Coulda, Woulda, Shoulda

STEPHEN YABLO

1 Terminology

A main theme of Saul Kripke's *Naming and Necessity* (1980) is that metaphysical necessity is one thing; apriority, analyticity, and epistemic/semantic/conceptual necessity are another. Or rather, they are others, for although the relations among these latter notions are not fully analyzed, it does emerge that they are not the *same* notion.

'Apriority' and 'analyticity' are for Kripke nontechnical terms. They stand in the usual rough way for knowability without appeal to experience, and truth in virtue of meaning. Examples of apriority are given that it is hoped the reader will find plausible. And a schematic element is noted in the notion of knowability without experience; how far beyond our own actual cognitive powers are we allowed to idealize? Beyond that, not a whole lot is said.

Analyticity, though, does come in for further explanation. The phrase 'true in virtue of meaning' is open to different interpretations, Kripke says, depending on whether we are talking about 'meaning in the strict sense' or meaning in the

looser sense given by a term's associated reference-fixing description. A sentence like 'Hesperus is visible in the evening' comes out loosely analytic but not strictly so, since the meaning proper of 'Hesperus' is exhausted by its standing for Venus.

Kripke stipulates that 'analytic' as he uses the term expresses *strict* analyticity, and he takes this to have the consequence that analytic truths in his sense are metaphysically necessary truths ('an analytic truth is one which depends on *meanings* in the strict sense and therefore is necessary' (1980: 122 n. 63)). He notes, however, that one might equally let the word express loose analyticity, and that on that definition 'some analytic truths are contingent' (ibid.).

Given the care Kripke takes in distinguishing the kind of analyticity that entails metaphysical necessity from the kind that doesn't, one might have expected him to draw a similar distinction on the side of apriority: there would be an apriority-entailing kind of analyticity and a kind that can be had by non-a priori statements. 'Hesperus is Phosphorus' is not a priori, but since its meaning is a proposition of the form $x = x$, and any proposition of that form is true, it could be considered true in virtue of meaning. I am not endorsing this particular example, just pointing out a move that could have been made.

Kripke seems, however, to take it for granted that analytic truths will be a priori knowable. In his characterization of loose analyticity he speaks, not of statements whose truth is guaranteed by reference-fixing descriptions, but ones whose '*a priori* truth is *known* via the fixing of a reference' (1980: 122 n. 63, italics added). A non-Kripkean line on the apriority of analytic statements will be elaborated below.

I said that apriority and analyticity were for Kripke (relatively) 'ordinary' notions. There are intimations in *Naming and Necessity* of a corresponding technical notion: a notion that explicates apriority/analyticity as metaphysical necessity explicates our idea of that which could not be otherwise. This technical notion—potentially a partner in full standing to metaphysical necessity—needs a name of its own. What should the name be?

'Epistemic necessity' is best avoided because, as Kripke says, to call S epistemically possible sounds like a way of saying that it is true (or possible) for all one currently knows.[1] A notion explicating apriority/analyticity should not be so sensitive to the extent of current knowledge. One doesn't know how to prove Goldbach's conjecture today, but one might tomorrow; it would then turn out to have been necessary (in the partner sense) all along.

'Semantic necessity' too is liable to mislead, since for some people, Kripke included, 'Hesperus' and 'Phosphorus' are semantically just alike, yet it is possible in the partner sense that Hesperus ≠ Phosphorus. As Kripke says, one is inclined to think that it could have turned out either way.

If a name is to be given, then, to the *non*metaphysical modality that features in *Naming and Necessity*, 'conceptual' is probably the least bad. It is true that Kripke doesn't use the word 'conceptual' and doesn't talk much about concepts. But his nonmetaphysical necessities do have their truth guaranteed by the way we have represented things to ourselves; and we can think of 'concept' as just evoking the relevant level of representation. Conceptual necessity will then be the technical or semi-technical notion that Kripke runs alongside, and to some extent pits against, metaphysical necessity.

2   Conceptual Necessity

An enormous amount has been done with the metaphysical/conceptual distinction. Yet, and I think this is agreed by everyone, the distinction remains not terribly well understood. One reason it is not well understood is that the conceptual side of the distinction didn't receive at Kripke's hands the same sort of development as the metaphysical side.

This might have been intentional on Kripke's part. He might have thought the conceptual notion to be irremediably obscure, but important to mention lest it obscure our view of metaphysical necessity. Certainly this is the attitude that many take about the conceptual notion today. It could be argued that much of the contemporary skepticism about narrow content is at the same time skepticism about conceptual possibility. Narrow content, if it existed, would give sense to conceptual possibility: holding its narrow content fixed, S could have expressed a truth. If one rejects narrow content, one needs a different explanation, and none comes to mind. Going in the other direction, one might try to define S's narrow content as the set of worlds w whose obtaining conceptually necessitates that S. Lewis remarks somewhere that whoever claims not to understand something will take care not to understand anything else whereby it might be explained. If you don't understand narrow content, you will take care not to understand conceptual possibility either.

But, although many people have doubts about conceptual possibility, a number of *other* people are entirely gung ho about it. Some even treat it (and narrow content) as more, or anyway no less, fundamental than metaphysical possibility (and broad content). An example is David Chalmers. He calls S's narrow content its 'primary intension', and its broad content its 'secondary intension'.

[1] DeRose (1991) argues that this familiar condition is not enough. If contrary information is there for the taking, and/or possessed by relevant others, then S is not epistemically possible, even if it could be true for all I myself know.

One suspects that the order here is not accidental. And even if the suspicion is wrong, the primary intension is certainly a partner in full standing.

In this paper I try not to take sides between the skeptics and the believers. My topic is how conceptual possibility should be handled *supposing it is going to be handled at all*. If I do slip occasionally into the language of the believers, that is because I am trying to explore their system from the inside, in order to see what it is capable of, and whether it can be made to deliver the advertised kinds of results. (I should say that my own leanings are to the skeptical side, though I think the issue is far from settled.)

3 Initial Comparisons

Kripke's theory (or picture) of metaphysical modality is familiar enough. He says that it holds necessarily that S iff S is true in all possible worlds. The word 'in' is, however, misleading. It suggests that S (or an utterance thereof) is to be seen as *inhabiting* the world(s) *w* with respect to which it is evaluated. That is certainly not Kripke's intent. His view is better captured by saying that S (that well-known denizen of *our* world), to be necessary, should be true *of* all possible worlds. Every world should be such that S gives a correct description of it. Every world should be such that the way S describes things as being is a way that it in fact is.

Conceptual possibility too is explained with worlds. To be conceptually possible is to be in some appropriate sense true with respect to—or, for short, true *at—w* for at least one world *w*. But what is the appropriate sense? Everyone knows the examples that are supposed to bring out how conceptual modality is different. It is conceptually possible, but metaphysically impossible, for Hesperus to be distinct from Phosphorus. This is because 'Hesperus ≠ Phosphorus' is true at a world that it fails to be true of. The metaphysical/conceptual contrast thus hangs on the contrast between true-of-*w* as just discussed and the notion of true-at-*w* that we must now attempt to develop.

Here is the obvious first stab: S is true at *w* iff S as uttered in *w* is true of *w*. 'Hesperus ≠ Phosphorus' uttered here in the actual world means that Venus isn't Venus; uttered in *w*, it might mean that Venus isn't Mars. If, in *w*, Venus indeed *isn't* Mars, then 'Hesperus ≠ Phosphorus' is true at *w*. And so *w* testifies to the conceptual possibility of Hesperus not being Phosphorus.

Compare now an S that strikes us as *not* conceptually possible: for instance, 'Phosphorus ≠ Phosphorus'. Uttered in *w*, this means that Mars ≠ Mars. Since that is false of Mars, in *w* or anywhere else, *w* does not testify to the

conceptual possibility of Phosphorus not being Phosphorus. Unless there are worlds where uttering 'Phosphorus ≠ Phosphorus' is speaking the truth, that Phosphorus ≠ Phosphorus is not conceptually possible.

But, and here is where the trouble starts, there *are* worlds like that. For there are worlds in which 'Phosphorus ≠ Phosphorus' means something other than what it actually means (say, that Phosphorus is *identical* to Phosphorus) and in which the other thing is true. So it looks like we reach the wrong result. It should not make 'Phosphorus ≠ Phosphorus' conceptually possible that there are worlds in which '≠' expresses identity!

One remembers this sort of problem from Kripke's discussion, not of conceptual possibility, but metaphysical possibility. Let it be, he says, that *w* contains speakers (maybe our counterfactual selves) who understand S eccentrically from our point of view. That has no bearing on the issue of whether S is true of *w*:

when we speak of a counterfactual situation, we speak of it in English, even if it is part of the description of that counterfactual situation that we were all speaking [another language] . . . We say, . . . 'suppose we had been using English in a nonstandard way'. Then we are describing a possible world or counterfactual situation in which people, including ourselves, did speak in a certain way different from the way we speak. But still, in describing that world, we use *English* with *our* meanings and *our* references. (1980:77)

By 'tail', for example, the inhabitants of *w* might mean *wing*. If so, then assuming *w*'s horses resemble ours, they speak falsely when they say 'horses have tails'. That is irrelevant, Kripke says, to the metaphysical necessity issue. 'Horses have tails' is as true of *w* as of the actual world. This is crucial if statements are to come out with the right modal status. 'One doesn't say that "two plus two equals four" is contingent because people might have spoken a language in which "two plus two equals four" meant that seven is even' (1980:77).

How much of this still applies on the conceptual side? Worlds where 'Hesperus ≠ Phosphorus' means that Venus ≠ Mars *can* (as we saw) bear witness to the conceptual possibility of Hesperus not being Phosphorus. So in judging conceptual contingency, we *do* want to look at *w*-speakers who, in a broad sense, mean something different by S than we mean by it here.

But there are limits; we are not interested in *w*-speakers who by 'Hesperus ≠ Phosphorus' mean that Hesperus is identical to Phosphorus, or that it's snowing in Brooklyn. It thus becomes important to know in what ways the meaning of S in the mouths of *w*-speakers can differ from the meaning of S in our mouths, for the truth of S as uttered in *w* to be relevant to the conceptual possibility of S here. Something has got to be held fixed, but what?

4 Holding Fixed

First try: S has got to mean the *very same* in *w* as it means here.

This holds too much fixed. 'Hesperus' and 'Phosphorus' as they are used here both mean Venus, and '≠' expresses nonidentity. A counterfactual utterance of 'Hesperus ≠ Phosphorus' that respected these facts would have to mean that Venus ≠ Venus; and so the utterance would not be true. But then it will not come out conceptually possible that Hesperus ≠ Phosphorus, as it should.

Second try: Corresponding expressions should mean the same, *or* have their references fixed by the same or synonymous descriptions.

This is all right as far as it goes, but there is a problem of coverage. If a reference-fixing description is one that picks out the referent no matter what, then reference-fixing descriptions are hardly ever available. One doesn't know of any description guaranteed in advance to pick out the referent of 'Homer' or 'water'. So the second proposal reduces in most cases to the first, which we've seen to be inadequate.

A third approach puts conditions not on S in particular, but on *w* as whole: *w* bears on S's conceptual possibility if and only if it is an 'epistemic counterpart' of our world, in the sense of confronting the speaker with the same evidential situation as he confronts here. If *w* is an epistemic counterpart of actuality, then S's meaning can change only in ways that leave the evidential situation as is; that is what it takes for S's truth in *w* to bear witness to its conceptual possibility here.

A seeming advantage of the proposal is that it no longer attempts to specify the relevant aspects of meaning (the ones that are supposed to be held fixed) explicitly. The thought is that those aspects, whatever they are, are fixed *inter alia* by fixing the entire evidential situation. This is also the proposal's problem, though. Mixed in with the semantical material we want to hold fixed will be nonsemantic circumstances that should be allowed to vary. One doesn't want to hold fixed that there seems to be a lectern present, or there seeming to be a lectern present will be classified as conceptually necessary. That is clearly the wrong result. Appearances are conceptually contingent if anything is.

5 Subjunctives

The kind of necessity we are calling conceptual is left by Kripke in a precarious state. Judging conceptual necessity is judging whether S *as uttered in w* is true of *w*. This collapses into triviality unless certain aspects of S's meaning are held fixed. And it is unclear which aspects are intended.

Why do the same problems not arise for metaphysical necessity? The usual answer is that with metaphysical necessity, one needn't bring in a counterfactual utterance at all. One considers whether *our* utterance, saying (or meaning) just what it actually says (means), gives a true description of *w*. But this doesn't give us much guidance in some cases.

Suppose we are trying to evaluate 'horses have tails' with respect to *w*. You maintain, reasonably enough, that what 'horses have tails' actually says is that tails are had by Northern Dancer, Secretariat, . . . (fill in here the list of all actual horses). You conclude that 'horses have tails' is true of *w* iff Northern Dancer, Secretariat, . . . (or perhaps just those of them that exist in *w*) have tails in *w*.

Someone else maintains, just as reasonably, that 'horses have tails' says that if anything is a horse, then it has a tail. She concludes that 'horses have tails' is true of *w* iff the things that are horses in *w* have tails in *w*. The two of you disagree, then, about how to evaluate 'horses have tails' at a world that contains all our horses (complete with tails) plus some *additional* horses that lack tails.

Who is right? What is really said by an utterance of 'horses have tails' and how do we tell whether it is true of a counterfactual world? These questions have no clear answers. One might, I suppose, look for answers in the theory of what is expressed, or what is said, by sentences in contexts. But it would be with a heavy heart (and not only because the notion of what is said is so slippery and vague). Almost every question in semantics can be framed as a question about what some S expresses in some context. It would be nice if we didn't have to do the full semantics of English before the truth-conditions of 'necessarily S' could be given.

If there were no way around this problem, I doubt that Kripke's approach would have found such widespread acceptance. One imagines, then, that the Kripkean has a response. Here is how I imagine it going: 'You are taking the "saying what it actually says" phraseology too seriously in some way. If any real weight were going to be laid on that way of putting it, then yes, a story would be needed about how it is determined what is said. But "saying what it actually says" is just a heuristic. It reminds us that it doesn't matter, in considering whether S is true of *w*, what the citizens of *w* mean by S. How in that case *is* true-of to be understood, you ask? One option is to treat it as primitive. But this option is problematic. It gives the skeptic about metaphysical possibility too big an opening: she can claim to find the primitive incomprehensible. It would be better if we could *explain* truth-of in terms that the skeptic, as a speaker of English, already understands. This can be done using the subjunctive conditional. To say that S is true of a world *w* is to say that *had w obtained, it would have been that S*.'[2]

---

[2] See in this connection Chalmers (2000).

