to justify some modal supervenience theses in metaphysics, theses which can be restated as entailments.

(2a)   But as we learned from Kripke and Putnam, whether or not a modal supervenience thesis holds is *a posteriori*. Where the water is supervenes on where the H_2O is, but this is not knowable *a priori*.

(3a)   Hence at the point where Jackson thinks that conceptual analysis is needed, what is really needed is an *a posteriori* identity thesis.

Which of these steps would Jackson reject? Why would he reject it?

The second objection turns on the apparently *a posteriori* output of Jackson's approach to analysis.

(1b)   Conceptual analysis, whatever it is, is something which delivers *a priori* results

(2b)   To perform what Jackson calls 'conceptual analysis' we need to take opinion polls.

(3b)   Any investigation which **requires** taking opinion polls cannot deliver *a priori* results.

(4b)   Hence, what Jackson calls 'conceptual analysis' is not really conceptual analysis.

Which of these steps would Jackson reject? Why would he reject it?

I'll call the first of these the Objection from *A Posteriori* Necessities, and the second the Objection from Linguistic Ignorance. To understand what Jackson says about each of these arguments, we need to look at the 2D apparatus he employs. And to do that we need to indulge ourselves with a little history lesson, starting roughly with *Naming and Necessity*.

2.   *The 2D Approach to Kripke Cases*

Let's start with two standard Kripke cases, the metre rod in Paris and the atomic number of gold. First the metre rod. Say that we stipulate that something is of length one metre iff it is of the same length as stick *S* at time *t*, where *S* is the standard metre rod. (Kripke uses t_0, but the subscripts get annoying, even on a modern word-processor.) The following will not be true, one fears.

(5)   If stick *S* had been heated at time *t*, it would have been one metre long at time *t*.

Rather, something like (6) will be true.

(6)   If stick *S* had been heated at time *t*, it would have been more than one metre long at time *t*.

Let's assume, for the sake of argument, that we buy the following premises:

P1:   Kripke is right about these conditionals; (5) is false and (6) is true.

P2:   Lewis is right about the logic of subjunctives; "If A were true then B would be" is true just in case *B* is true in the nearest possible world in which A is true.

P3:   The possible worlds account of propositions is correct; the proposition *p* is the set of worlds in which *p* is true. As a corollary, *p* is necessary iff it is the set of all possible worlds.

I suppose I have accepted all of these at some stages, so I can't really complain about these assumptions. It follows that (a) there are possible worlds in which stick *S* is not a metre long at time *t* and hence (b) the proposition *Stick S is a metre long at time t* is not necessary.

One lesson you might draw from my paper on conditionals is that the second premise here is important, and not obviously correct. (David Chalmers independently came to similar conclusions to me about conditionals, and he makes quite a bit of this point.) If we think, *contra* Lewis and Kripke, that the way to tell what happens in other possible worlds is *via* so-called 'indicative' conditionals, rather than *via* these so-called 'subjunctive' conditionals, we don't get to (a). The idea Chalmers and I are pushing is that when we rephrase the Kripke conditionals as indicatives, the intuitive judgements about their truth values are reversed. Making this a little more concrete, intuitively (7) is odd but true and (8) is false.

(7)   If stick *S* was heated at time *t*, it was one metre long at time *t*.

(8)   If stick *S* was heated at time *t*, it was more than one metre long at time *t*.

Chalmers's argument here (and I do wish I had also thought of *this* step) is that if we replace P2 with a claim about indicatives, we might take the truth of (7) and like conditionals as evidence that it really is true that stick *S* is a metre long in all possible worlds. Indeed, we may take the intuitive plausibility of the claim that it is necessary that *S* is a

metre long (after all, it is just a matter of stipulation) as evidence that P2 is wrong and should be so replaced. So the manuscript Chalmers is working on is tentatively titled "The Tyranny of the Subjunctive." Anyway, this is all a long digression to get to the main point that if we are obedient little neo-philosophers, and accept that P1, P2 and P3 are correct, we should accept that there are worlds in which *S* is not a metre long at *t*, and hence that it is not necessary that *S* is a metre long at *t*.

