Coulda, Woulda, Shoulda

S. Yablo

draft of March 22, 2001

1. Terminology

A main theme of Kripke's Naming and Necessity 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 non-technical 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, 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" (122)). He notes, however, that one might equally let the word express loose analyticity, and that on that definition "some analytic truths are contingent" (122).

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. Some might want to distinguish an apriority-entailing kind of analyticity from a kind that does not entail apriority. 'Hesperus is Phosphorus' is not apriori, they might say, but since its meaning is a proposition of the form x=x, and any proposition of that form is true, it should be counted true in virtue of meaning. (I am not saying they would be right!)

Kripke seems, however, just to take it for granted that analytic truths will be apriori knowable. In his characterization of loose analyticity he speaks, not of statements whose truth is guaranteed by reference-fixing descriptions, but ones whose "apriori truth is known via the fixing of a reference" (122, italics added). A non-Kripkean line on the apriority of analytic sentences will be defended 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, that something is epistemically possible is naturally taken to mean that it is true (or possible) for all one knows. A notion explicating apriority/analyticity should not be so sensitive to one's state of

knowledge. One doesn't know how to prove Goldbach's conjecture today, but one might tomorrow; if that happens we might well say that it had been necessary (in the partner sense) all along.

"Semantic necessity" too is liable to mislead, since for some people, Kripke included, 'Hesperus' and 'Phosphorus' have the same meaning, yet it is possible in the partner sense that Hesperus ≠ Phosphorus. As Kripke says, 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 examples do seem to fit with the idea of sentences that have got to be true because of 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 vs. conceptual distinction. Yet, and I think this is agreed by everyone, the distinction remains not terribly well understood. It is not well understood because 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 necessary to mention lest it obscure our view of metaphysical necessity. Certainly this is the view many people would take about the conceptual notion today. A lot of the contemporary skepticism

about narrow content is at the same time skepticism about conceptual possibility. Narrow content if it existed would give meaning 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 it's not clear there are any. 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. Loosely quoting from Lewis: Whoever claims not to understand something will take care not to understand anything else whereby it might be explained. If one doesn't understand narrow content, one 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." One suspects that the order here is not accidental. And even if that suspicion is wrong (as he says it is), the primary intension is certainly a partner in full standing.

This paper tries not to take sides between the skeptics and the believers. Our topic is how conceptual possibility should be handled supposing it is going to be handled at all. If we do slip occasionally into the language of the believers, that is because we are trying to explore their system from the inside, in order to see what it is capable of, and whether properly developed it delivers 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 closed.)

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 w should be such that S gives a correct description of it; every world w should be such that the way S describes things as being is a way that w in fact is.

Conceptual possibility too is explained in terms of worlds To be conceptually possible is to be (in some appropriate sense) true-with-respect-to-w for at least one world w. Our job is to specify the sense. Conceptual modality is supposed to be different; and everyone knows of the famous examples that are supposed to bring out how, e.g., the example of Hesperus being "possibly" distinct from Phosphorus in the one sense but not the other. But the nature of the difference hangs on the contrast between true-of-w as just discussed, and the notion of true-w.r.t.-w that we must now attempt to develop.

Here is the obvious first stab: S is true w.r.t. 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; the same sentence uttered in w might mean that Mars isn't Saturn. If, in w, Mars indeed isn't Saturn, then 'Hesperus ≠ Phosphorus' is true w.r.t. 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, 'Hesperus ≠ Hesperus.' Uttered in w, this means that Mars ≠ Mars. Since that is not true of Mars, in w or elsewhere, w does not testify to the conceptual possibility of Hesperus not being Hesperus. If there are no worlds where uttering 'Hesperus ≠

Hesperus' is speaking the truth, then it is not conceptually possible that Hesperus ≠ Hesperus.

Trouble is, 'Hesperus ≠ Hesperus' is true uttered in some worlds. For there are worlds v in which 'Hesperus ≠ Hesperus' means something quite different from what it actually means (say, that Hesperus is identical to Hesperus) and in which the quite different thing is true. This is clearly the wrong result. It shouldn't count make ' Hesperus ≠ Hesperus' conceptually possible that there are worlds in which '≠' expresses identity!