Consider in this light the 'controversy' about horses and their tails. When we evaluate 'horses have tails' with respect to w, is it only the actual horses that matter, or do horses found only in w have to be taken into account as well? Suppose that although actual horses have tails, w's additional horses include some that are tail-less. Is 'horses have tails' true of w?

The subjunctive account makes short work of this conundrum. Had w obtained, it would *not* have been that horses had tails; there would have been some horses with tails and some without. So 'horses have tails' is false of w.

Return now to the case of a w where 'tail' means *wing*. Does the fact that w-people speak falsely when they say 'horses have tails' show that 'horses have tails' is false of w? It doesn't, and we can now explain why in a theoretically uncontroversial way. The question is whether horses would still have had tails, if people had used 'tail' to mean wing. They clearly would have; how people talk doesn't affect the anatomy of horses. Had 'tail' meant wing, 'horses have tails' would not have been true, but horses would still have had tails.

6 Disparity

All this is to emphasize the *disparity*, in the immediate aftermath of *Naming and Necessity*, between metaphysical and conceptual necessity. The first was in good shape—because it went with 'S is true of w', which could be understood as 'it would have been that S, had it been that w'. The second was in bad shape—because it went with 'S is true when uttered in w', which had to be understood as 'it would have been that S was true, had it been that w, *and had S retained certain aspects of its actual meaning*'.

Then a brainstorm was had that seemed to restore parity.[3]

Recall what we do to judge metaphysical necessity. We ask of various worlds w whether S (*our* S, natch) is true *of w*. The Kripkean tells us that to judge conceptual necessity, we need to ask, not whether S is true *of w*, but whether it is true (as spoken) *at w*. But maybe it wasn't really necessary to move S over to w.

---

[3] At least three ideas were involved. (1) Instead of moving S over to w, bring w back to S. To do that, (2) evaluate S on the hypothesis that w actually obtains. To do that, (3) evaluate the indicative conditional 'if w actually obtains, then S'. (1) and (2) are present to some degree in Evans (1979) and Davies and Humberstone (1980), and are explicit in Chalmers (1994). I am not aware of any discussion of (3) before Chalmers (1996, 2000). See also Segerberg (1972), White (1982), and Stalnaker (1972, 1990, 1991).

A different option is to move w over to actuality: to the place where the token of S that we want evaluated in fact occurs.[4]

All right, but how do we do that? It looks at first very simple. Just as, when judging metaphysical necessity, we consider w as counterfactual, so, when judging conceptual necessity, we consider it as *counteractual*. We consider it as a hypothesis about what *this* world is like. Of course, we do not in general *believe* the hypothesis. But that should not deter us; we are masters at working out how matters stand on hypotheses we reject. Evaluating S with respect to counteractual w is asking whether S holds on the hypothesis that w is (contrary to what we perhaps think) this very world.

For example, it is conceptually possible that Hesperus $\neq$ Phosphorus because, if we suppose for a moment that this world is one in which Hesperus-appearances are due to Mars and Phosphorus-appearances to Venus, then clearly (on that supposition) we are *wrong* to think that Hesperus = Phosphorus. It is not that counterfactual people are wrong about *their* world. It is we who are wrong about *our* world, on a certain hypothesis about what our world is like.

This sounds like progress, but we should not celebrate too soon, because the disparity with metaphysical modality is not entirely gone.

I said that everyone would (should!) have been unhappy if they had been asked to treat 'true of counterfactual w' as a semantic primitive. We are willing to rest so much on true of because of the *explanation* we have been given of that notion: S is true of w iff, had w obtained, it would have been that S. It is this biconditional, with 'true of' on the left and a counterfactual on the right, that convinces us that there's a there there.

Apart, though, from some suggestive talk about what to say 'on the supposition' that w obtains, we have no comparable explanation of what is involved in S's being true with respect to counteractual w. If we use 'true if w' for truth with respect to a world conceived as actual, the problem is that 'true of' has been translated into English and 'true if' has not.

---

[4] A third option is to leave S and w where they are, and treat 'true if' as a trans-world primitive. This is one possible reading of Chalmers's remark (1994) that 'we can retain the thought from the *real* actual world and simultaneously ask its truth-value in other actual-world candidates without any loss of coherence'. He adds in a footnote that 'Doing things this way . . . avoids a problem . . . raised by Block (1991) and Stalnaker (1991). The problem is that of what must be "held constant" between contexts . . . On my account, nothing needs to be held constant, as we always appeal to the concept from the real world in evaluating the referent at [an actual-world candidate]' (1994: 42). This is certainly one way to go. But it has its costs. If taking 'true of' as primitive is obscurantist, primitivism about 'true if' borders on mysticism (our pre-theoretical grip on the second is that much weaker).

7 Indicatives

One proposal about this suggests itself immediately. Since 'true of' goes with a *counterfactual* conditional, 'true if' perhaps goes with the corresponding *indicative* conditional. 'S is true if *w*' says that *if w in fact obtains (evidence to the contrary notwithstanding), then S.*[5]

The proposal is intriguing because it offers to link two deep distinctions: metaphysical versus conceptual necessity, on the one hand, and subjunctive versus indicative conditionality, on the other. The reason it is only metaphysically necessary that Hesperus = Phosphorus is that there are worlds *w* such that, although Hesperus would have been Phosphorus *had w* obtained, it is not Phosphorus if *w does* obtain.

Do the two conditionals really 'predict' the two types of necessity? Before attempting to decide this, we need to remember how we got here. It was important for metaphysical necessity to keep what–is–said fixed as we evaluate S at *w*. Subjunctives are valued because they in effect do this, without dragging us into controversies about what is in fact said. It is *not* important to conceptual necessity to keep what–is–said fixed; indeed, we are willing and eager that it should change in certain respects under the impact of this or that counteractual hypothesis. (For example, we are eager for 'Hesperus = Phosphorus' to take on a content having to do with Venus and Mars.) Crucially, though, we do *not* want S's meaning to be changeable in *all* respects. (We don't want 'Hesperus = Phosphorus' to acquire a content having to do with nonidentity.) Indicatives are attractive because they seem to deliver an appropriate measure of meaning-fixation, just as subjunctives did on the metaphysical side.

Indicatives *appear* to deliver an appropriate measure of meaning-fixation. But when you look a little closer, the appearance fades. Indicatives don't in fact deliver *anything* in the way of meaning-fixation. The meaning of S as it occurs in the consequent of an indicative conditional can be changed all you want by putting the right kind of misinformation into the antecedent. Example: If 'tail' had meant wing, horses would still have had tails. But suppose that 'tail' *does* mean wing; it has meant wing all along, not only in others' mouths but also our own; a brain glitch (or demon) leads us systematically astray when we reflect on the meaning of that particular word. Then, it seems clear, horses do not have tails. If 'tail' as a matter of fact means wing, then to say that horses have tails is to say that they have wings. Horses do not have wings. So if 'tail' means wing, then horses do not have tails.[6]

[5] Chalmers (2000).

[6] Indicative conditionals are conditionals with antecedent and consequent in the indicative mood. Philosophers have proposed various theories of these conditionals. One, defended by

You may say: why should it be a problem if there are counteractual worlds at which horses lack tails? That is not the problem. The problem is that there are worlds where horses lack tails *not for anatomical reasons but on account of 'tail' not meaning tail.* If horses can lose their tails that easily, then take any S you like, it is true in some counteractual worlds and false in others. It is true in worlds where S means that X, and X is the case, and false in worlds where S means Y, and Y is not the case. This spells disaster for the indicative approach to conceptual possibility. It should not make 'Hesperus ≠ Hesperus' conceptually possible that there are worlds where people use '≠' to express identity.

8 Narrow Content

The indicative is not the conditional we want. But it is close. We want a conditional A → C that is *like* the indicative except in one crucial respect: C is protected from a certain sort of meaning shift brought on by A.

An example of the 'good' or 'permitted' sort of meaning shift is the kind exhibited by 'Hesperus ≠ Phosphorus' on the supposition that Phosphorus-appearances are caused by Mars. An example of the 'bad' sort of meaning shift is that exhibited by 'Phosphorus ≠ Phosphorus' on the supposition that '≠' expresses identity.

It may seem that the answer is staring us in the face. The 'bad' kind of meaning shift is the kind that mucks with *S's narrow content.* Our conditional should be such that S's narrow content is the same when we condition on *w* as when we don't. (The indicative is wrong because the narrow content of 'horses have tails' is one thing if 'tail' means wing, another thing if it doesn't.) Calling the actual narrow content NC, attention is to be restricted to worlds such that *w* obtains → S (still) means NC.

But, although helpful as an intuitive constraint, this doesn't solve our problem. This is partly because one doesn't know what the narrow content in fact is; NC has been pulled out of a hat. Second, though, to appeal to narrow content in this context gets things the wrong way around. The reason for being interested in 'S is true if *w*' was to get a better handle on conceptual necessity.

Grice (1989), is that they are 'material', or truth–table, conditionals. Another, defended by Adams (1975), is that they are probability conditionals. Chalmers in recent work declares a preference for the material conditional, regardless of its relation, if any, to the indicative. (He requires the material conditional to hold a priori.) The objection in the text applies regardless. However the indicative is interpreted, A's a priori entailing C suffices for the apriority of 'if A then C'. The conditional 'if horses are wingless and "tail" means wing, then horses do not have tails' has A a priori entailing C, so the conditional is a priori.

But, as noted above, conceptual necessity and narrow content are two sides of the same coin. The idea is to explain narrow content using →, not → using narrow content.

9 Turning Out

Our problem now is similar to one faced earlier in connection with metaphysical necessity. It seemed that an account of true of would have to appeal to the notion of what is said. That would be unfortunate, because it would reverse the intended order of explanation. The what-is-said of an utterance (its broad content, nearly enough) is given by the worlds of which it is true. The special case in which S's broad content takes in *all* worlds is what is otherwise known as metaphysical necessity. That is why we don't want to use broad content to explain true of. Our current worry is the same, except that it concerns true if rather than true of, and narrow content rather than broad.

How did we deal with that earlier problem? By calling in the subjunctive. We said that S is true of w iff it would have been that S, had w obtained. The claim was that this construction *automatically* targets the agreement or lack thereof between w and S's broad content. Can a construction be found that automatically targets the agreement or lack thereof between w and S's *narrow* content, as the subjunctive does for broad content?

One that comes pretty close occurs in *Naming and Necessity* itself. Kripke notes that we're at first inclined to think that Hesperus and Phosphorus (although in fact identical) could have been distinct. Then we learn about metaphysical versus other types of necessity, and we lose the inclination; Hesperus and Phosphorus could not have been distinct. Even now, though, apprised of the metaphysical facts, we are still inclined to think that it *could have turned out* that Hesperus was distinct from Phosphorus.

It is this phrase 'could have turned out' that I want to focus on. Kripke is right to represent us as still inclined to think that it could have turned out that Hesperus was distinct from Phosphorus, even after we have taken on board that it could not have *been* that Hesperus was distinct from Phosphorus. The inclination persists even among practicing modal metaphysicians (who ought to know better, if there is better to know). This suggests that 'could have turned out' is special in ways we should try to understand.

It suggests it to me, anyway. Kripke apparently does not agree. He maintains that the second inclination is just as mistaken as the first. Not only could it not have been, it could not even have turned out that Hesperus was distinct from Phosphorus. This is only to be expected if 'it could have turned out that S'

means, as Kripke hints it does mean, 'it could have been that: S and we believed that S and with justification'. This interpretation, however, leaves it a mystery why the second inclination outlasts the first—why we persist in thinking that it could have turned out that Hesperus wasn't Phosphorus even after giving up on the idea that Hesperus could have been other than Phosphorus.

I propose that the persisting thought is correct. Kripke to the contrary, it could indeed have turned out that Hesperus wasn't Phosphorus. That is what *would* have turned out had it turned out that Phosphorus-appearances were appearances of Mars. It could *not*, however, have turned out that Phosphorus ≠ Phosphorus, even granting that '≠' could have turned out to express identity. That is a way for it to turn out that that 'Phosphorus ≠ Phosphorus' is true, not a way for it to turn out that Phosphorus ≠ Phosphorus.[7]

10 Conceptual Possibility

It would have turned out that C, had it turned out that A shares features with both the indicative conditional and the subjunctive. It resembles the indicative in making play not with counterfactual worlds, but with suppositions about *our* world. It resembles the subjunctive in that the consequent C is protected from a certain kind of semantic influence on the part of A. The way C (narrowly) represents things as being is left untouched by 'had it turned out that A'. The role that the antecedent plays is all on the side of whether things are, on the hypothesis that A, the way that C (in actual fact, given that the hypothesis is false) narrowly represents them as being.

If 'tail' means wing, we said, then horses lack tails. → is supposed to be different in this respect. It should not be that w *(in which 'tail' means wing) obtains* → *horses lack tails*. That is the result we get if → is a 'would have turned out' conditional. For it is *not* the case that horses would have turned out to lack tails, had it turned out that 'tail' meant wing. It is not for linguistic reasons that horses have tails; so they are not deprived of their tails by the linguistic facts turning out differently.

---

[7] Chalmers employs similar wording when he introduces primary intensions: 'there are two quite distinct patterns of dependence of the referent of a content on the state of the world. First, there is the dependence by which reference is fixed in the actual world, depending on *how the world turns out: if it turns out* one way, a concept will pick out one thing, but *if it turns out* another way, the concept will pick out something else' (1996: 57, italics added). I applaud the use of 'turns out', but I think the mood should be subjunctive—if it had turned out—rather than indicative—if it does turn out. If it turns out that 'tail' means wing, then horses lack tails. But that 'tail' means something different in w should be irrelevant to the question of whether w's horses have tails. Otherwise conceptual necessity is trivialized. See also Jackson (1994, 1998).

One can come at → from the other direction. If Phosphorus-appearances had been due to Mars, Phosphorus would still have been Hesperus. → is supposed to be different in this respect too. We want there to be worlds w such that w *obtains* → *Hesperus* ≠ *Phosphorus*. That cannot happen unless the broad content of 'Hesperus ≠ Phosphorus' can be changed by conditioning it on the hypothesis that w obtains. Here too, 'would have turned out' delivers the goods. Had it turned out that Phosphorus-appearances were due to Mars and Hesperus-appearances (still) to Venus, it would have turned out that Hesperus ≠ Phosphorus.