Despite this, we have no reason to accept that it is not *a priori* that stick *S* is a metre long at *t*. Given the stipulation, what could we possibly discover that would tell us that *S* is not a metre long at *t*. Not that it was heated; as we saw in (7) even if it was heated at *t*, it was a metre long at *t*. Maybe that it didn't have a length, because it had been smashed into smithereens, in which case possibly the term 'a metre long' is non-referring. So we really ought to qualify our *a priori* claim to "If *S* had a length at *t*, it was a metre long at *t*." But that isn't as immediately appealing, so let's ignore such pedantic qualifications. With that wonderfully rigorous move made, we can say we have discovered a contingent *a priori* truth: Stick *S* is a metre long at *t*.

Second standard Kripke case, after all that. Gold has atomic number 79, or so several disreputable sources tell me. The same sources tell me that it boils at 2970°C, and let's assume that's also true. Say that something is goldie iff it has many of the surface properties we have come to associate with gold, like being very dense and very hard to melt, let alone boil. It seems to be almost stipulative that gold is the goldie stuff. If you doubt this, think how the scientists discovered that gold really does have atomic number 79. They found some stuff, knew it was gold, tested it, and found it had atomic number 79. How did they know the stuff was gold before they ran the tests? Well, because it was goldie, and it is stipulative that goldie stuff is gold. Quick obedience test: what is the truth value for the following conditionals?

(9)   If there were stuff which was goldie, but didn't have atomic number 79, it would be gold.

(10)   If all the stuff which is goldie doesn't have atomic number 79, the goldie stuff isn't gold.

The answers here are taken as evidence that there are necessary *a posteriori* truths. The reasoning is relatively simple. From conditionals like (9), we learn that however the world were like, it would be the case that gold has atomic number 79. By P2, we conclude from that that gold has atomic number 79 in all possible worlds. So by P3, we conclude that gold necessarily has atomic number 79. But clearly it is *a posteriori* that gold has atomic number 79, this is a scientific discovery. So there are necessary *a posteriori* truths.

Here's one intuitive response to that data. There are various possible worlds which are conceptually possible. That is, they are worlds which can't be ruled out by *a priori* reasoning alone. In some of these, for example, gold has atomic number 42. We can't tell *a priori* that gold doesn't have atomic number 42, so such worlds are conceptually possible by definition. But only some of these conceptually possible worlds are metaphysically possible. In particular, only the worlds in which gold has atomic number 79 are really metaphysically possible. So there are two important types of possible worlds, the conceptually possible worlds and the metaphysically possible worlds, and whether something is necessary or contingent depends merely on what happens in the metaphysically possible worlds.

As it stands, this is a dreadful response to the data. We might explain the necessary *a posteriori* truths that way, but we certainly can't explain the contingent *a priori* truths so easily. For those, there must be worlds which are conceptually impossible but metaphysically possible. Whatever intuitive merit the 'two types of world' view may have had seems to have disappeared by this point, at least for me, but some brave souls fight on. So some people argue that there really are two distinct types of possible worlds, and neither is a special case of the other. We have the metaphysically possible worlds, the conceptually possible worlds, and rarely the twain shall meet. This world is metaphysically and conceptually possible, but it may well be the only such world.

The bulk of chapter 3 is a response to this kind of ontological extravagance. Jackson argues that rather than having two types of worlds, we really have one type of worlds, the possible worlds, under two types of descriptions. This isn't too hard to motivate intuitively. When we thought about the world where stuff with atomic number 42 (presumably Molybdenum) was goldie, we really were thinking about a possible world, both metaphysically and conceptually. What may have gone wrong pre-theoretically was that we thought this goldie stuff was gold. (Jackson thinks this didn't go wrong. Why?) What we learn from Kripke is some facts about how to describe this world that we all agree exists. Under one perfectly respectable disambiguation of English, the proper way to describe any possible world is to use 'gold' to refer to the stuff in it with atomic number 79.