This is a point Kripke lays great stress on, in his discussion of metaphysical possibility. Let it be, he says, that w contains speakers (maybe counterfactual versions of ourselves) 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 all been speaking German" or "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 (NN, 77).

So, by 'tail' speakers in w might mean wing. If so, then assuming w's horses resemble ours, w-people speak falsely when they say 'horses have tails.' That's irrelevant, Kripke says, to the metaphysical necessity issue; 'horses have tails' is as true of w as it is true 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" (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.

4. Holding Fixed

First try: S has got to mean the very same in w as it means here. This would be asking too much. After all, 'Hesperus' and 'Phosphorus' as they are used here both mean Venus, and '≠' stands for 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. Yet we want it to come out conceptually possible that Hesperus ≠ Phosphorus.

Second try: corresponding expressions should either mean the same, or have their references fixed by the same or synonymous descriptions. This is better but still not enough. 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 descriptions guaranteed in advance to pick out

the referent.) So the second proposal reduces in most cases to the first, which we've seen to be inadequate.

A third approach stipulates that w should be an "epistemic counterpart" of actuality, in the sense of confronting the speaker with the same "evidential situation." Any variations in S's meaning will then be subliminal. If meanings change at most subliminally in w, then S's truth as uttered in w suffices for its conceptual possibility here.

On the plus side, this proposal no longer attempts to specify the relevant aspects of meaning -- the ones that are supposed to be held fixed -- explicitly. The idea is that we can catch all of that by holding fixed the entire evidential situation, including presumably the speaker's semantically relevant memories. The problem is that mixed in with the relevant aspects of meaning will be non-semantic circumstances that should be allowed to vary. If the fact that there seems to be lectern present is held fixed, then it will come out conceptually necessary that there seems to be a lectern present. Why should it make S necessary that it describes part of the evidential situation?

5. Subjunctives

Conceptual necessity is left by Kripke in a more precarious state than metaphysical. Judging conceptual necessity means considering 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 the metaphysical notion?

The usual answer is that with metaphysical necessity, we needn't bring in a counterfactual utterance at all. All we need ask is 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' w.r.t. w. Jones maintains, reasonably enough, that what 'horses have tails' actually says is that tails are had by Northern Dancer, Secretariat, etc. (fill in here the list of all actual horses). She concludes that 'horses have tails' is true of w iff Northern Dancer, Secretariat, etc. (or perhaps just those of them that exist in w) have tails in w.

Smith 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. They accordingly disagree 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 a given sentence in a given context. But it would be with a heavy heart, and not only because the notion of what-is-said is vague and context-sensitive. What semantic question cannot be framed as a question about what some S expresses in some context? It would be nice if we didn't have to complete the project of semantics before the truth conditions of 'necessarily S' could be explained.

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 it might go:

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 you're right, a story would be needed about how it is determined what is said. But the phraseology is there only for heuristic purposes. It is there to remind us that it doesn't matter, in considering whether S is true of w, what the citizens of w say with S.

Of course we can't let the Kripkean off the hook yet. If "saying what it actually says" is just a heuristic, then how is true-of-w to be explained?

> One option would be to treat "true of" as a primitive. But that option is problematic. It gives the skeptic about metaphysical possibility too big an opening. (She can claim to find the primitive incomprehensible.) It seems important to try to explain "true of w" using form of words the skeptic qua English speaker already understands. This can be done using the subjunctive conditional. To say that S is true of a world is just to say that had that world obtained, it would have been that S .[1]

Consider in this light the "controversy" about horses and their tails. When it comes to evaluating 'horses have tails' with respect to a world w, is it only the (actual) horses that matter, or do the additional horses that would have existed had w obtained have to be taken into account as well? That is, suppose that although actual horses have tails, w contains in addition to the actual horses some "extra" horses which are tailless. Is 'horses have tails' true of w?