What these examples suggest is that 'would have turned out' conditionals exhibit just the right combination of (i) openness to shifts in broad content, (ii) intolerance of shifts in narrow content. I therefore propose *it would have turned out that C, had it turned out that A* as the proper interpretation of A → C. And I make a hypothesis:

(M)   It is metaphysically possible that S iff some world w is such that it would have been that S, had w obtained.

(C)   It is conceptually possible that S iff some world w is such that it would have turned out that S, had w turned out to be actual.

More simply, S is *metaphysically* possible iff it could have *been* that S, and *conceptually* possible iff it could have *turned out* that S.

11   Analyticity and Apriority

A priori truths are truths that can be known not on the basis of empirical evidence. How well that accords with the Kripkean notion of apriority depends on one's theory of justification. There is a danger, though, of its according very badly.

One theory says that all spontaneously arising beliefs start out justified. They can lose that status only if evidence arises against them. Suppose that this view is correct, and suppose that, on pulling the curtains open, I spontaneously come to think that the sun is shining. (I don't infer that it is shining from premises about how things perceptually appear to me.) Then I know that the sun is shining, and not on the basis of empirical evidence. And yet it certainly isn't a priori, as Kripke uses the term, that the sun is shining.

Another theory has it that our most 'basic' beliefs lack empirical justification, because they are epistemically prior to anything that might be said in their support. So, the belief that nature is uniform lacks empirical backing. If we know that nature is uniform, and let's assume we do, the knowledge is not empirical. But it isn't a priori in Kripke's sense that nature is uniform.

Apriority, then, is not *any* old kind of not-empirically-based knowability, as judged by any old theory of justification. That would let far too much in. A (very familiar) objection from the other side helps us to clarify matters. If experience cannot be appealed to at all, then shouldn't it be enough to stop S from being a priori if it is through experience that we *understand* S? The answer to this is that our interest is in how S is *justified*, our understanding taken for granted.

If that is the one and only concession made, then we wind up with a roughly Kripkean notion of apriority. S is a priori iff it is *knowable just on the basis of one's understanding of S*. Or, better, it's a priori *for me* iff *I* can know it just on the basis of my understanding of S. This is why the originator of a name is apt to know more a priori than someone picking the name up in conversation. The mental state by which Leverrier understands 'Neptune' tells him that Neptune, if there is such a thing, accounts for the perturbations in the orbit of Uranus. The mental state by which others understand 'Neptune' is liable to be much less informative about Neptune's astronomical properties.

Apriority is knowability on the basis of understanding. Understanding is, one assumes, knowing the meaning. But what meaning?

Perhaps understanding is knowing meaning 'in the strict sense': the sense that ignores reference-fixing descriptions. But Kripke calls it a priori that Hesperus = Hesperus, and a posteriori that Hesperus = Phosphorus, though the strict meanings are the same. More likely, then, it is knowledge of meaning in the *loose* sense that makes for understanding. The closest thing to loose meaning in our framework is narrow content. So it does not do *too* much violence to Kripke's intentions to say that S is a priori iff one can know that it is true just on the basis of one's grasp of its narrow content.

Kripke calls S analytic iff 'it's true in virtue of meanings in the strict sense'. This definition has to be treated with some care, since the strict meaning of 'Hesperus = Phosphorus' is a singular proposition of the form $x = x$, and Kripke does not want 'Hesperus = Phosphorus' to come out analytic. (It is not a priori, and Kripke thinks that analytic truths are a priori.) Then what is his intent in speaking of 'meanings in the strict sense'? He cannot have been trying to *include* statements ('Hesperus = Phosphorus') that are true in virtue of strict meaning as opposed to loose. He must have been trying to *exclude* statements ('Hesperus is visible at night') that are true in virtue of loose meaning as opposed to strict. This is, in effect, to limit analyticity to 'Fregean' sentences: sentences to which the loose/strict distinction does not apply. S is analytic iff it is true in virtue of its Fregean meaning, that being the only meaning it has.

Now, though, one wants to know: why should it stop S from being analytic if *in addition* to its truth-guaranteeing Fregean meaning, it has a (possibly not

truth-guaranteeing) Kripkean meaning? Or, to put it in narrow/broad terms, if S has a truth-guaranteeing narrow content, why isn't that enough to make it analytic, quite regardless of whether it has a broad content in addition?

True-blue Kripkeans will reply that narrow content is not (except *per accidens*, when it agrees with broad) part of *meaning*. Narrow content is *metasemantical*, not semantical.

But this, one may feel, is just terminological fussiness.[8] Even Kripke considers it a *kind* of meaning—meaning in the loose sense—and he says explicitly that some might want to define analyticity as truth in virtue of *that*. So, it does not do *too* much violence to Kripke's intentions to let analyticity be truth in virtue of narrow content. (This fits with our account of Kripkean apriority as knowability in virtue of grasp of narrow content.)

Now, finally, we can ask the question that matters: Is conceptual necessity a kind of apriority, or a kind of analyticity, or both?

I do not think there can be much doubt that it is a kind of analyticity. A conceptually necessary sentence is one true in all counteractual worlds. These worlds comprise what Chalmers calls the sentence's *primary intension*, and primary intension is his candidate for the role of narrow content. So, a conceptually necessary sentence is one whose narrow content is such that, no matter which world is actual, it comes out true. Truth guaranteed by narrow content is analytic truth.

Is conceptual necessity *also* perhaps a kind of apriority? As just discussed, the narrow content of a conceptually necessary sentence is such as to guarantee its truth. Does it follow that someone *grasping* the content is thereby in a position to *see* that S is true?

That depends on what is involved in grasping a content (let 'narrow' be understood). S's content is, roughly, a bunch of conditionals of the form: it would (or wouldn't) have turned out that S, had *w* turned out to be actual. Someone who grasps the content is in a position to know the conditionals. So if S is conceptually necessary, then she is in a position to see, for each *w*, that had *w* turned out to be actual, it would have turned out that S. Doesn't this show that she can determine a priori that S?

No; in fact, we are still miles from that conclusion. Let it be that the speaker knows for each *w* that *w* obtains → S. It is wide open so far whether this knowledge is a priori. Someone who grasps S's meaning is in a position to come to know the conditionals *somehow or other*. A priori or a posteriori is an open question.

[8] I myself feel it is more than that, but this is the charge made by the narrow content enthusiast whose part I am playing.

You might think that the knowledge *has* to be a priori. If grasping S's content gives me knowledge of the conditionals, then I know the conditionals based on my grasp of S's content. Knowledge based on grasp of content is a priori knowledge.

This is entirely unconvincing. Grasping S's content 'gives me' knowledge of the conditionals only in the sense of putting me in a position to come to know them; my advantage over non-graspers is that I have 'what it takes' to know. That is roughly to say that understanding S is *necessary* if one wants to know whether S if *w*, or the most important necessary condition, or the only necessary condition one has to worry about. Apriority requires that understanding be *sufficient*. I have granted that understanding suffices for being in a position to *work out* whether S if *w*. If the working out involves experience, though, then the knowledge will not be a priori.

12   Peeking

I said that our understanding of S might not be enough to go on, when it comes to working out whether S holds in a world *w*. The 'official story' about evaluation at counteractual worlds strongly denies this. But the possibility has a way of sneaking in uninvited. Here is Chalmers:

> [A]s an in-principle point, there are various ways to see that someone (a superbeing?) armed only with the microphysical facts and the concepts involved could infer the high-level facts. The simplest way is to note that in principle one could build a big mental simulation of the world and watch it in one's mind's eye, so to speak. (1996: 76)

Say that this is right; I am able to build a mental model of *w*, and judge whether S is true in *w* by viewing the model with my mind's eye. The question is whether viewing a model of *w* and asking myself 'how it looks' S-wise is a way of coming to know S's truth-value in *w* a priori.

Here is a reason to think not. Asking yourself how something strikes you is using yourself as a measuring device. Information acquired by use of an external measuring device is a posteriori on anybody's account. Information acquired by use of an internal one seems no different. What matters is that an experiment is done, the outcome of which decides your response.

It might be argued that mental experimentation *is* different. Knowledge gained from it is acquired within the privacy of one's own mind. You determine that S without appealing at any point to information about the outside world. Shouldn't that be enough to make the knowledge a priori?

No, for you determine that you have a headache the same way. Knowledge of headaches is certainly not a priori. The modal rationalist in particular should agree, for my headache, if a priori, would be a counter-example to the proposed equation between apriority and truth in all counteractual worlds. 'I have a headache' fails in some counteractual worlds. A priori truths are supposed to hold everywhere.

Some internally acquired knowledge presumably is a priori. If you think up a counter-example to argument form F in your head, then you know a priori that F is invalid. What distinguishes this sort of case, where you do know a priori, from the case of looking at a mental model with the mind's eye?

Two things. First, when you conjure up an image of *w*, you are *simulating* the activity of really looking at it. Simulated looking is not a distinct process, but the usual process run 'off-line'. Knowledge gained by internal looking is not a priori because it is acquired through the exercise of a perceptual faculty rather than a cognitive one.

Second, some imagined reactions are a better guide to real reactions than others.[9] Imagined shape reactions are a good guide, you say, and you are probably right. But it is hard to see how the knowledge that they are a good guide could be a priori. If the mind's eye sees one sort of property roughly as real eyes do, while its take on another sort of property tends to be off the mark, that is an empirical fact known on the basis of empirical evidence. I know not to trust my imagined reactions to arrangements of furniture, because they have often been wrong; now that I see the wardrobe in the room, I realize it is far too big. It is only because they have generally been right that I am entitled to trust my imagined judgments of shape.

The temptation to think of simulation as a source of a priori knowledge is due in part to there not being much that we are able to simulate. There might be beings who, given only the microphysical blueprint of, say, an exotic fruit, are able to imagine its color in much the way that we are able to imagine its shape. They come to know that rambutans are red, without ever laying eyes on one. I take it that no one would consider the knowledge to be a priori. These beings did not deduce the color from microphysics. Information was also needed about how that microphysics appears to human eyes. They obtained

---

[9] Stepping into the lake, you say, 'It's colder than I thought.' The earlier thought might have been a real judgment based on partial information (it's August, lots of people are swimming), but it might also have been a simulated judgment based on full information about the water's kinetic properties. You imagine yourself stepping into water with those properties, and it seems to feel warmer than water like that really does feel. (Most of us do something like this with temperature properties; 80 degree water is surprisingly cold.)

this information experimentally, by simulating an encounter with a rambutan, and using it to predict the outcome of a real encounter.

Suppose that we had been able to simulate reactions in other modalities. Suppose we could determine the taste and smell of a microphysically given item with the mind's tongue and nose. Would that make it an a priori matter how rambutans [insert chemical description here] tasted? No. How a thing tastes is an empirical question. One does not feel that it escapes being a priori only because of a contingent incompleteness in our nature. It would still have been an empirical matter how rambutans tasted, even if God had been more generous in the mind's sense-organ department.

These claims might be accepted but shrugged off as irrelevant. It doesn't matter if self-experimental knowledge is a posteriori, for any suggestion of self-experimentation was inadvertent. 'I looked at *w* and saw it to contain so-and-so's' is only a colorful description of something far more innocent: intellectually contemplating a world *description* and *thinking* my way to a conclusion about whether there are so-and-so's in *w*.

That is fair enough, on one condition. Self-experimentation had better not be *needed* to work out whether S holds in *w*. It had better be that one can reason from a microphysical description of *w* to a conclusion about whether or not S. *No peeking*. I assume that Chalmers would agree; for, if peeking is allowed, the inference from 'S holds in all candidates for actuality' to 'it is a priori that S' clearly does not go through. This inference is crucial to the view that Chalmers calls 'modal rationalism'.

Given how much hangs on our ability to evaluate S without peeking, one might have expected a show of vigilance on this score. If we are playing 'pin the tail on the donkey', you watch me like a hawk. You know how hard I find it to ignore information right in front of my nose. The same should apply when the game is 'decide the truth value of S'. If it is difficult to infer S (¬S) from microphysics, I will be tempted to switch to sensory imagining. Knowing this, you will take pains that my mind's eye is completely shut, or completely covered by my mind's blindfold.

The need for vigilance is never mentioned, as far as I know, in the modal rationalist literature. Here is how the passage quoted above continues:

Say that a man is carrying an umbrella. From the associated microphysical facts, one could straightforwardly infer facts about the distribution and chemical composition of mass in the man's vicinity, giving a high-level structural description of the area. One could determine the existence of a male fleshy biped straightforwardly enough. . . . It would be clear that he was carrying some device that was preventing drops of water, otherwise prevalent in the neighborhood, from hitting him. Doubts that this device is really an umbrella could be assuaged by noting from its physical

structure that it can fold and unfold; from its history that it was hanging on a stand that morning, and was originally made in a factory with others of a similar kind. (Chalmers 1996: 76)

When I try to 'determine' these higher-level facts, I find myself relying on visual imagining at every turn. 'Keep your mind's eye scrunched tight,' I am told. I can try, but then the higher-level facts go all mysterious. The feeling intensifies when I read how 'doubts that the device is an umbrella can be assuaged'. Never mind how they are assuaged; I do not see how the umbrella idea came up in the first place.

I realize how it's supposed to go. I start with objective, geometrical information. A chain of a priori inferences leads to 'it's shaped like an umbrella'. That conclusion combines with a host of others to establish its umbrella-hood beyond any doubt. Visualization is barred, so I have no idea of how the object looks. (Eventually it may strike me that since the object is an umbrella, it probably looks like one.)

Is this possible? It helps to look at a simpler case. I am to infer a plate's shape (it's in fact round) from premises about the arrangement of its microphysical parts. The premises might take various forms, but assume for definiteness that the arrangement is specified in analytic geometry terms. I am told that the object's teeny-tiny parts occupy the points (x, y) such that $x^2 + y^2 < 63$. (The plate is two-dimensional, no pun intended.) If I am to reason from this to the object's shape, I must know, implicitly at least, conditionals like the following:

if R is circumscribed by the points (x, y) such that $x^2 + y^2 = 63$, then R is round;

if R is circumscribed by the points (x, y) such that $x^4 + y^4 = 63$, then R is not round.