So on Jackson's picture, here's how we get sentences coming out necessary *a posteriori*. For some terms, what they refer to (in all worlds) is determined by the way the world is. Given that the world is the way it actually is, the reference of these terms will be rigid. For example, given that gold actually has atomic number 79, 'gold' refers to stuff with atomic number 79 in all possible worlds. Hence the sentence 'Gold has atomic number 79' is true in all possible worlds, as required. But it can't be determined by first philosophy just what gold refers to. This is because, as we've said, what it refers to is determined by facts which are hidden to first philosophy. Hence it is *a posteriori*. This is a rather metalinguistic explanation; we get the results by going via facts about the *reference* of terms in the language. And this feature of the explanation may come back to haunt us rather soon. But for now let's just rest in the haunted house for a few minutes to catch our breath. Surely nothing can go wrong in that time!

We can give a similar style of explanation of the existence of contingent *a priori* truths. Whatever 'gold' refers to, it is the stuff which is goldie. We know this because we know that to find out anything about gold, we have to find goldie stuff and then examine it. But this project can't help but conclude that gold is goldie. So it is *a priori* that 'Gold is goldie'. But when we conclude this investigation, however it concludes, we will have discovered something about the nature of gold. And if we are good modern philosophers, we will concede that that stuff, which we agree deserves the name 'gold', may have had different surface properties. So that stuff need not have been goldie. Hence 'Gold is goldie' turns out to express a contingent truth. So there are contingent *a priori* truths, as required.

In sum then, the truth of some sentences in some worlds depends not just on how those worlds are, but on how this world is. To tell whether it is true in a world that my yacht is larger than your yacht actually is, we don't just have to know the size of my yacht in that world, but also the size of your yacht in this world. If an acid is whatever plays a certain functional role in this world, then whether there are acids in the New York City drinking water in *w* depends not just on facts about *w*, but on what plays a certain functional role in this world. Stalnaker gives us a nice way of presenting these facts. There are other approaches, for example in my "Indicatives and

Subjunctives" I borrow a strategy from Martin Davies and Lloyd Humberstone, but Stalnaker's is probably the easiest to learn.

Say S is true at a pair of worlds $<w, v>$ iff S is true at world w given that world v is actual. A sentence is *a priori* iff for all worlds w, S is true at $<w, w>$. And letting @ stand for this world, it is necessary iff for all worlds w, S is true at $<w, @>$. Stalnaker suggests a graphic representation of this. Set up an e by e matrix, where e is the number of possible worlds there are. Each cell corresponds to a pair of worlds. The matrix for a sentence has T at a cell iff that sentence is true at the world in the column, given the world in the row is actual. So say we have three worlds w_1, w_2 and w_3, with the following properties.

- In w_1 proton donors play the acid role, and there are proton donors but no proton takers in the NYC drinking water.
- In w_2 proton donors play the acid role, and there are no proton donors in the NYC drinking water, but there are proton takers.
- In w_3 proton takers play the acid role, and there are proton takers in the NYC drinking water, but there are proton donors.

The matrix for the sentence "There are acids in the NYC drinking water" looks like this.

	w_1	w_2	w_3
w_1	T	F	T
w_2	T	F	T
w_3	F	T	F

To get the hang of these, we should work through a few more examples. Since these are from last week's notes, I suppose everyone will already have worked them out! Assume there are just three worlds, @ the actual world, *tw* being Twin Earth as usually described, and *pw* being a world where half the watery stuff is H_2O, and half is XYZ. (On Twin Earth there is some H_2O and it is gooey and sticky.) Draw the matrix for each of the following sentences:

1. All water is H_2O.
2. All water is watery.
3. Water is a gooey, sticky substance.
4. The stuff called 'water' is a gooey, sticky substance.
5. The stuff actually called 'water' is a gooey, sticky substance.