The subjunctive account makes short work of this conundrum. What we need to ask ourselves is whether it would have been that horses had tails, had w obtained. The answer is clear. 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.

Consider next the case of a world w with speakers who by 'tail' mean wing. Does the fact that the w-people speak falsely when they say 'horses have tails' show that 'horses have tails' is false of w? No, and we can now see why. The w-speakers' irrelevance follows automatically from the subjunctive explanation of relative truth. The question is whether horses would have still have had tails, even if people had used 'tail' to mean wing. Of course, they would have; how people talk doesn't change the anatomy of horses. Had people spoken differently, they might not have uttered 'horses have tails,' but horses would still have had tails.

6. Disparity

All of this is to emphasize the disparity, in the immediate aftermath of Naming and Necessity, between metaphysical and conceptual necessity. Metaphysical necessity was in pretty good shape; because it went with "true of w," and "true of w" could be explained using "it would have been that S, had it been that w." Conceptual necessity was in comparatively bad shape; because it went with "it would have been that S was true, had it been that w, and certain facts about S's meaning held fixed ." And no satisfactorily general story existed about which facts were to be held fixed.

Then a brainstorm was had that seemed to restore parity.[2]

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's true (as spoken) at w. But maybe it wasn't really necessary to "move S over" to counterfactual world w. A different option is to "move w over" to actuality: to the place where the token of S that we want evaluated actually occurs.[3]

All right, but how do we move w over to actuality? It looks at first very simple. Just as, when judging metaphysical necessity, we consider w as counterfactual, 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; a good part of cognitive life consists of working out what is the case on hypotheses that we reject. Evaluating S with respect to w construed as actual is asking whether S is the case on the hypothesis that w is (contrary to what we perhaps think) this very world.

For example: it is conceptually possible that Hesperus ≠ Phosphorus because, if we hypothesize for the moment that this world is one in which, contrary to what we had thought, Hesperus-appearances are due to Mars, Phosphorus-appearances to Venus, then clearly (on that hypothesis) we were/are wrong to think that Hesperus=Phosphorus. It is not that counterfactual people are wrong about their world. It's 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 still there.

I said above that everyone would (should!) have been quite unhappy if they had been asked to treat "true of counterfactual world w" as a semantic primitive. Our willingness to rest so much on "true of" was thanks to the ordinary-language explanation we had 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 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 w.r.t w considered as

counteractual. If by analogy with "true of" as an expression for truth with respect to a counterfactual world, we use "true if" for truth with respect to a world conceived counteractually, then the problem is this: we have an explanation of "true of" but nothing similar for "true if."

7. Indicatives

One proposal suggests itself immediately. Given that "true of" goes with a certain counterfactual conditional, it stands to reason that "true if" would go with the corresponding indicative conditional. S is true if w just on the condition that if w in fact obtains (evidence to the contrary notwithstanding), then S.[4]

This approach has the nice feature of linking two deep distinctions: metaphysical vs. conceptual necessity, on the one hand, and subjunctive vs. indicative conditionality on the other. The reason it is metaphysically but not conceptually necessary that Hesperus=Phosphorus is that there are worlds w such that, although Hesperus would have been Phosphorus had w obtained, it isn't Phosphorus if w obtains.

But do the two conditionals really "predict" the two types of necessity? It helps to remember how we got here. It was important for metaphysical necessity to keep what-is-said fixed as we evaluate S w.r.t. w. Subjunctives are valued because they in effect do this, without dragging us into deep and controversial issues about what is said. When we switch to conceptual necessity, it stops being important 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. But we do not want S's meaning to be "changeable" in all respects, or nothing will come out conceptually necessary. The attraction of indicatives was that they

seemed to deliver an appropriate measure of meaning-fixation -- just as counterfactuals did on the metaphysical side .