I should know many, many conditionals of this nature, one per lower-level implementation of roundness, and, I suppose, one per implementation of non-roundness. And, most important of all, I should know the conditionals a priori, just through my grasp of the relevant English words.

But, it isn't clear that I *do* know many conditionals like these. (I am tempted to say that it's clear that I don't.) And the few that I do know, I don't seem to know a priori. It wasn't learning the meaning of 'round' that taught me the formula for circles. I worked it out empirically by graphing the formula, *looking* at the figure I had just drawn, and then *reflecting* on how I was inclined to describe the figure. (I take it that no one has their first encounter with roundness in a geometry class.)

I do not say that the above shows that you *have* to peek. There may be other ways of proceeding that haven't occurred to me. All I mean to be claiming for

now is that 'one can find the umbrellas in *w* without peeking, just by virtue of one's competence with the word' is a *substantive and surprising thesis*. Theses like this need to be argued for, and no argument has been given. A priori entailment has been presented as what you would expect, unless a skeptical philosopher had got to you first.

13 Recognitional Predicates

Now let me move on to urging in a positive way that there is only so much we can judge with the mind's eye averted. I think that one *can't* always tell, just by drawing inferences from a world description, whether the world is one where it turns out that S. If that is right, then the method that Chalmers didn't really mean to be advocating, and that figures only inadvertently in his narrative, is in some cases the only possible method. This will be argued for *observational* predicates (starting with the subtype *recognitional*), then *evaluative* predicates, then, finally, *theoretical* predicates.

What marks a predicate P as observational? The usual answer is that understanding P involves an ability to work out its extension in *perceptually* (as opposed to intellectually) presented scenarios. To determine P's extension in a world, I have to cast my gaze over that world—at candidate Ps in particular—and see how it strikes me.

Nothing has been said about the kind of appearance that marks a thing as P. Sometimes *x* is judged P because our experience of *x* has a quality Q notionally independent of P. So, *x* is tantalizing if, roughly, the experience of it makes one want to get closer and know more. Other times the experience that marks *x* as P is the experience of it as being precisely P. One judges *x* to be P because P is how it looks or feels or sounds . . . . This is what I am calling a *recognitional* predicate.

Examples are bound to be controversial, so let me just follow Kripke. Kripke says that 'the reference of "yellowness" is fixed by the description "that (manifest) property of objects which causes them, under normal circumstances, to be seen as yellow" ' (1980: 140 n. 71). We understand by yellowness whatever property it is that makes objects look yellow, or gives rise to the sensation of yellow. The predicate 'yellow' is recognitional on this view, since the yellow objects are picked out by their property of looking yellow.

Suppose Kripke is right about our understanding of 'yellow'. What are the implications for the way yellow things are identified in a candidate *w* for actuality? It's clear that *x* has to look yellow to be counted into the predicate's extension. But to whom? Perhaps it needs to look yellow to the *w*-folks,

including one's counteractual self. If it is counteractual Steve's reactions that matter, then I don't need to experience x myself to determine if x is (in w) yellow. I can infer x's color a priori from what the relevant world description says about the experiences Steve has when experiencing x.

But what *does* the world description say about counteractual Steve's experiences? Suppose, first, that it describes them in intrinsic phenomenological terms; banana-caused visual experiences are said to have intrinsic phenomenological property K. This doesn't yet tell me whether bananas are yellow, for I don't know that K is the phenomenology appropriate to experiences of *yellow*. I can't determine that without giving myself a K-type experience and checking its content: do I feel myself to be having an experience of yellow or of green?

Suppose, on the other hand, that counteractual Steve's experiences are described intentionally, as 'experiences of yellow', 'yellow' being the predicate whose corresponding property we are trying to identify. Then we would seem to be caught in a circle. The referent of a compound expression depends on the referents of its parts. So any intelligence we might have about what it is to be an 'experience of yellow' must come from prior information about (among other things) what it is to be 'yellow'. But then the referent of each of these two phrases depends on that of the other.[10]

Kripke must have been aware of this problem. He notes that '[s]ome philosophers have argued that such terms as "sensation of yellow", "sensation of heat", . . . and the like, could not be in the language unless they were identifiable in terms of external observable phenomena, such as heat, yellowness'. And he says that 'this question is independent of any view argued in the text' (1980: 140 n. 71). Kripke doesn't mind, in other words, if one can't identify sensations of yellowness until one has identified the property they are sensations of. How, if that is so, can we hope to identify yellowness by way of sensations of yellow?

Here is what I think Kripke would say. Yellowness is identified not by a *condition* on experience ('such as to give rise to sensations of yellow'), but by the experience itself. The objects I call yellow are the ones that *look* yellow. If the yellow things were identified by an experiential *condition*, then we would face the problem of working out which experiences were of the indicated type. But that is not our situation. Far from being something in need of discovery, the experience of yellow is *part of the discovery process*.[11] I don't have to *identify* my

---

[10] One option is to say that yellowness and the sensation of it are identified together by means of a gigantic Ramsey-type theoretical definition. This is filed under the heading 'just a pipe dream until somebody supplies details'.

[11] The issue here is much like the one raised by Putnam's 'descriptivist' interpretation of the causal theory of reference. Putnam suggests that words have their reference fixed by a causal

yellow-experiences in order to learn by their exercise, any more than I have to identify my eyes in order to learn by use of them.[12]

There is a second reason why Kripke would (should) not take 'yellow' to have its reference fixed by an experience-implicating description. What will the description say about proper viewing conditions?

This is a problem that he himself raises for a related view: the view that 'yellow' is *defined* as 'tends to produce such and such visual impressions'. Tends to produce them under what circumstances, Kripke asks? Any answer will be unsatisfactory: 'the specification of the circumstances C either circularly involves yellowness or . . . makes the alleged definition into a scientific discovery rather than a synonymy' (1980: 140 n. 71). If C-type circumstances are circumstances where we are not deceived as to yellowness, then (while it may be analytic that x is yellow iff it looks yellow in C-type circumstances) the definition uses 'yellow', so cannot explain its meaning. If C-type circumstances are ones where (say) the light is of such-and-such a composition, no one is suffering from jaundice, the object is not a Benham's disk rotating at such-and-such a rate, etc., then, while it may be true that x is yellow iff it looks yellow in C-type circumstances, it is not definitionally true, but empirically so.

If this is a good objection to the idea that 'tends to . . . in circumstances C' defines 'yellow', it would seem to be equally hard on Kripke's own claim that 'yellow' has its reference fixed by that description. Either C-type circumstances are ones where we are not deceived as to yellowness, or they are ones where the light has such-and-such a composition, etc. If the first, then, while it may be a priori that x is yellow iff it looks yellow in C-type circumstances, the reference-fixer presupposes yellowness, and so cannot be used to identify it. If the second, then, while it may be true that x is yellow iff it looks yellow in C-type circumstances, it is not a priori true, as it would be if the description fixed 'yellow''s reference.

---

condition. One finds the referent by looking for whatever stands in the right causal relation to speech. This makes for circularity problems, since one needs to know which relation causation is to work out what 'causation' denotes. From here it is a short step to radical indeterminacy of reference. The almost universal response was that reference is fixed *causally*, not *descriptively* by a condition alluding *inter alia* to causal relations. Kripke as I am reading him says something similar: reference is fixed experientally, not descriptively by a condition alluding *inter alia* to a certain sort of experience.

[12] I like what Colin McGinn says about perceptual concepts. Some think that 'When a concept is applied to a presented object that is always a further operation of the mind, superadded to the mere appearance of the object in perceptual consciousness. On my way of looking at it, concepts figure as *substitutes* for perceptual appearance—. . . they are needed for intentionality only when the object is not being perceived' (1999: 324).

One can reply in the same way as before. What marks a thing *x* as yellow isn't the *condition* 'tends to produce . . . under circumstances C'. What marks *x* as yellow is that that is how it looks. Someone can of course ask, how do you know the perceptual circumstances (including the condition of the perceiver) are right? But we do not say to this person, 'the present circumstances are of type C, and C defines rightness'. That would open us up to all the problems raised above. Our answer is, 'Why shouldn't they be right? What is it that leads you to suspect trouble?' It may not be a priori that what looks yellow under conditions C is yellow, but it does seem to be a priori that what looks yellow is yellow assuming nothing funny is happening. And that is an assumption we are always entitled to, unless and until we run into specific objections.

I hope this makes clear how our grasp of a predicate can be *recognitional* rather than intellectual. I do not reason my way to the conclusion that something is yellow from premises about what looks yellow under which conditions. The belief arises spontaneously in me when I look at a thing. That *has* to be how it works, for I have in general no a priori reliable information about which viewing conditions are appropriate. The most that is a priori is that *these* conditions are appropriate, unless there is reason to think otherwise.

If P is a recognitional predicate, then I have an a priori entitlement to '*These* conditions are (funny business aside) such that what seems P is P'. This is an entitlement that, by its nature, does not travel well. It lapses when we move from the world that really is actual to worlds only treated as actual for semantic evaluation purposes. For in lots of those worlds, we find (what from our actual–actual perspective is) funny business.

A few special cases aside, what looks yellow, is yellow. But things could have turned out so that whipped cream looked yellow—say, because a jaundice-like staining was characteristic of healthy eyes rather than diseased ones. This would not bother the people we turned out to be (they think *our* eyes are problematic), but it does bother us as we are. Whipped cream is white, and so whoever sees it as yellow is to that extent getting it wrong.

This has two semi-surprising consequences, which for now I'll just state without argument.

(1) Something known a priori need not hold in all counteractual worlds. It is a priori that funny business aside, what looks yellow, is yellow. But had our eyes turned out as described, objects would have turned out to look yellow that were in fact white. There is no mistake here, nor is anyone misled. Whipped cream is indeed what *they* mean by the word. It is just not what we mean by it, that is, it is not yellow.

(2) Something holding in all counteractual worlds might be knowable only a posteriori. Let F be a complete intrinsic characterization of some white chalk.[13] Could an F have turned out to be other than white? The chalk could have turned out yellow-*looking*, as already discussed. To have turned out *yellow*, however, it would have needed different (non-F-ish) intrinsic properties. So although it is a posteriori what color Fs in fact are, their color is conceptually necessary in the sense that it could not have turned out any different.

14   Observational Predicates

Everyone knows what it is for a figure to be *oval*. It is not hard to distinguish ovals from polygons, figure-eights, and so on. It is not even all that hard to distinguish ovals from otherwise ovular figures that are too skinny or too fat to count. To a first approximation, a figure is oval if it has the proportions of an egg, or a two-dimensional projection of an egg. I take it that few of us know in an intellectual way what those proportions are. What marks a figure as oval is not its satisfaction of some objective geometric condition, but the fact that when you look at it, it looks egg-shaped.[14]

Because our grasp of *oval* is constituted in part by how its instances look, one might be tempted to group it with 'response-dependent' concepts like *ticklish* or *tantalizing*. That would be a mistake. There are several respects in which *oval* is quite *un*like *ticklish*, which, once pointed out, make the label 'response-enabled' seem much more appropriate. Another term I shall use is 'grokking concept'. (I apply these labels to concepts, but, depending on one's other commitments, they could speak more to how the concept is grasped.)

Constitution: Why are ticklish things ticklish? That might mean 'what is the evidence that they are ticklish?' If so, the answer is that we respond to them in a certain way; they tickle us. If it means 'what qualifies them to be so regarded?', the answer has again to do with our responses. So far there is no contrast with *oval*. But suppose we now ask, 'in what does their ticklishness consist?' Eliciting or tending to elicit a certain reaction in us is 'what it is' to be ticklish. To be oval, though, is simply to have a certain shape.

---

[13]   More may have to be packed into F, such as prevailing natural laws.

[14]   I say *looks egg-shaped* and not *looks oval* because I want 'oval' to be an example of an observational predicate that is not recognitional.

Tracking: Our responses do not track the extension of 'ticklish'; they dictate it. It makes no sense to suggest that our tendency to be tickled by various things might not have been, or might have turned out not to be, a good guide to what is really ticklish. It is different with 'oval'. Our responses give us *access* to the extension of 'oval', but they do not dictate the extension.

Motivation: Why are the ticklish things picked out experientially? There is an in-principle reason for this: we want to classify as ticklish whatever is experienced in a certain way. Why are the oval things picked out experientially? There is no in-principle reason, but only a practical one: we have no other way of roping in the intended shapes.

Evaluation: Externalities are the same in *w* as here, but our responses are different. Suppose that our world had turned out to be *w*. What would have turned out to be ticklish? That which turned out to elicit the tickle response. What would have turned out to be oval? That which *does* elicit the oval response; that which *does* look egg-shaped. For dimes to have turned out oval, they would have had to turn out a different shape.

The 'evaluation' contrast is the one that matters, so let me dwell on it a little. Imagine someone who thinks that 'oval' applies to whatever strikes the locals as egg-shaped, in any *w* you like, considered as counteractual or counterfactual. This person has misunderstood the concept. If he were right about counterfactual worlds, then

dimes would have been oval, had they (although still round) looked egg-shaped.

If he were right about counteractual worlds, then

dimes would have turned out to be oval, had they (although still round) turned out to look egg-shaped.

This is false, too. The way to a thing's ovality is through its shape; you can't change the one except by changing the other. You can't make something oval by tinkering only with our responses.

What can we say to our confused friend to straighten him out? 'Oval' stands for things like *that*, the kind that we *do* see as shaped like eggs. The concept uses our responses as a tool—a tool that, like most tools, stops working if it's banged too far out of shape. The concept presupposes that our responses are what they are, and then leans on that presupposition in marking out the class of intended shapes. This is why its turning out that we saw dimes as egg-shaped would be a way for it to turn out (not that they were oval, but) that we were taking non-ovals for ovals.

A better analogy for our concept of oval is the concept expressed by '*that* shape' when we say, pointing at a sculpture, that 'that shape is eerily familiar'—or the one expressed by '*this* big' in 'a room has to be at least *this* big [gesturing at the surrounding walls] to hold all my furniture'.[15] The role of '*this* big' is not to pick out whatever old size one might turn out to be perceiving: *tiny* if one turned out to have been in a tiny room suffering an optical illusion. It is, rather, that one takes oneself to be perceiving a room of a certain size, and one has no way of knowing the size other than via its perceptual appearance.