Now for some more general questions:

- What will the matrix for a necessary true sentence look like? Why?
- What will the matrix for an *a priori* sentence look like? Why?
- What will the matrix for an analytically true sentence look like? Why?
- Is there a distinction, within the matrix, between being necessary and *a priori*, and being analytic?
- Give examples of sentences which are necessary and *a priori* but are not analytic.

Jackson mentions that there are two types of proposition which may be expressed by a sentence. In our terminology, the A-proposition is the set of worlds corresponding to T's on the diagonal. That is, the set of worlds such that if that world is actual, that sentence is true. These are sometimes referred to in the literature as diagonal propositions. The C-proposition is the set of worlds corresponding to T's on the top line. That is, the set of worlds where the sentence is true (given the fact that this is the actual world). These are sometimes referred to as horizontal propositions. A sentence is necessary iff the C-proposition it expresses is the set of all worlds; it is *a priori* iff the A-proposition it expresses is the set of all worlds.

With all that technical apparatus at our command, we can finally do a bit of philosophy. In this approach, we can get all the Kripke results, and we can get it using just one set of worlds. We do, to be sure, use those worlds twice over, once on each axis. But only one type of world, and so our general love of minimal ontology should drive us towards accepting this story over the story which posits different types of worlds.

3. How Jackson uses the 2D model to respond

I'll look first at the Objection from *A Posteriori* Necessities, because Jackson's response to this objection seems reasonably clear.

(1a)   Jackson's argument for the importance of conceptual analysis relies on the need to justify some modal supervenience theses in metaphysics, theses which can be restated as entailments.

(2a)   But as we learned from Kripke and Putnam, whether or not a modal supervenience thesis holds is *a posteriori*. Where the water is supervenes on where the H_2O is, but this is not knowable *a priori*.

(3a)   Hence at the point where Jackson thinks that conceptual analysis is needed, what is really needed is an *a posteriori* identity thesis.

Since (1a) is just a summary of what's gone on so far, it can't really be rejected. Our options are either than (2a) is false, or that (3a) doesn't really follow from (2a). Jackson uses the 2D apparatus to argue that (2a) is false. What Jackson needs to show here is that there is a disanalogy between the case mentioned in (2a), the supervenience of the water facts on the H_2O facts, and the case in which he is interested, the supervenience of all the psychological facts on the physical facts.

If the two types of world theory were correct, this move would be impossible. Here's how the argument would play out. The Kripke cases show us that it is *a posteriori* whether or not there are a pair of metaphysically possible worlds in which the physical facts are alike (and there are only physical facts) but the psychological facts are different. Even if such worlds are conceptually impossible, the case of contingent *a priori* truths show us that there are metaphysically possible worlds which are not conceptually possible. So the most conceptual analysis could tell us is that the supervenience thesis holds for all conceptually possible worlds, but what is needed is proof that it holds in all metaphysically possible worlds, and we need to do some real *a posteriori* research to show that.

Given the premise, that there are these two distinct types of worlds, the argument looks fairly compelling. If there is just one kind of world, but two or more ways of describing it, Jackson has more room to manoeuvre. Reflect again on why it is that "Water is H_2O" is necessary but *a posteriori*. Some sentences express different propositions in different worlds, indeed some express different propositions in different contexts. In this context, the proposition that sentence expresses is the necessary one, but we can't know *a priori* that we are in such a context. If we had enough information to know which context we are in, we would know that proposition was necessary.