Indicatives seem to deliver an appropriate measure of meaning-fixation. But when you look a little closer, you see that they don't 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 appropriately meaning-altering 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, we were just been confused on the point. 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. [5]

You may say: why should it be a problem if there are counteractual worlds in which horses lack tails? That is not the problem. The problem is that there worlds where horses lack tails not for anatomical reasons but on account of 'tail' not meaning tail. If horses can lose their tails like that, then take any S you like, it is true in some worlds and false in others. This spells disaster for the attempt to explain conceptual possibility in indicative terms. One doesn't want 'Hesperus ≠ Hesperus' to be conceptually possible just because there are worlds w where people mean identical by '≠'.

8. Narrow Content

The indicative is not the conditional we want. But it is close. The conditional we are after should be like the indicative except in one crucial respect: its consequent should be protected from a certain sort of meaning-shift brought on by the antecedent.

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 'Hesperus ≠ Hesperus' on the supposition that '≠' expresses nonidentity. But to identify the desired sort of conditional, we'll need to proceed beyond the example stage.

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 assuming w as it is absent that assumption. The narrow content of 'horses have tails' if 'tail' means wing is not the same as its narrow content given that 'tail' means tail. Calling that actual narrow content NC, the result we want is that 'tail' means wing ➔ 'horses have tails' (still) means NC.

But, although helpful as an intuitive constraint, this doesn't solve our problem. This is partly because, well, it just doesn't; no explicit interpretation of 'A➔ B' has been put forward.. Second, though, to appeal to narrow content in this context gets things the wrong way around. Remember why we became interested in "S is true if w" in the first place. It was to get a better handle on conceptual necessity. But conceptual necessity is, more or less, necessity in virtue of narrow content. So narrow content is, if anything, part of what we wanted explained in terms of "true if w." To appeal to it in our account of "true if w" would get things the wrong way around.

9. Turning Out

Our problem now is similar to one faced earlier in connection with metaphysical necessity. It seemed that in explaining truth-of we would have to make use of the notion of what is said by someone uttering S. That would reverse the intended order of explanation.

The what-is-said of an utterance -- its broad content -- is given by the worlds of which the sentence is true. Metaphysical necessity is the special case in which S's broad content holds in all worlds, or, as some would have it, is (the set of) all worlds. That is why we don't want to be relying on broad content in explaining what is involved in a sentence's being true of a world. Our current worry is the same except that it involves "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, and stipulating that S is true of w iff had it been that w, it would have been that S. The beauty of this construction is that it automatically focuses on the agreement or not between w and S's broad content. Can a construction be found that automatically focuses on the agreement or not between w and S's narrow content, as the subjunctive does for broad?

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

It's 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. That the inclination persists, even among practicing modal metaphysicians, suggests that there is a bona fide difference between "could have been" and "could have turned out" that we should try if at all possible to respect in our semantics.

It suggests it to me, anyway. Kripke apparently does not agree. He maintains that that the second inclination -- the inclination to say it could have turned out that Hesperus was distinct from Phosphorus -- 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. Which 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."

If this were the only available interpretation, then it seems a mystery why the second inclination above should have outlasted the first -- why we should have persisted in the idea that it could have turned out that Hesperus wasn't Phosphorus even after giving up the idea that it couldn't have been that Hesperus wasn't Phosphorus.

So, proposal: on the most natural construal of the phrase, it could indeed have turned out that Hesperus ≠ Phosphorus. That indeed is what would have turned out had it turned out that Phosphorus-appearances were appearances not of Venus but of Mars. By contrast, it could not have turned out that Hesperus ≠ Hesperus. Had it turned out that '≠' meant identical, it would have turned out that the sentence we use to express that Hesperus ≠ Hesperus was true; but it would not have turned out that Hesperus ≠ Hesperus.[6]

10. Conceptual Possibility

I have no theory of "could/would have turned out" to offer. The attraction of the phrase is that it's a piece of ordinary English that seems to work in the desired way.

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

If 'tail' means wing, we said, then horses lack tails. But we don't want the fact that 'tail' means wing in w to have the consequence that w obtains ➔ horses lack tails. 'Tail's meaning wing in w won't have that consequence -- it won't make it the case that w ➔ horses lack tails -- 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. Unless our reasons for thinking that horses have tails are metalinguistic in nature, revelations about the meanings of words cannot undermine them.