15 Analyticity without Apriority

First there are the response-*dependent* concepts: ticklish, aggravating, tantalizing, painful-to-behold. Then there are the response-*enabled* concepts: oval, aquiline, jagged, crunchy, smiley-faced. Response-enabled concepts have their own distinctive pattern of evaluation at counteractual worlds. If *oval* were response-dependent, then one could determine its extension in *w* by asking what the people there saw as egg-shaped.[16] If it is response-*enabled*, then those counteractual responses are irrelevant. Ovality is to be judged not by *as-if* actual observers, but by *actual* actual observers. A thing in *w* is oval if it is of a shape that would strike *me* as egg-shaped were I (with my sensibilities undisturbed) given a chance to look at it.

This has consequences for what comes out analytic, or conceptually necessary. Consider a world *w* about which all I'm going to tell you is that it contains Figure 1. Is 'oval' true of this figure in *w* considered as actual? The answer is clear. All we need do to determine that it is oval is look at the figure, and note that it looks *like that*—the way that ovals are supposed to look.

Once again, I have not said anything about how observers in *w* see Figure 1. Maybe there are no observers in *w*, or maybe there are, but they do not think Figure 1 has the right sort of look. It doesn't matter, for we evaluate the figure with respect to our word 'oval', understood as we understand it. Our dispositions figure crucially in that understanding, so they are part of what we (imaginatively) bring to bear on the figure in *w*.

Now let's bring in our conditional →, the conditional used to define conceptual necessity. Is it or is it not the case that *w obtains* → *Figure 1 is oval*? Would Figure 1 have turned out still to be oval, had it turned out to be shaped as shown? You bet it would. Whether an as-if actual figure is oval is completely

[15] Peacocke (1989).

[16] So-called rigidified response dependency is for our purposes a minor variant of the unrigidified kind.
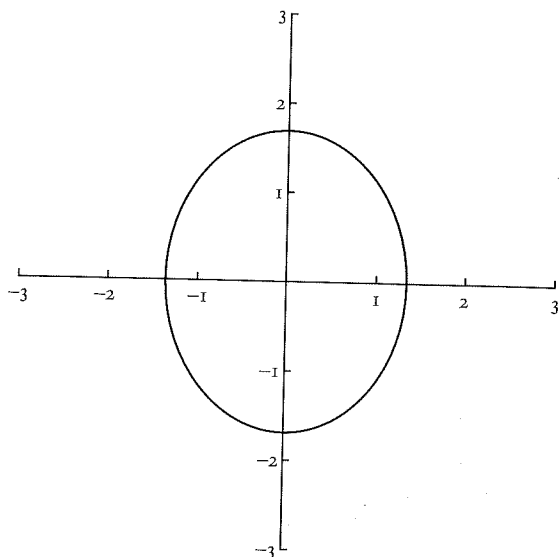
Fig. 1. Could this shape have turned out not to be oval?

determined by its shape. Things could have turned out so that Figure 1 did not look egg-shaped: we could have wound up with greater powers of visual discrimination, and as a result been 'bothered' by departures from an egg's precise shape that, as we are, we find it easy to ignore.[17] But Figure 1 would not in that case have turned out not to be oval. One wants to say, rather, that ovals would have turned out not to look the way they do look; ovals would have turned out to lack the feature by which things are recognized as ovals.[18]

Suppose we do some measurements and determine that Figure 1 is defined (up to congruence) by the equation $(x^2 + y^2)^2 - (x^2 - y^2) = 5$. Figures like that can be called *cassini*-shaped, or, for short, *cassinis*. (Giovanni Cassini (1625–1712) studied a class of figures of which this is one.[19]) 'Cassini-shaped' is an objective, third-personal predicate applying to all and only figures with

---

[17] Suppose for argument's sake that phenomenological similarity goes with the number of jnd's—just noticeable differences—separating one figure from another.

[18] I assume that the label 'oval' continues in the imagined case to be applied on the basis of egg-looking-ness. Does the fact that different things turn out to look that way make us (in that case) bad judges of ovality? Yes and no. Our counteractual responses are an excellent guide to what 'oval' would have turned out to mean. They are a bad guide, however, to what 'oval' does mean; they are a bad guide to what is in fact oval.

[19] The class of 'Cassinian ovals', although not all are really oval; indeed, not all are topologically connected.

---

the geometrical properties (that we all correctly take to be) exemplified by Figure 1.

Could things have turned out so that cassinis were not oval? If ovality in a world is purely a function of shape, then the answer is no. 'Cassinis are oval' is true in all worlds-taken-as-actual, which makes it (given our definition above) conceptually necessary.

But, of course, it is very far from a priori that cassinis are oval. To determine whether they are oval, you have to cast your eyes over (some of) them, and see how they look to you. There is no other way to do it. 'Cassinis are oval' is an analytic (conceptually necessary) truth that we are in no position to know a priori.

16 Other Intensions

If every world w is such that its cassinis are (to us) eggish-looking, then 'cassinis are oval' is analytic. Its meaning as encoded in our reactive dispositions guarantees its truth. But this is a kind of analyticity that we would not expect to make for apriority, because the route from understanding to extension and hence truth-value is inescapably observational.

To put it the other way around, one can't conclude from the fact that 'cassinis are oval' fails to be a priori that there is a counteractual world some of whose cassinis aren't oval. The premise you need for that is that 'cassinis are oval' is not *analytic*. But it *is* analytic. Given what the sentence means, it has got to be true.

Once again, the inference from (i) failure of apriority to (ii) a world that 'witnesses' the failure is crucial to modal rationalism. One might almost be forgiven for thinking that the main thing people *value* in the doctrine is its ability to deliver a counter-world. I assume, then, that modal rationalists would like, if possible, to *plug* the gap that seems to have opened up between analyticity (conceptual necessity) and apriority.

One approach harks back to the indicative account of truth in a counteractual world. For S to hold in counteractual w is, on that account, for it to be the case that if w obtains, then S. We rejected this account on the ground that it makes every sentence conceptually contingent. (If 'sibling' means parent, then sisters are *not* always siblings.) But, you may say, there is an obvious fix. It should be not merely true but a priori that if w obtains, then S. It is not a priori that if 'sibling' means parent, then sisters aren't always siblings. So a world where 'sibling' means parent is not on the new definition a world where the problematic sentence (some sisters are not siblings) holds.

Suppose we let S's *epistemic* intension be the set of worlds such that it's a priori that if w obtains, then S. And suppose that conceptual necessity is understood as necessity of the epistemic intension so defined. What happens to the argument above that conceptual necessity is a kind of analyticity but not a kind of apriority?

It might seem to fall apart. 'Cassinis are oval' may have a necessary primary intension, but its epistemic intension is contingent. (It is not generally a priori that if w obtains, then cassinis are oval; perhaps it is never a priori.) But then, if conceptual necessity goes with the epistemic intension, 'cassinis are oval' is not conceptually necessary. And so it no longer serves as a counter-example to the idea that whatever is conceptually necessary is a priori.

This assumes, however, that intensions built on a priori indicatives avoid the problems that were raised for intensions built on ordinary indicatives. Do they? What does seem clear is that the old examples no longer work. But this is for a correctable reason: namely, that sisters might, for all we know a priori, be one and all parents. It *is* a priori (let's assume) that sisters (if there are any) are not numbers. And so it is a priori too that if 'sibling' means number, then sisters (if there are any) are not siblings.

I have said that if 'sibling' means number, then sisters aren't siblings. Suppose that claim was based on empirical evidence. What would the evidence be? The only empirical fact in the neighborhood would seem to be this: 'sibling' does not in fact mean number. Call that the actual-meaning fact. Does it form part of my justification for believing that if 'sibling' means number, then sisters aren't siblings?

If it does form part of my justification, then should I *forget* 'sibling''s meaning, or come to hold an erroneous view of it, my justification would be compromised. Say I fall under the impression that 'sibling' does mean number. Have I now lost my grounds for thinking that *if* it means number, then sisters aren't siblings? Surely not. My reasons for thinking that *if 'sibling' means number, sisters are not siblings*, are just the same whether I believe the antecedent or not. How could forgetting what 'sibling' *does* mean compromise my ability to make inferences from a certain *hypothesis* about its meaning?[20]

---

[20] This is intuitive on its face, but it can also be argued for in the following way. It's agreed that I know that *if 'sibling' means number, then sisters aren't siblings*. The question is whether my justification is a posteriori, because based on the actual-meaning fact. If it is, then I lack the knowledge we've just agreed I have. Here is why. You are not said to know that *if A, then B* unless you know something from which B can be inferred, should it be discovered that A. Your justification for the conditional should therefore be 'robust' with respect to A: it should be such as to *stay in place* should one come to believe that A. (See Jackson 1979.) Your justification would not be robust if the conditional were based on ¬A. Conclusion: you don't know that *if A, then B* if your belief is

Where does this leave us? If my belief in the conditional is a priori, then there is a world that is not in the epistemic intension of 'sisters have siblings.' The same argument shows that *no* statement S, however a priori in appearance, has a necessary epistemic intension. I conclude that the a priori indicative strategy is no great advance over the plain indicative strategy. Both have the same basic problem: they make all intensions contingent, and so drain the class of conceptual necessities of all its members.

It might be held that the problem is not with the aprioritizing as such, but with the type of conditional aprioritized. A second option is to call S true in w-considered-as-actual iff it holds a priori that (w obtains → S)—it holds a priori that it would have turned out that S, had w turned out to be actual. The intensions that result can be called *priory* intensions. If conceptual necessity is necessity of the priory intension, maybe the inference to a counter-world can be saved. Certainly it isn't refuted by the cassini example; for although 'cassinis are oval' has a necessary primary intension, its priory intension is *not* necessary. (You need experience to establish that, had it turned out that w, it would have turned out that cassinis are oval.)

The priory intension is more than unnecessary, however. One can *never* tell a priori whether cassinis would have turned out to be oval, had it turned out that w. (I ignore the case where there are no cassinis.) 'Cassinis are oval' has, therefore, *nothing* in its priory intension. The same goes for 'cassinis are not oval'. It goes in fact for most sentences whose predicates express response-enabled concepts. If one can't determine a priori whether a counteractual object is P, then that object can't be put into P's priory intension, or ¬P's either. If the priory intensions of P and its negation are empty, then so in all likelihood are the priory intensions of sentences built on P.

Concepts like *oval* are not well-represented by their priory intensions. Still, you might say, why should that matter? The point of priory intensions is to predict epistemic status: if S fails to be a priori, there should be a world that is not in its priory intension. Why should the modal rationalist want any more?

One can see why more is wanted by considering the modal rationalist's refutation of physicalism. How does that argument go, with intensions understood as priory? First premise: it is not a priori that if PHYSICS, then PAIN. Second premise: if it is not a priori that if PHYSICS, then PAIN, then there are worlds that are not in that conditional's priory intension. Third premise: worlds not in

---

based on the premise that ¬A. Since I do know that *if 'sibling' means number, then sisters are not siblings*, my belief is not based on the premise that 'sibling' does not mean number. But that is just to say that my belief is not based on the actual-meaning fact. If it is not based on that, then it is not based on any empirical evidence.

that priory intension are zombie worlds—worlds physically like ours in which no one feels pain. Conclusion: there are zombie worlds.

The argument needs priory intensions to be like primary intensions in a certain respect. If PIs are primary intensions, then worlds that are not in a sentence's PI are worlds in which S is false. Does the same hold for priory intensions? It doesn't. If PIs are priory intensions, all we can say is that there is a *w* such that it *fails to be a priori that* it would have turned out that S, had it turned out that *w*. It might still be *true* that it would have turned out that S! There might be no way for it to turn out that PHYSICS without its also turning out that PAIN.

I present this as a problem for priory intensions, but epistemic intensions are every bit as vulnerable to it. That there are worlds lying outside S's epistemic intension does not show that there are worlds in which S is false, but only that you can't always get to S a priori. (The response will come that that is enough, since any S which cannot be verified by a priori means can be falsified by a priori means. But we have examples to the contrary, such as 'this equation describes an oval'.) I don't think that anything is gained, then, by switching to an aprioritized notion of truth at a world. The balloon just bulges in a different place. Yes, there is a world outside the intension, but there is no reason to think that it falsifies S, as opposed to just failing to a priori verify it. Better to stick with primary intensions as defined above. S is conceptually necessary iff it holds however things turn out.

17   Grasping Meaning

Why expect an analytic (conceptually necessary) sentence to be knowable a priori? Why expect a sentence whose meaning guarantees that it is true to have the further property that we can *see* that the sentence is true just from our grasp of its meaning? There might be ways of grasping meaning that do not tell us outright whether S is true if *w*, but only how to *work out* whether S is true if *w*. If S's meaning is grasped like that, then its not being a priori that S does not establish the existence of a falsifying world. The sentence might be (like 'this equation describes an oval') a posteriori but true in every world considered as actual.

The only way out is to maintain that the indicated kind of grasp is not possible. One will have to maintain that grasp of meaning always takes a certain form, a form that discloses to the grasper whether the meaning is truth-guaranteeing. If all I can do is *work out* whether $w \rightarrow S$, then I don't understand S. To understand, I have to *know* that $w \rightarrow S$.

Say that my understanding of S is *rationalistic* if it consists in whole or part of my *knowing* the conditionals. The road from analyticity to apriority would be a lot smoother if all understanding was rationalistic.

On what basis, though, can other forms of understanding be ruled out? What is the problem with grasping a word's meaning other than rationalistically? The closest thing I've found to an explicit discussion is Chalmers's reply to Loar in *The Conscious Mind* (1996).

Summarizing greatly, Loar (1990) thinks that *pain* is a recognitional concept[21] and that *C-fiber firings* is a theoretical concept, and that that is enough to make them cognitively distinct. Their distinctness notwithstanding, 'it is reasonable to expect a recognitional concept R to "introduce" the same property as a theoretical [concept] P.' So we cannot conclude from the non-apriority of 'C-fiber firings are pains' that C-fiber firings aren't pains. The failure of apriority might be because *pain* is recognitional and *C-fiber firings* isn't. If their a priori inequivalence is explained thus, then there is nothing to stop them from co-referring. These are fine things to claim, Chalmers says, but it is not clear that they can all be reconciled.

[Loar] gives the example of someone who is able to recognize certain cacti in the California desert without having theoretical knowledge about them. But this seems wrong: if the subject cannot know that R is P a priori, then reference to R and P is fixed in different ways and the reference-fixing intensions can come apart in certain conceivable situations. (Chalmers 1996: 373)

This might seem to be based on a misunderstanding. Observational concepts (of which recognitional concepts are a subtype) do not have their reference fixed in *any* epistemically available way; hence they do not have it fixed in a different way than holds for theoretical concepts.