The vital move now is made in the first paragraph on page 83. The reason we can't infer *a priori* from "Here's some H_2O" to "Here's some water" is that the extra proposition we need for the entailment, that water is H_2O, is not available *a priori*. That is only given by extra empirical research. The situation is quite different in trying to get from "The complete physical story of the world is thus-and-so" to "The complete psychological story of the world is blah". For this to be *a priori* we may need, in some sense, some context fixing sentences that play the role of "Water is H_2O". But if we're physicalists, we should think that whatever fixes context, it had better be physical. So we can infer the needed context-fixers from "The complete physical story of the world is thus-and-so". And by analogy when we know the context-fixers, the sentences like "Water is H_2O", the inferences in question will be *a priori*.

So there is a disanalogy between what we might call 'partial' supervenience theses and 'global' supervenience theses. On the first kind, the inference from the subvenient base to facts about the supervenient properties may be *a posteriori*. The reason is that the facts in the subvenient base may not be sufficient to ensure that the supervenience relationship holds. Just knowing all about H_2O will probably not tell you that H_2O is water. At least it won't tell you that it is the only kind of water. So even knowing all about the H_2O will not suffice to prove that there is no water in a particular area. But when we are discussing global supervenience theses, like the supervenience of everything on the physical, the subvenient base does include sufficient facts to ensure that the supervenience thesis holds, so the inference in question may be *a priori*.

So quickly reviewing where we've been so far. Grant, for the sake of the argument, that physicalism is committed to there being certain entailments between physical facts and psychological facts (and economic facts, and semantic facts, and so on). This still doesn't show that there is an important role for conceptual analysis, because to show that we need to show that the entailments in question are *a priori*. As we learned from Kripke, some entailments, like "This is water" so "This is H_2O" are not *a priori*, and we have no reason to think that the Jackson entailments will be any different. Indeed, the paradigm instances of *a priori* entailments are where the entailment holds because of syntactic facts about the relata, as in *modus ponens*, so we may have some reason for thinking that the entailments here will be *a posteriori*. Jackson's first move is to use the 2D apparatus to show that necessary *a posteriori* sentences, or equivalently *a posteriori* entailments, only arise in a specific kind of case. The case is where we don't know which C-proposition is expressed by a particular sentence, because the C-proposition expressed depends on facts not settled by the truth of that sentence. The second point to note is that, if physicalism is true, the entailment of the psychological story by the physical story cannot be like this, because if physicalism is true the truth of the physical story will settle which C-proposition is expressed by every sentence.

With all that behind us, we can return to the objection from linguistic ignorance. There are two kinds of ignorance that might worry us. The first is that we know the A-proposition expressed by a sentence, but because we

don't know which context we are in, we don't know the C-proposition it expresses. That we can solve by reduction to the previous case. The second is that we may not even know the A-proposition a sentence expresses. The most obvious reading of the paragraph on page 47 setting out the objection suggests this is the problem Jackson takes himself to be addressing. But if that is the problem, it is hard to know just what Jackson has to say about it. The answer he gives all turns on the differences between A and C-propositions, just as we've been discussing so far. But it is hard to extract a solution to the puzzle from that material.

To see this, consider a different example. Say we are trying to determine, by conceptual analysis, whether it is true that all squares have four sides. Ideally, we would like to say it is *a priori* that squares have four sides, but it isn't at all clear Jackson can do this. The problem, as Jackson originally notes, is that to elucidate the concept *square*, we have to do just what cognitive psychologists do when they work out the young child's concept *faster than*, or what political scientists do when they work out the American's concept *socialist*. Since these are clearly empirical investigations, it seems that working out the concept of *square* will be an empirical investigation, but since we need to do this to work out whether all squares have four sides, this will not be *a priori*.

Jackson's solution turns on the fact that some predicates have different A-intensions and C-intensions. The A-intension of a predicate is a function from possible worlds to the things which would satisfy that predicate were that world actual. So the A-intension of 'water' is the watery stuff at each world. The C-intension of a predicate is a function from possible worlds to things satisfying that predicate at that world. So the C-intension of 'water' is the H_2O at each world. This is a very pretty theoretical apparatus, but it seems completely irrelevant to the problem at hand. For the problem can arise, as we just saw, with predicates like 'square' which apparently have identical A-intensions and C-intensions.