To come at ➔ from the other direction, we said that we don't want our conditional to leave broad content alone. For we want there to be worlds w such that w obtains ➔ Hesperus ≠ Phosphorus. And that can't happen unless the broad content of 'Hesperus ≠ Phosphorus' is allowed to change under the impact of the hypothesis that w obtains. Here too "would have turned out" delivers the goods, for it would indeed have turned out that Hesperus ≠ Phosphorus, had it turned out that Hesperus-appearances and Phosphorus-appearances were appearances of different planets.

So, to judge by our examples, "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 B, had it turned out that A as an analysis of A ➔ B. 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. (E) It is conceptually necessary that S iff some world w is such that it would have turned out that S, had w turned out to be actual.

A simpler way to put it is that S is metaphysically possible iff it could have been that S, and conceptually possible iff it could have turned out that S.

1I. Analyticity & Apriority

Apriori truths are truths that can be known not on the basis of empirical evidence. How well that accords with the Kripkean notion depends on one's theory of justification. There is a danger though that the definition takes in too much, for it can be argued that a very great deal is known "not on the basis of empirical evidence."

Some say that all of our 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 I spontaneously come to think that the sun is shining. (I don't infer that the sun is shining from premises about how things perceptually appear to me.) Then I may well know that the sun is shining, and not on the basis of any empirical evidence. But, it certainly isn't apriori, as Kripke uses the term, that the sun is shining. Again, some say that our most "basic" beliefs lack empirical justification, because they are epistemically priori to anything that might be said in their support. But it isn't apriori in Kripke's sense that nature is

uniform, or that the more explanatory hypothesis is other things equal likelier to be true.

Apriority then is not any old kind of not-empirically-based knowability, as judged by any old theory of empirical justification. That would let far too much in. We can get guidance on how to clarify the notion by looking at a (familiar) objection from the other side. If experience were strictly off-limits, then it would be enough to stop S from being apriori if it were through experience that we understood S. That would keep too much out. Phrases like "not on the basis of experience" are brought in to signal that our interest is in how S is justified, our understanding of S taken for granted.

Kripke's notion of apriority is what you get if that is the one and only concession made. S is apriori iff it's knowable just on the basis of one's understanding of S. Or, better, it's apriori for me iff I can know it just on the basis of my understanding of S. This explains among other things why the person fixing a word's reference is sometimes in a position to know more apriori about the referent 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.

So, apriority is "knowability on the basis of understanding." "Understanding" is knowing the meaning. But what meaning? It can't be meaning "in the strict sense": the sense that ignores reference-fixing descriptions. For Kripke calls it apriori that Hesperus = Hesperus, and aposteriori that Hesperus = Phosphorus; yet 'Hesperus' and 'Phosphorus' are exactly alike in strict meaning. Presumably then we're talking about meaning in the loose sense, the sense that corresponds to knowing the reference-fixing descriptions. Loose meaning though corresponds to what we are calling narrow

content. So it does not do too much violence to Kripke's intentions to say that S is apriori 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; for it is not apriori and he thinks analytic truths are apriori. Then what is his intent in speaking of "meanings in the strict sense"? He meant not to include statements (like 'Hesperus = Phosphorus') true in virtue of strict meaning as opposed to loose, but to exclude statements (like 'Hesperus is sometimes visible at night') 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 well feel, is just terminological fussiness. Even Kripke considers it a kind of meaning -- meaning in the loose sense -- and he notes 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 complements our account of Kripkean apriority as knowability in virtue of grasp of narrow content.)

Now finally we can ask the really important question: 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 Chalmers's 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.

Is conceptual necessity also perhaps a kind of apriority? 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 the "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, I assume, in a position to know the conditionals. So if S is conceptually necessary, then the speaker 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 the speaker can determine apriori that S?