What can Chalmers be thinking, then? He knows that Loar *says* that 'recognitional concepts refer "directly" . . . without the aid of reference-fixing properties' (1996: 373). He just thinks Loar is wrong about this. 'The very fact that a concept *could* refer to something else (a different set of cacti, say) shows that a substantial primary intension is involved' (1996: 373).

But, Loar can concede a substantial primary intension. The directness he is talking about is *epistemic*; one doesn't (and couldn't) *infer* that the cactus is R from its lower-level properties. A substantial primary intension is at odds only with *semantic* directness, as I now explain.

Fact: R applies to these things and not those. Why? What explains the differential treatment? If the question has an answer, as let's assume it does, it will

---

[21] This section is sloppy about recognitional versus observational, and also about Loar-recognitional versus recognitional in our sense.

be a truth of the form: R applies to *x* if and only if *x* is so-and-so. Consider this property of being so-and-so. It might be considered a reference-fixer for R; like a reference-fixer, it tells you how a thing has to be for R to refer to it. *Oval* too has a reference-fixer in this sense. Whether a figure is oval is not a brute fact about it, but depends on its shape.

A reference-fixer in the *theoretical* sense is a statement of the qualifications for being referred to by R, as these might be judged by a (smart enough) semantic theorist. A reference-fixer in the *ordinary* sense, though, is a statement of the qualifications for being referred to by R, as these might be explained by a (smart enough) user of the concept, trying to enumerate the factors she takes to make R applicable.

The claim about recognitional concepts is that they lack *ordinary* reference-fixers. Speakers do not apply *oval* on the basis of a condition that they know (even implicitly) that sums up the requisite features. Speakers do not know any conditions like that. They do not know any conditions that get the extension right no matter what. The condition that comes closest is *looks egg-shaped*. But, as we have seen, things could have turned out so that some *bona fide* oval had the wrong looks, and/or a non-oval had the right looks. I know an oval when I see one, and that seems to be enough.

Chalmers is right about one thing: it would be a mistake to deny recognitional concepts reference-fixers in the theoretical sense. That would be to deny that a thing's status as *oval* was a function of its lower-level properties. But if the claim is that recognitional concepts lack reference-fixers in the ordinary sense, then it would seem to be true. Speakers don't (and often can't) determine extensions a priori by asking what has the R-making properties.

How does all this bear on the issue that Loar and Chalmers are primarily interested in: the issue of physicalism? Chalmers, you will recall, argues as follows. It is not a priori that if PHYSICS, then PAIN; so the primary intension cannot contain every world; so there are worlds physically like this one in which pain is lacking; so physicalism is false.

The problem is (once again) with the inference from *not a priori* to *less than full primary intension*. With certain concepts the link between apriority and primary necessity breaks down. And the way it breaks down gives the physicalist an opening. She can say this: *Pain* (like *oval*) is a grokking, or observational, concept. That being so, whether an objectively described state is a case of pain cannot be determined just by rational reflection. One has to 'sample' the state by experiencing it from the right sort of first-personal perspective.

Two consequences should be noted. First, suppose there were a world *w* physically like ours but without pain. That world would do nothing to explain the non-apriority of 'if PHYSICS, then PAIN'; or rather, it would do nothing

that couldn't be done just as well by a world *with* pain. For *w* to help, our intuition of non-apriority would have to be owing to our awareness of *w*. But the relevant fact about *w* (that it lacks pain) is not available to us as students of its microphysical description. Just as you can't tell whether *w* has ovals except by sampling its shapes, so you can't tell whether it has pain except by sampling its brain states.

Second, not only is a world like *w* of no particular help in explaining the failure of apriority, it isn't needed. Suppose that *v* is a world just like ours in every physical respect. The question of whether there is (say) pain in *v* is the question of whether there is anything there that *hurts* if sampled in the right sort of first-personal way. Whether a state hurts when sampled by someone in the state is not the kind of thing that can be decided from the armchair. If we are trying to explain why PHYSICS doesn't a priori entail PAIN, a world whose zombie-ness can't be a priori ruled out works just as well as a true zombie world would.

18   Evaluative Predicates

Our grasp of a concept is *rationalistic* if it consists in whole or in part of a certain kind of knowledge: knowledge of conditionals of the form *w* obtains → *x, y, z, . . . are the Cs*. Suppose that your conditionals put *x, y, z, . . .* into a concept's extension in *w*, while mine count *x, y, z, . . .* out. Then, by Leibniz's Law, your concept and mine are not the same. A single concept cannot have conflicting extensions in the same world.

Now, in some cases, it seems quite right that disagreements about what goes into the extension should make for differences in the identity of the concept. If you and I can't agree about whether to call a certain almost-round figure oval, and this is not because of misinformation, error, or oversight on either side, then probably we have different concepts; probably we mean slightly different things by the word. There is no question of trying to work out who is really correct, because our beliefs are not really in conflict.

Similarly, if we can't agree about whether recently widowed 98-year-old males are bachelors, and not because either of us is misinformed or confused, then probably we mean slightly different things by 'bachelor'. There is no question of trying to work out who is really right, because we aren't really disagreeing.

A phrase sometimes used for concepts of this kind is *intolerant of brute disagreement*; if we have the same concept, we should not 'brutely disagree' about what falls under it. Are all concepts like that? Imagine that we disagree about whether it was wrong of Smith to tell a lie in hopes of saving his child

embarrassment. The disagreement can't be traced back to differences in factual information, or miscalculation or oversight on either side. Does this show that we mean different things by 'wrong'?

The usual view is that it doesn't. People who disagree about the extension of 'wrong' (and where the disagreement does not trace back to . . .) do not necessarily mean different things by the word. Likewise for disputes about what is beautiful or fitting or reasonable. You will get people angry if you brand these disputes 'merely verbal', just because you can't see any good way to bring the two parties into line. Some concepts, then, are *tolerant of brute disagreement*.

A lot of philosophers would claim something even stronger. So far is the meaning of 'right' from dictating a particular view of its extension that it positively *rejects* the notion that such dictation is possible. If I try to represent your side of a moral controversy as based in a misunderstanding of 'right', then I am the one who misunderstands. Questions of rightness are supposed to be *contestable* in the (rather minimal) sense that someone who brutely disagrees with you can't be charged on that basis alone with meaning something different by 'right'. Some concepts, then, seem to be *intolerant of intolerance of brute disagreement*.

How do we grasp the meaning of 'right'? If our grasp is rationalistic, then (assuming we mean the same by 'right') all of us know the same conditionals *w obtains → x, y, z, . . . are right and other things aren't*. Someone operating with different conditionals attaches a different meaning to the word. In that case, though, the concept is intolerant of brute disagreement. And our concept of rightness is, on the contrary, intolerant of such intolerance.

That is one argument for the conclusion that we do not grasp evaluative concepts in a purely rationalistic way. Here is another. Recall a well-known puzzle about right and wrong. On the one hand, you can't derive an ought from an is. 'If N then M', where N is descriptive and M is evaluative, cannot be known a priori. On the other hand, it does seem to be a priori that the evaluative facts are fixed by the descriptive ones. There is a tension here; we have trouble seeing how the two claims are supposed to hang together. But we do not get an outright contradiction unless it is supposed that our grasp of evaluative concepts is rationalistic.

Assume with the rationalist that if it is not a priori that S, then there's a counteractual *w* such that ¬S. Then, from the fact that N does not a priori entail M, we can infer the existence of a *u* such that *u obtains → (N & ¬M)*. Since it's also not a priori that if N, then ¬M, there should be a world *v* such that *v →* (N & M). But if N is descriptively complete (as we are free to suppose), then these two worlds taken together constitute a counter-example to the thesis that there can be no moral differences without underlying descriptive differences.

It could be objected that all *u* and *v directly* show is that things could have turned out so that N & M, and they could have turned out so that N & ¬M. To get to ◊ (N & M) and ◊ (N & ¬M), one needs to assume that M does not change in broad content between *u* and *v*. But that is a fair assumption, for the facts relevant to reference determination are descriptive facts, and these are the same in both worlds. Hence we can argue as follows:

(1) It is not a priori that if N, then M, or that if N, then ¬M.
(2) If it is not a priori that S, then there's a *w* such that *w → ¬S*.
(3) There are *u* and *v* such that *u →* (N & ¬M) and *v →* (N & M).
(4) M does not change in broad content between *u* and *v*.
(5) ◊ (N & M) and ◊ (N & ¬M).

But (5) is an a priori falsehood. Somewhere or other a big meta-ethical mistake has been made.

I claim that the puzzle has nothing essentially to do with ethics. Consider the conditionals, 'if *x* is cassini-shaped, then it is oval', and 'if *x* is cassini-shaped, then it is not oval'. Neither is knowable a priori. Shouldn't there then be a pair of worlds *u, v*, exactly the same in geometrical respects but such that *u → cassinis are oval*, while *v → cassinis are not oval*? These worlds threaten to show that there can be differences in respect of ovality without underlying geometrical differences.

Where the ovality argument goes wrong is easy to see. The problem is (2). You can't get a world where cassinis are not oval out of the fact that it's not a priori that they are oval. If our grasp of ovality were purely rationalistic, then the failure of apriority *would* call for a counter-world. But it isn't, so it doesn't.

The morality puzzle can be pinned on the same mistake. You can't get a world where N and ¬M out of the fact that it's not a priori that if N, then M. It would be different if our grasp of rightness were rationalistic; then we would have a genuine paradox on our hands. I conclude that it isn't rationalistic. A similar argument can be given for other evaluative concepts. None, I claim, are grasped rationalistically. None are grasped in what modal rationalists consider to be the one way in which a concept can be grasped.

19   Theoretical Predicates

Consider, finally, theoretical predicates: acid, energy, force, mass, species, cause, mereological sum, essential nature. What can be said about our understanding of these? Do we understand 'energy' by knowing a lot of conditionals of the

form 'had it turned out that *w*, such-and-such would have turned out to be the energy'?

Here are two arguments to the contrary, both harking back to the discussion of evaluative predicates. Suppose that we *do* (*qua* understanders of 'energy', etc.) know all these conditionals—that our concept of energy not only fixes, for each possible scenario, but discloses to us, for each of these scenarios, where the energy is to be found. How is it, then, that you and I continue to disagree about where the energy is to be found? (You say there is energy stored up in the curvature of space, while I deny it.) After all, there is a conditional known to both of us (as understanders of 'energy') that decides the matter. The explanation must lie in one of two places. It must be that

(i) someone is misconstruing the lower-level facts, and so picking the wrong conditional,

or:

(ii) someone is misconstruing their own mental states, specifically, the belief with that conditional as its content.

Whichever of these applies, our disagreement has the character of a misunderstanding. One or the other of us is laboring under a misimpression, and will (or should) change his or her tune when the mistake is pointed out. Of course, there is always the possibility that we associate *different* conditionals with 'energy'. In that case, though, we are not disagreeing at all; we mean different things by the word, so are talking past each other. None of the three scenarios allows for substantive disputes. Someone has made a mistake of type (i) or (ii), or else we are arguing over words.

This is almost as hard to accept here as it was in the evaluative case. Some disagreements *are* merely verbal, and some are based in correctable false impressions. The usual view, though, is that there's a third category: honest-to-God conflicts about what it is reasonable to believe, between people in command of the same lower-level facts. The effect of the rationalistic theory of grasp is to eliminate this third category.

The extension of 'energy' in a world is a function of what the correct scientific theory is. To find that theory, one must appeal at some point to considerations of naturalness, simplicity, nonarbitrariness, and the like—in a word, to considerations of *reasonableness*. (The positivists were the last to seriously question this.) Reasonableness is an evaluative concept and, as such, response-enabled. You can't hand responsibility over to 'rules of reasonableness'; there are no such rules, or at any rate not enough of them. You have to let yourself be led to some extent by your gut.

There are places where Chalmers sounds this theme himself. Figuring a concept's extension, he says, is not just grinding out conclusions. Judgment and discretion may be called for:

the decision about what a concept refers to in the actual world [may] involve[ ] a large amount of reflection about what is the most reasonable thing to say; as, for example, with questions about the reference of 'mass' when the actual world turned out to be one in which general relativity is true, or perhaps with questions about what qualifies as 'belief' in the actual world. Consideration of just what the primary intension picks out in various actual-world candidates may involve a corresponding amount of reflection. But this is not to say that the matter is not a priori: we have the ability to engage in this reasoning independently of how the world turns out. (1996: 58)

I suppose that we do have this ability. We can ask ourselves what is the most reasonable thing to say on various hypotheses about how the world turns out. It is not clear, though, how that argues for the matter's being a priori. We can also ask ourselves where the ovals are on various hypotheses about how the world turns out. Our conclusions in the second case aren't a priori, so why should they be a priori in the first?

If the oval example shows anything, it's that the move from 'we can tell independently of how things turn out' to 'we can tell a priori' is a *non sequitur*. For 'we can tell independently' may just mean that we can stage simulated confrontations with nature on various hypotheses about the form nature takes. It may not be obvious that searchers after the most reasonable hypothesis are doing this. But it seems to me that they are. Judgments of reasonableness and plausibility are arrived at by exercising a type of sensibility.

To be sure, the sensibility involved is not a perceptual one. And there seems less cause for worry about simulated plausibility judgments being a bad guide to real such judgments.[22] But the fact that sensibility is required should still give pause. It means that if you and I disagree about a sentence's truth-value in *w*, there may be no more we can say to each other than 'I find your position unreasonable'. The claim that everything but consciousness is a priori entailed by physics thus comes down to this: if two people disagree about a sentence's truth-value in *w*, each will find his or her own position to be the more reasonable one, unless the sentence is about consciousness, in which case each side concedes the rational defensibility of the other. Even if this were true, it is hard

---

[22] The moral case is arguably intermediate in these respects. Sensitivity to the moral aspects of things has often been likened to good vision or a keen sense of smell. And our horror at an observed case of, say, euthanasia or abortion may catch us by surprise, given our approving reaction to the imagined case. (Why else would right-to-lifers work so hard at getting us to *look* at what is being done?)

to see an argument for metaphysical dualism in it. And it is not true; the zombie hypothesis is *much less reasonable* than the hypothesis that what people seem to be feeling, they are feeling.