There is a way out here for Jackson, but it isn't very honourable. A sentence is *a priori* iff understanding that sentence is sufficient to know that it is true. This is roughly how Ayer uses *a priori* in *Language, Truth and Logic*, and we all know everything in that book is true. So a sentence expresses an *a priori* truth iff once we know the A-intensions of all the words in it, we know that it is true. There appears to be empirical work involved because empirical work is needed sometimes to know A-intensions. So on this picture Jackson rejects (3b); some investigations, namely investigations into A-intensions, which require taking opinion polls nevertheless lead to *a priori* knowledge.

4.   Stalnaker's Objections

This isn't particularly satisfactory as it stands. If Jackson needs to reject (3b), why not take that rejection as a *reductio* of his theory? Well, the *a priori* is a bit of a mystery I suppose, so we shouldn't be too surprised to learn it has new hidden features. There may be a serious problem lurking here though.

How do we tell whether a world in the A-proposition expressed by a sentence *S*? Here's two ways not to answer this question. First way: ask whether *S* is true at *A*. That is the right question to ask when discussing C-propositions, but the wrong question to ask when discussing A-propositions. Second way: Ask whether the speakers there would endorse *S*. That will tell us that the A-proposition expressed by the sentence "Water is a liquid" is not

the set of all worlds. For at many worlds, the sentence "Water is a liquid" is used to express the proposition that Republicans lost the last Presidential election, and at many of those worlds this is false.

This vague rhetorical question can be turned into a serious worry. First, according to the official formulation of 2D modality, the way to determine whether a *sentence*, and it is now important that we are talking about sentences, is true at a world considered as actual (i.e. a point on the diagonal) is to ask whether that sentence would be endorsed there. If this looks like confusing use and mention, or making some other equally ghastly mistake, recall that the whole point of the framework was to give a formal way of expressing the fact that sentences express different propositions in different contexts, or worlds. So the first worry is that there may not be any way at all of making sense of the 2D apparatus in the way Jackson needs. The problem is that we need the following to be true:

- In all worlds considered as actual, "Water is a liquid" is true
- In some worlds considered as actual, the sentence "Water is a liquid" expresses a falsehood
- Which proposition is expressed at a world considered as actual is determined by the nature of that world

Jackson may have an 'out' here. Maybe we can say that whether a sentence is true in a world considered as actual depends on whether a sentence expressed there *with the same meaning* as it would have if expressed here, is true there. That's a bit messy, but I'm sure you get the idea. The problem now is that we have to find a notion of 'same meaning' which gives us the desired result. It can't be the Kripkean concept of same meaning, because then 'water' and 'H$_2$O' will have the same meaning. More importantly, it can't be any concept which we explain by using the 2D apparatus. For we need this concept of same meaning in order to use that apparatus the way Jackson wants us to use it.

Chalmers makes the point that we can't take the first problem at face value. The fact that there is one formal interpretation of the apparatus according to which the sentence "Water is a liquid" is not true at all worlds considered as actual shows neither that the sentence is not *a priori*, nor that there is no other interpretation of the formal structure. The problem which remains is that it seems incumbent on users of this apparatus to say just what this other interpretation is, and to say this in a way which only uses concepts we can explain without this apparatus.

5. *For Next Week*

At this stage we have a choice. We can either spend more time on this material (more than that short sketch), call this Option A. Or we can move on to the material on moral realism, call this option B.

If you (collectively) choose Option A, the reading for next week will be the papers already in the filing cabinet that I recommended people skim for this week, with the exception of my paper on conditionals. That would still be background reading.

If you choose Option B, the reading will be chapters 5 and 6 of Jackson, and two Lewis papers. One, "Psychophysical and Theoretical Identifications" is in **Lewis**. The other "How to Define Theoretical Terms" will appear I suppose in the filing cabinet.

If we do Option A next week, we will still do Option B the week after.