No; in fact we are still miles from the conclusion that S is apriori knowable. There are many gaps that would need filling but the most important is this. Let it be that the speaker knows for each w that w obtains ➔ S It is wide open so far whether this knowledge is apriori. Someone who grasps S's meaning is in a position to come to know the conditionals somehow or other. Apriori or aposteriori is an open question.

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. And knowledge based on grasp of content is apriori knowledge.

But this is 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. But this only shows 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. And the question that matters to us is whether understanding is sufficient. Understanding might suffice for being in a position to "figure out" whether S if w. But the figuring out could depend on experience, and then the knowledge will not be apriori.

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 in The Conscious Mind:

> [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 [76].

Say this is right; I am able 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 apriori.

Not on the face of it. Asking yourself how a world strikes you is using yourself as a measuring device. Information acquired by use of a regular measuring device is aposteriori; information acquired by use of yourself qua device is 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. Isn't that enough to make the knowledge apriori?

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

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

Two things. (1) 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." One reason knowledge gained by internal looking is not apriori is that it is

acquired through the use of a perceptual faculty rather than a cognitive one.

(2) Some imagined reactions are better guides to our "real" reactions than others. 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: the season, the fact that people are swimming. But it could also have been an imagined judgment based on full information. You are given that the water has such and such kinetic properties; when you imagine water like that, and yourself stepping into it, you imagine it feeling warmer than it turns out in fact to feel. (Just as water at 79 degrees farenheit feels colder than you expect.) The point is this. For all you know apriori, your imagined judgments of shape are just as unpredictive of how you would really react as your imagined judgments of temperature. (One can think of examples where your apriori worry is borne out.) If the mind's eye sees (some) things more or less as real eyes see them, this is an empirical truth, not an apriori one

The temptation to think that perceptual simulation makes for apriori knowledge is due in part to the fact that our powers of perceptual simulation are extremely limited. There might be beings who, given a microphysical blueprint for X, find that when they simulate looking at X it looks not only triangular but also yellow. They are able to see X's color in much the way that we are able to see its shape. They come to know in this way that objects of the specified kind are yellow.

I take it that no one would consider that knowledge to be apriori. Our beings did not deduce the color from the microphysics. Information was also needed about how that microphysics appears to human eyes. They obtained this information experimentally, by simulating an encounter with X, and using it to predict the outcome of a real encounter.

Imagine that we had been able to simulate reactions in other modalities besides vision. We could determine the taste of a microphysically given item with the mind's tongue, its smell with the mind's nose, and so on. Would it then be an apriori matter how pineapples [insert chemical description here] tasted? The answer seems clear. How a chemical substance tastes is the paradigm of an empirical question. One does not feel that the question escapes being apriori only because of a contingent incompleteness in our nature. It would still have been aposteriori how pineapples tasted, even had God 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 the self-experimental knowledge is aposteriori, for any suggestion of self-experimentation was inadvertent. "I looked at w and saw that it was 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 S holds in w.

That is fair enough, on one condition. Envisaging w had better not be needed to work out whether S holds in w. It had better be that one can work out whether S holds by reasoning from a microphysical description of w to the conclusion that S, or ¬S. No peeking.

I assume that Chalmers would agree, for if peeking were allowed, then the inference from "S holds in all candidates for actuality" to "it's apriori that S" would no longer go through. That inference is central to the view Chalmers calls "modal rationalism." It's of the essence of modal rationalism that apriority goes with truth in all counteractual worlds.

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. You need to be careful, then, that my mind's eye is completely shut, or completely covered by my mind's blindfold. Strangely, the need for vigilance is never mentioned 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 (76)

I don't know about you, but 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 on me. The feeling intensifies when I read how "doubts that the device is an umbrella can be assuaged." Never mind how they are assuaged; I cannot see how the umbrella idea came up in the first place.

I see how it's supposed to go. I start with objective, geometrical, information.. A chain of apriori 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, and 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 elliptical) 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 $x^2 + 2y^2 \leq 63$. (It's 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 $x^2 + 2y^2 = 63$, then R is elliptical;
>
> if R is circumscribed by the points (x,y) such $x^3 + 2y^2 = 63$, then R is not elliptical; .....