20   Logical Empiricism and Modal Rationalism

There were two dogmas of empiricism. One was the analytic/synthetic distinction. The other was 'semantic reductionism'—the idea that each statement is linked by fixed correspondence rules to a determinate range of confirming observations. Quine held that the two dogmas are 'at bottom the same'. For the correspondence rules are in a sense analytic. They give the sentence its meaning, so cannot fail as long as that meaning holds fixed. The dogmas are at least notionally different, though, and my focus will be on the second: the conception of correspondence rules as analytic, and therefore a priori. Although I will follow Quine in speaking mostly of analyticity, it is the apriority that is my real concern.

How is a modal rationalist like a logical empiricist? They seem initially very different. The empiricist has analytic correspondence rules connecting theory to experience. Modal rationalists aren't proposing anything like *that*. Yes, people have to be able to tell a priori whether S is true in a presented world. Gone, though, is any thought of that world being presented *in experiential terms*. There is no case, then, for a charge of *phenomenalistic* reductionism.

If one looks, though, at Carnap's writings on protocol sentences, it turns out that his sort of reductionism did not have to be terribly experiential either. Under the influence of Neurath, Carnap thinks that it is somewhat of an open question which sentences ought to be counted as protocols. Sometimes a protocol sentence is said to be any sentence 'belonging to the physicalistic system-language' which we are prepared to accept without further tests.[23] Often it is said to be a matter of *convention* which sentences will count as protocols. The important point for us is that Carnap thinks there are a priori rules connecting theoretical statements with protocols, whatever protocols turn out to be.

Another seeming difference emerges from Quine's complaint that Carnap overlooks the 'holistic nature of confirmation'. The complaint might be understood like this: One never knows whether S is really correct until all the observational evidence is in. Hence any rules portraying S as verifiable on the basis of limited courses of experience—courses of experience small enough to be enjoyable by particular observers—would be untrue to the way in which confirmation actually works.

[23] Ayer (1959: 237).

This complaint the rationalist can rightly claim to have answered. He never represents *partial* information as enough to ensure that S; the rules he contemplates take as input *complete* information:

[Quine says that] purported conceptual truths are always subject to revision in the face of sufficient empirical evidence. For instance, if evidence forces us to revise various background statements in a theory it is possible that a statement that once appeared to be conceptually true might turn out to be false.

This is so for many purported conceptual truths, but it does not apply to the supervenience conditionals that we are considering, which have the form 'If the low-level facts turn out like this, then the high-level facts will be like that.' The facts specified in the antecedent of this conditional effectively include all relevant empirical factors. . . . The very comprehensiveness of the antecedent ensures that empirical evidence is irrelevant to the conditional's truth-value. (Chalmers 1996: 55)

This is a good answer as far as it goes. But there are aspects of Quine's critique that it does not address. Quine says that

the dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion, issuing essentially from Carnap's doctrine of the physical world in the Aufbau, is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body. (1951: 41 in rpt.)

The problem here is not that S's confirmational status is underdetermined until all the empirical evidence is in. The problem is that S's confirmational status is not fully determined even by the full corpus E of empirical evidence. The degree to which E confirms S, Quine thinks, is tied up with the extent to which E or aspects of E are deducible from S. But nothing of an observational nature is deducible from S except with the help of a background theory T. Hence the degree of support that E lends to S depends on which background theory we use.

This complaint would be easily evadable if there were an analytically guaranteed fact of the matter about which theory E selects for. One could simply ask whether E supports S relative to the E-preferred theory, whatever it might be.

One has to assume, then, that this is what Quine is really concerned to deny. He denies that there are analytic connections between total corpuses E of empirical evidence and theories T of nature. Without these, there can be no analytic connections between E and particular statements S. A number of things suggest that analytic confirmation relations are indeed the target:

I am impressed, apart from prefabricated examples of black and white balls in an urn, with how baffling the problem has always been of arriving at any explicit theory of the empirical confirmation of a synthetic statement. (Quine 1951: 49 in rpt.)

This could be taken to mean just that the sought-after theory of confirmation would have to be very complicated. But Quine has something different in mind. He is aware, after all, of Carnap's attempts to work out a logic of confirmation which would tell us what to believe on the basis of given evidence. He is aware, too, that the attempt failed even for the simplest sort of examples. Carnap came up with a whole array of confirmation functions, none of them looking a priori better than the rest.

Where does this leave us? One problem with analytic confirmation relations concerns total evidence. This the rationalist has addressed. But there's a second problem: 'total science, mathematical and natural and human, is underdetermined by experience' (Quine 1951: 45 in rpt.). The version of underdetermination Quine needs is really a rather mild one. He needn't deny that there is an objectively best theory relative to a given body of evidence. He needn't even deny that there's a single most rational theory to adopt. All he need claim is that the choice between theories compatible with the evidence cannot be based just on our grasp of meaning. It 'turns on our vaguely pragmatic inclination to adjust one strand of the fabric of science rather than another. Conservatism figures in such choices, and so does the quest for simplicity'. (Quine 1951: 49 in rpt.).

This can be reconciled with the analytic view of confirmation relations only by supposing that my grasp of the language tells me how conservative I should be, and how important simplicity is, and how these sorts of desiderata trade off against one another. If two scientists judged the trade-offs differently, at most one could be considered to be speaking correctly—that is, in accordance with the meanings of her words. That, however, is not how the science game is played.

The interesting thing is that Carnap *agrees* that it's not how the science game is played. His goal, as he usually describes it, is not to uncover the true nature of meaning, but to give us tools for making our discursive practice more rational and efficient. He thinks that disputants should pick a common framework and then resolve their disagreements by reference to its assertion rules:

it is preferable to formulate the principle of empiricism not in the form of an assertion . . . but rather in the form of a proposal or requirement. As empiricists we require the language of science to be restricted in a certain way. (Carnap 1936–7: sect. 27)

Based on passages like this, one recent commentator has summarized the view as follows:

Criticisms of the meaning/belief distinction rest on the lack of a principled criterion for [semanticality]—no empirical method can be found for making it. However, for Carnap, such a distinction is to be reached by agreement in a conflict situation. Maximize agreement on framework issues and situate disagreement on either empirically answerable problems or on questions of a pragmatic nature about the framework. (O'Grady 1999: 1026)

One can argue about whether this would really be helpful. All I am saying right now is that not even Carnap believes that it is how we really operate: that our actual practice lends itself to a distinction between semantic factors in assertion and doxastic ones.

Is there anyone who does believe that this is how we operate? The modal rationalist does, or at least, such a view is not far from the surface. We are told that grasp of S's meaning, or at least the kind of grasp you need to count as understanding S, is knowing which worlds w are such that had this turned out to be w, it would have turned out that S. This applies not just to observation-level statements, but to theoretical statements as well. It is part and parcel of knowing T's meaning to know what the world would have had to be like for it to be the case that T. And that is not obviously different from Carnap's idea of analytic confirmation rules.

I say 'not obviously different', because there may be room for maneuver on the issue of what is involved in 'knowing which worlds are S-worlds'. I have been assuming that worlds are given in 'lower-level' terms, whatever exactly that might mean. What if worlds are described more fully than that, perhaps as fully as possible? There would be no need to infer that theory T applied; it would be given that it applied in the world's initial presentation. This seems tantamount to saying that one knows the S-worlds as, well, the S-worlds, or the worlds such that if they turned out actual, it would turn out that S.

But, if a 'homophonic' grasp of the set of verifying worlds were all one needed, then there would be no reason to expect a sentence to be knowable a priori just because its primary intension contained all worlds.

This is clear from Chalmers's discussion of physicalism. Consider again the conditional 'if PHYSICS, then PAIN'. It is claimed that the only way for this to be non-a priori is for there to be worlds not in its primary intension: there have got to be zombie worlds. If our grasp of primary intensions was homophonic, the failure of apriority would present no puzzle, hence no puzzle to which zombie-worlds might be offered as a solution. The reason I don't know a priori that if PHYSICS, then PAIN is that I can't tell a priori whether the primary intension of 'if PHYSICS, then PAIN' contains all worlds. I can't tell that because I can't tell a priori whether the PHYSICS worlds are a subset of the PAIN worlds. If they are a subset, there is no puzzle as to why the understander doesn't realize it, because it is assumed from the outset that PHYSICS worlds are, for all she knows a priori, worlds without PAIN.

How, then, are worlds presented to the meaning-grasper? She must be able to pick out the S-worlds on the basis of their ground-level properties. 'If the low-level facts turn out like this, then the high-level facts will be like that' (Chalmers 1996: 55). These conditionals are thought to be analytic; indeed, they are true in virtue of the aspect of meaning to which we have a priori access. This is why I say that modal rationalists are committed to something *like* the analytic confirmation relations advocated by Carnap and rejected by Quine. The rationalist who wants to escape Quine's criticisms has got to (a) show that the criticisms don't work even against logical empiricism; (b) show that the cases are relevantly different.

To accomplish (a) would be to find a mistake in Quine's reasoning. Maybe, for example, it's just untrue that theory is underdetermined by evidence. To accomplish (b) would be to show that what the modal rationalist says is different enough from what the logical empiricist says that the Quinean critique doesn't generalize. Maybe, for example, the lower-level facts on the basis of which we can tell a priori whether S are quite unlike the 'empirical' facts on the basis of which we can't tell a priori whether S. I won't pursue the matter any further here, but I suspect that the prospects for doing either of these things are not terribly good.

21 Digression: Imaginative Resistance

Hume, in 'The Standard of Taste', points out something surprising about our reactions to imagined circumstances. Reading a story according to which S, I try to imagine myself in a situation where S really holds. The surprising thing is that we can do this quite easily if S is contrary to *descriptive* fact, but have a great deal of trouble if S is contrary to *evaluative* fact. Reading that Franco drank from the Fountain of Youth and was made young again, you don't blink twice. But reading that it was good that little Billy was starved to death since he had, after all, forgotten to feed the dog, you want to say, 'it was *not* good, I won't go along.'

Call that *imaginative resistance*.[24] Why does it happen? A number of explanations have been tried. Do we resist because what we're asked to imagine is conceptually false? No, because (i) counter-moral hypotheses are *not* conceptually false (remember essential contestability), and (ii) lots of conceptually false scenarios are *not* resisted (as readers of Calvino and Borges will attest).

[24] On imaginative resistance, see Gendler (2000), Moran (1989), and Walton (1994).

Do we resist because what we're asked to imagine is morally repugnant? No, because we balk at aesthetic misinformation as well. 'All eyes were on the twin Chevy 4×4's as they pushed purposefully through the mud. Expectations were high; last year's blood bath death match of doom had been exhilarating and profound, and this year's promised to be even better. The crowd went quiet as special musical guests ZZ Top began to lay down their sonorous rhythms. The scene was marred only by the awkwardly setting sun.' Reading this, one thinks, 'If the author wants to stage a monster truck rally at sunset, that's up to her. But the sunset's aesthetic properties are not up to her; nor are we willing to take her word for it that last year's blood bath death match of doom was a thing of beauty.'[25]

Do we resist because the scenario is repugnant along *some evaluative dimension or other*? No, because it is not only evaluative suggestions that are resisted. You open a children's book and read as follows: 'They flopped down beneath the great maple. One more item to find, and yet the game seemed lost. Hang on, Sally said. It's staring us in the face. This is a *maple* tree we're under. She grabbed a five-fingered leaf. Here was the oval they needed! They ran off to claim their prize.' Reading this one thinks, 'If the author wants it to be a maple leaf, that's her prerogative. But the leaf's physical properties having been settled, whether it is oval is not up to her. She can, perhaps, arrange for it not to have the expected mapley shape. But if it does have the expected shape, then there is not a whole lot she can do to get us to imagine it as oval.'

Imaginative resistance arises not only with evaluative predicates, but also with (certain) descriptive ones: 'oval', 'aquiline', 'jagged', 'smooth', 'lilting'. What do these predicates have in common? P makes for imaginative resistance if, and because, the concept it expresses is of the type I have called 'grokking', or response-enabled.[26]

Why should resistance and grokkingness be connected in this way? It's a feature of grokking concepts that their extension in a situation depends on how the situation does or would strike us.[27] 'Does or would strike us' *as we are*: how we are represented as reacting, or invited to react, has nothing to do with it. Resistance is the natural consequence. If we insist on judging the extension ourselves, it stands to reason that any seeming intelligence coming from elsewhere is automatically suspect. This applies in particular to being 'told' about the extension by an as-if knowledgeable narrator.

[25] She knows this, moreover. Why make a suggestion you know will not be accepted? There might be any number of reasons, but most likely she is just pulling our leg.

[26] [For an alternative diagnosis of this phenomenon, see Currie, Ch. 4 above—eds.]

[27] I assume that fictional situations are presented as counteractual, not counterfactual. One is to think of them as really happening.

22 (Conceptually) Contingent A Priori

I have called a lot of claims a priori. But not much has been done to explicate the notion; the focus has been more on conceptual necessity. I doubt that it is possible to explain apriority in all its guises with the materials at hand. But I'll try in the next few sections to clarify a particular type of apriority as far as I can. (Nothing argued so far depends on what is coming next.)

'Water contains hydrogen' is touted in *Naming and Necessity* as an example of an a posteriori metaphysical necessity. 'Cassinis are oval' has been touted here as an example of an a posteriori *conceptual* necessity. A posteriori conceptual necessities are the counterpart in our system of the a posteriori metaphysical necessities that Kripke emphasized.

One might wonder whether we have anything to correspond to Kripke's *other* famous category: the category of a priori but (metaphysically) contingent truths like 'Neptune is the planet if any responsible for . . .'.

I suggested above that 'unless we are greatly misled about the circumstances of perception, a figure is oval iff it looks egg-shaped' was a priori, or close enough for present purposes. But of course things *could* have turned out so that we were unable to see eggs in oval figures. Things could have turned out so that we never saw anything as egg-shaped.

Had things turned out so that nothing looked egg-shaped, would the world have turned out to be oval-free? The answer seems clear. How we see things is irrelevant to how they are shaped. It would have turned out that there were ovals which, however, did not look the way ovals are supposed to look.

I make no prediction about what we would have *said*. It may be that we would have said 'there are no ovals'. That is irrelevant unless the meaning that 'oval' would have turned out to have in that circumstance is the meaning it has actually. And it seems clear that the meanings are different. If people say 'there are no ovals' in a world geometrically just like ours, they do not mean the same thing by 'oval' as we do.