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

But, it isn't clear that I <u>do</u> know many conditionals like these. (I am tempted to say that it's clear I don't.) And the few that I do know, I don't seem to know apriori. It wasn't learning the meaning of 'ellipse' that taught me the formula for ellipses. I worked it out empirically by graphing the formula, <u>looking</u> at the figure I had just drawn, and then <u>reflecting</u> on how I was inclined to describe the figure. (Perhaps some acquired the word at the same time as the formula. If so, let them substitute 'circular.' I take it that no one has their first encounter with circles in geometry class.)

I don't 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 that need to be argued for, and no argument has been given; on the contrary, the thesis has been presented as pretty close to what you'd expect.

13. Recognitional Predicates

Now let me move on to urging in a positive way that there's only so much we can judge with the mind's eye averted. I will be arguing that one indeed 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 -- the one that figures only inadvertently in his narrative -- is in some cases the only possible method. This will be argued first 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 this. 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 perceptually appears.

Nothing has been said about the kind of appearance or experience that marks a thing as P. Sometimes x is judged P because our experience of x has a quality Q quite 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 /feels/sounds/.... Then P is a recognitional predicate.

Examples are bound to be controversial. But, for instance, we find Kripke saying 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' " (140). Kripke says that we understand by yellowness whatever property it is that makes objects look yellow, or rise to the sensation of yellow. The predicate 'yellow' is recognitional on this view, since the yellow things are identified as the ones that look precisely 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, look yellow to whom? Perhaps what matters is whether it looks yellow to the w-folks, including in particular my counteractual self. If it's counteractual Steve's reactions that matter, then I don't need to experience x myself; I can infer x's color apriori from what the relevant world description says about the experiences Steve has when experiencing x.

But what does the world-description say about counter-actual 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 intentional content: do I feel myself to be having an experience of yellow or of green?

Suppose on the other hand that counter-actual Steve's experiences are described intentionally, as "experiences of yellow," 'yellow' here 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.[7]

How does Kripke avoid this problem, one might ask? He doesn't positively advocate the 'yellow'-before-'experience of yellow' position. He does however note that "some 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,...." (140). And he says that "this question is independent of any view" that he does positively advocate (140). 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 (should) say. Yellowness is identified not descriptively, 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. The experience of yellowness is not something to be discovered but part of the discovery process. And so the problem of identifying it doesn't arise. I don't have to identify my yellow-experiences in order to learn by their exercise, not any more than I have to identify my eyes in order to learn by the use of them. [8] [9]

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

This is a problem that Kripke 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" (140). If C-type circumstances are ones 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' and so cannot explain its meaning. If C-type circumstances are ones where (say) the light has such and such a composition, no one is suffering from jaundice, and there are no spinning black and white disks, 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 etc...in circumstances C" defines 'yellow,' why isn't it equally hard on the idea 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 apriori 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 apriori, as it would be if the description fixed 'yellow's reference.

Kripke 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 that the viewing conditions 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 apriori that what looks yellow under conditions C is yellow, but it does seem to be apriori 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.

So much is to argue that our grasp of recognitional predicates is, well, recognitional rather than intellectual. I do not reason my way to '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 apriori-reliable information about which conditions are deceptive. The most that is apriori is that these conditions are non-deceptive, unless there is reason to think something funny is going on.

If P is a recognitional predicate, then I have an apriori entitlement to "These conditions are (funny business aside) such that what seems P is P." This is an entitlement that by its very nature does not travel well. It does not stay with me as I move through counteractual worlds, for in lots of those worlds, one finds what from the perspective of this world is funny business. As matters stand, what feels smooth, is smooth. But that is how it did turn out, not how it had to. It could have turned out that some of the things felt as smooth were not smooth at all. They only seemed to be smooth due to (what from our present perspective is) some perceptual wires getting crossed.