'Unless . . . , a figure is oval iff it looks egg-shaped' is an a priori but conceptually contingent truth. It could have turned out that we were not prone to see ovals as egg-shaped, perhaps because we were not prone to see anything as egg-shaped. And, approaching it from the other end, it could have turned out that almost-circular figures looked to us egg-shaped, despite not being oval.

This seems at first puzzling: how can it be a priori that 'oval iff looks egg-shaped' when it could have turned out otherwise? One has to remember that the scenario where it turns out otherwise is *also* a scenario where it turns out that 'oval' doesn't mean what we all know it does mean. A scenario in

which 'oval' changes meaning can no more stop 'oval iff looks egg-shaped' from being a priori than one in which '=' means nonidentity can stop 'Phosphorus = Phosphorus' from being a priori.

23 Apriority versus Conceptual Necessity

I said that 'oval' could have turned out not to mean what we all know it does mean. What we all know it does mean is *oval*. So I could equally have said that it could have turned out that 'oval' did not mean oval. I do not shrink from this way of putting it, or even the claim that it could turn out (though it won't) that 'oval' *doesn't* mean oval.

I admit, however, that these claims sound funny. If we accept that 'oval' could have turned out not to mean oval, then it seems like we should regard as not completely insane someone (Crazy Eddie) who says that 'oval' *doesn't* mean oval. He could turn out to be right! Intuitively, though, there is no chance whatever of Crazy Eddie's turning out to be right.

What does it take for Crazy Eddie to be vindicated? It is not enough that, letting S be the sentence he uttered, it could have turned out that S. The scenario in which it turns out that S could be a scenario in which S has changed meaning. You are not vindicated unless what you said turns out to be right; it's not enough that what you turn out to have said turns out to be right. Otherwise Warrenites would be vindicated if 'Oswald acted alone' turned out to mean that Oswald had help, and he did. There is no danger of Crazy Eddie turning out to be right, because, letting M be the (actual) meaning of his words, had it turned out that M, it would have turned out that M was not what he said!

I assume that 'it could turn out that . . .' is an intensional context—that is, a context treating synonyms alike. Since 'sister' is synonymous with 'female sibling', and it could turn out (though it won't!) that 'sister' does not mean female sibling, it could turn out that 'sister' does not mean sister. The reason why it sounds funny to say it is that the statement strongly *suggests* something absurd: namely, that someone who conjectures that 'sister' doesn't mean sister could turn out to be right.

Another (not incompatible!) way to explain the funniness is this. There is a use of 'it could turn out that S' on which it means that it is not a priori that ¬S. In that (alternative) sense of the phrase, it really *couldn't* have turned out that 'sister' didn't mean sister. For we know a priori that 'sister' means sister. If it doesn't sound as bad to say that 'sister' could turn out not to mean female sibling, that might be because we don't know a priori that it does mean female sibling.

Compared to conceptual necessity, apriority is an elusive notion. One reason has already been noted. If it is a priori that 'sister' means sister, but not that it means female sibling, then 'it is a priori that . . .' is not an intensional context; it cares about the difference between synonyms. ('It could have turned out that . . .' (in the alternative epistemic sense) is therefore not intensional either.)

Stranger even than the failure of intensionality is the following. The class of a priori truths is often claimed to be closed under (obvious) logical consequence. This can't be right, if a well-known account of apriority is even roughly correct. It is a priori that S, according to the well-known account, if one can know that S is true just on the basis of one's grasp of S's meaning. Suppose I know that A and that A ⊃ B just through my grasp of the two sentences' meanings, and then I infer B. If this is my reason for believing B, then I do not know it a priori. For my belief is based in part on my grasp of A's meaning, and A is a different sentence from B.

The failure of logical closure helps us resolve a puzzle. There are many things I know a priori. For instance, I know a priori that sisters are sisters, and that Hesperus = Hesperus. If 'S' is a sentence I understand, then I would seem to know a priori that 'S' is true iff S.[28] (More on this claim below.)

But I rarely, if ever, know a priori that a sentence 'S' is true; for truth-value depends on meaning, and my knowledge of meaning is a posteriori. I have to learn what a sentence means, even a sentence of my own idiolect. And my views on the topic are rationally defeasible under the impact of further evidence.

The question is, why can't I combine my a priori knowledge that sisters are siblings with my a priori knowledge that if they are siblings, then 'sisters are siblings' is true, to arrive at a priori knowledge that 'sisters are siblings' is true?

The problem is not that I can't *modus ponens* my way to the conclusion that 'S' is true, starting from premises known a priori. The problem is that, having done so, it is not just in virtue of understanding ' "S" is true' that I know that 'S' is true. The understanding I have of 'S' plays a role too, and that is something over and above my understanding of ' "S" is true'. (I can understand the latter while momentarily forgetting what 'S' means, or while entertaining a skeptical hypothesis to the effect that it means something other than I had thought.) Since I cannot claim to know that 'S' is true just in virtue of my understanding of that very sentence, I cannot claim to know a priori that 'S' is true.

---

[28] Notice the quotation marks. Use/mention distinctions that had been left to context are here marked explicitly.

24 Apriority

What can we say about apriority to explain these puzzling features? Since apriority is a matter of what my grasp of a sentence's meaning tells me, our account has got to bring in grasp explicitly. What aspect of grasp could function to tell me that the sentence is true? A state that tells me something is a state whereby I possess information. So our account should be in terms of the information I possess whereby I grasp meaning. Call this my *grasp-constituting information* about 'S'. The proposal is that

(AP) it is a priori (for me) that S iff for some G

(a) that 'S' is G is part of my grasp-constituting information,

and

(b) being G conceptually necessitates being true.

Let's revisit some earlier questions with (AP) in hand.

How can it be a priori that 'sister' means sister yet not a priori that it means female sibling? That 'sister' means sister is part (all?) of the information whereby I grasp 'sister'. I do of course realize 'on the side' that to be a sister is none other than to be a female sibling. But that is a collateral belief which does not figure in my grasp. Suppose the belief changed in response to some *outré* counter-example; that would be a change in what I thought sisters were, but not a change in what I meant by 'sister'.

Why are the a priori truths not closed under logical consequence? Having deduced B from A and A ⊃ B, I am in possession of information given which B has to be true. But there is no reason to expect the information to be grasp-constituting with respect to B; on the contrary, the information by which I grasp A is likely to be involved. To know B a priori, I need to know it on the basis of the information whereby I grasp B.

How can an a priori truth fail to be conceptually necessary? The information G that conceptually necessitates that 'S' is true might not be conceptually necessary information. If 'S' has a conceptually contingent property that conceptually necessitates that 'S' is true, all I can conclude about 'S' truth-wise is that it *is* true given how matters actually stand. Conceptual necessity requires more than this: 'S' must be true on *any* hypothesis about how matters stand, including the false ones.

Example: I am newly arrived in the royal court. A helpful attendant explains that 'the king' is to be understood so that 'the king is the guy giving orders, wearing the crown, and so forth' comes out true. I come as a result to know a priori that the king is the guy giving orders, and the rest. Now, as a matter of

fact, it is Richard who is doing all these things; as a matter of fact, it is Richard who is the king. But things *could* have turned out so that it was an impostor Richerd who was giving orders, and so on. Would the king then have turned out to be Richerd?

I have certainly been given no reason to think so. I was told that 'the king' stood for the order-giver by someone who supposed (correctly) that the order-giver was Richard. They leaned on that supposition in defining 'the king' as the order-giver. Leaning on a supposition that they knew could turn out to be false, they were careful *not* to say that the king would still have been the order-giver however things had turned out. And indeed, he wouldn't: things could have turned out so that the king was Richard, while the order-giver was Richerd.[29] It is conceptually contingent that the king = the order-giver. Still, I know it a priori.

Why are some conceptually necessary truths not a priori? Sometimes the information that a speaker possesses about 'S' whereby she grasps its meaning is information that exhibits 'S' as true. Other times, it isn't. I am not sure what a typical understanding of 'cassinis are oval' involves, but one is not expected to realize that it is true. You should perhaps know that things looking egg-shaped are to be counted oval. But that doesn't enable you to work out that cassinis are oval until you've laid eyes on one.

If 'sisters are siblings' can turn out not to be true, yet sisters cannot turn out not to be siblings, then in some counteractual world sisters are siblings and 'sisters are siblings' is untrue. Why isn't this a world in which the T-biconditional fails? It *is* a world where the T-biconditional fails. It could have turned out that 'sisters are siblings' is untrue although sisters are siblings. This seems odd until we remember that it can happen only if 'sisters are siblings' turns out not to mean what it does mean. A world where it turns out not to mean what it does mean is a world where my grasp-making information fails. A world where that information fails is irrelevant to the issue of whether that information entails the truth of the

---

[29] A related sort of presupposition is discussed by Putnam: 'Suppose I point to a glass of water and say "this liquid is called water". . . . My "ostensive definition" of water has the following empirical presupposition: that the body of liquid I am pointing to bears a certain sameness relation to . . . most of the stuff I and other speakers in my linguistic community have on other occasions called "water." If this presupposition is false because, say, I am without knowing it pointing to a glass of gin and not a glass of water, then I do not intend my ostensive definition to be accepted. Thus the ostensive definition conveys what might be called a defeasible necessary and sufficient condition. . . . If it is not satisfied, then one of a series of, so to speak, "fallback" conditions becomes activated' (1975: 225). I would add only that the series tends to be a finite one. Some defeats you recover from; an (itself defeasible) backup condition kicks in. Eventually, though, the backups are exhausted, and the definition just fizzles.

---

biconditional—and so to the issue of whether it holds a priori that 'sisters are siblings' is true iff sisters are siblings.

Why does the feeling persist that if it is not a priori that S, there is a counteractual world in which ¬S? There is an argument to this effect that almost works. If there are no counteractual worlds in which ¬S, then every counteractual world is an S-world. A fact like that surely figures in the information whereby we understand 'S'.[30] The fact entails that 'S' is true, and so grasp-making information entails that 'S' is true, and so S is a priori. Contraposing, if S is not a priori, then it does not hold in all counteractual worlds. But, the sentence beginning 'a fact like that surely figures' assumes our grasp is rationalistic. The feeling persists because we forget that there are other ways to understand.

REFERENCES

Adams, E. W. (1975), *The Logic of Conditionals* (Dordrecht: D. Reidel).

Almog, Joseph, Perry, John, and Wettstein, Howard (1989) (eds.), *Themes from Kaplan* (New York: Oxford University Press).

Ayer, A. J. (1959) (ed.), *Logical Positivism* (New York: Free Press).

Carnap, Rudolf (1936–7), 'Testability and Meaning', *Philosophy of Science*, 3: 419–71; 4: 1–40.

Chalmers, David (1994), 'The Components of Content', Philosophy/Neuroscience/Psychology Technical Report 94-04, Washington University, http://www.u.arizona.edu/~chalmers/papers/content.html.

—— (1996), *The Conscious Mind: In Search of a Fundamental Theory* (New York: Oxford University Press).

—— (2000), 'The Tyranny of the Subjunctive', http://www.u.arizona.edu/~chalmers/papers/tyranny.txt.

Davies, Martin, and Humberstone, Lloyd (1980), 'Two Notions of Necessity', *Philosophical Studies*, 38: 1–30.

DeRose, Keith (1991), 'Epistemic Possibilities', *Philosophical Review*, 100: 581–605.

Evans, Gareth (1979), 'Reference and Contingency', *Monist*, 62: 161–89.

Gendler, Tamar Szabó (2000), 'The Puzzle of Imaginative Resistance', *Journal of Philosophy*, 97(2): 55–81.

Grice, H. P. (1989), 'Indicative Conditionals', in *Studies in the Way of Words* (Cambridge, Mass.: Harvard University Press).

Jackson, Frank (1979), 'On Assertion and Indicative Conditionals', *Philosophical Review*, 88: 565–89; repr. in Jackson (1991), 111–35.

—— (1991), *Conditionals* (Oxford: Oxford University Press).

---

[30] This is a bit of an exaggeration, since knowing of each w that $w \rightarrow S$ is not yet knowing that $w \rightarrow S$ for all w.

Jackson, Frank (1994), 'Armchair Metaphysics', in Michaelis Michael and John O'Leary-Hawthorne (eds.), *Philosophy in Mind* (Dordrecht: Kluwer), 23–42.

—— (1998), *From Metaphysics to Ethics: A Defense of Conceptual Analysis* (Oxford: Clarendon Press).

Kripke, Saul (1980), *Naming and Necessity* (Cambridge: Harvard University Press).

Loar, Brian (1990), 'Phenomenal States', *Philosophical Perspectives*, 4: 81–108.

McGinn, Colin (1999), *Knowledge and Reality* (Oxford: Clarendon Press).

Moran, Richard (1989), 'Seeing and Believing: Metaphor, Image, and Force', *Critical Inquiry*, 16: 87–112.

O'Grady, Paul (1999), 'Carnap and Two Dogmas of Empiricism', *Philosophy and Phenomenological Research*, 49: 1015–27.

Peacocke, Christopher (1989), 'Perceptual Content', in Almog *et al.* (1989), 297–330.

Putnam, Hilary (1975), 'The Meaning of "Meaning" ', in *Mind, Language, and Reality* (Cambridge: Cambridge University Press), 215–71.

Quine, Willard van Orman (1951), 'Two Dogmas of Empiricism', *Philosophical Review*, 60: 20–43; repr. in Quine (1961), 20–46.

—— (1961), *From a Logical Point of View*, 2nd edn. (New York: Harper & Row).

Segerberg, Krister (1972), 'Two Dimensional Modal Logic', *Journal of Philosophical Logic*, 2: 77–96.

Stalnaker, Robert (1972), 'Assertion', in Peter Cole (ed.), *Syntax and Semantics*, ix (New York: Academic Press), 315–32.

—— (1990), 'Narrow Content', in C. A. Anderson and J. Owens (eds.), *Propositional Attitudes* (Stanford, Calif.: Center for the Study of Language and Information).

—— (1991), 'How to Do Semantics for the Language of Thought', in Barry Loewer and Georges Rey (eds.), *Meaning in Mind: Fodor and his Critics* (Oxford: Blackwell).

Walton, Kendall (1994), 'Morals in Fiction and Fictional Morality', *Proceedings of the Aristotelian Society*, supp. vol. 68: 27–50.

White, Stephen (1982), 'Partial Character and the Language of Thought', *Pacific Philosophical Quarterly*, 63: 347–65.

INDEX