This has two semi-surprising consequences, which for now I'll just state without argument. (1) Something known apriori might hold only in some counteractual worlds. An example is "funny business aside, what looks yellow is yellow." This fails in worlds where a physiological switcheroo makes us bad judges of yellowness. Likewise "what feels smooth, is smooth," "what sounds loud, is loud." (2) Something holding in all counteractual worlds might be

knowable only aposteriori. Let x be something smooth (a baby's cheek, the side of an ice cube) and let F be a complete intrinsic description of x. I claim there is no way Fs could have turned out not to be smooth. They could have turned out not to be felt as smooth, but whoever who finds a baby's cheek rough is a bad judge of smoothness. I will not press these points now but they will be returned to below.

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 fat to count. To a first approximation, a figure is oval if it has the dimensions of an egg, or a two-dimensional projection of an egg. Of course, few of us know in an intellectual way what those dimensions 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.[10]

Because our grasp of oval is constituted in part by our responses, one might be tempted to group it with "response-dependent" concepts like ticklish or aggravating. That would be a mistake. There are several respects in which oval is quite unlike ticklish, which once pointed out make the label "response-enabled" seem much more appropriate. Another term I shall use is "grokking concept."

Constitution: Why are ticklish things ticklish? If that means "what is the evidence that they are ticklish?" the answer is that we respond to them in a certain way. 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. The interesting thing is that the answer still has to do with our giggling if the question means, "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.

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. What about 'oval'? Our responses may give us a special sort of access to the extension of 'oval,' but they do not dictate the extension.

Motivation: Why are the ticklish things picked out experientially? There's 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? For a practical reason: we have no other way of roping in the intended shapes.

Evaluation: Suppose that this had turned out to be w instead of @. External objects are the same in w as here, but our responses are different. What would have turned out to be ticklish? That depends on what would have turned out to tickle us. What would have turned out to be oval? The things that do look egg-shaped, quite regardless of our counteractual responses. For dimes (e.g.) 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 actual 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.

And that is false. If he were right about counteractual worlds, then

dimes would have turned out to be oval, had things turned out differently in the perception department, so that circular objects looked egg-shaped

That 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. In particular, you can't change x's status as 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 see basketballs as egg-shaped would be a way for it to turn out (not that they were oval, but) that they were misperceived.

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."[11] The role of "this big" is not to pick out whatever old size one turns out to be perceiving. "This big" wouldn't turn out to mean 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 has no way of reporting on the size other than via its perceptual appearance.

15. Analyticity without Apriority

First there are the response-dependent concepts: ticklish, aggravating, painful-to-behold. Then there are the response-

enabled concepts: oval, aquiline, crunchy, rotund. Response-enabled concepts have their own distinctive pattern of evaluation at counteractual worlds. If oval were response-dependent, then I could determine its extension in w by asking what the people there see as egg-shaped. If it is response-enabled, then those counteractual responses are irrelevant. Ovality is to be judged not by as-if actual observers, but 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 One:

INSERT FIGURE ONE (CASSINIAN OVAL)

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 One. Maybe there are no observers in w, or maybe there are but they do not think Figure One 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 One is oval? Would Figure One have turned out still to be oval, had it turned out to be shaped as above? You bet it would.

Whether an as-if actual figure is oval is completely determined by its shape. Had it turned out that people (we) did not experience Figure One as having the right look, that would just show that people had turned out to be bad judges of ovality.[12]

Suppose we do some measurements and determine that Figure One is defined (up to congruence) by the equation $(x^2 + y^2)^2 - (x^2-y^2) = 5$. A figure like that can be called $(x^2 + y^2)^2 - (x^2-y^2) = 5$-shaped, or for short cassini-shaped, or for shorter cassinis. (Giovanni Cassini (1625-1712) studied a class of figures of which this is one.[13]) 'Cassini-shaped' is an objective, third-personal, predicate applying to all and only figures with the geometrical properties (that we all correctly take to be) exemplified by Figure One.

Consider the statement 'cassinis are oval.' Could it have turned out otherwise? Is there a w such that had things turned out as in w, cassinis would have turned out not to be oval? If ovality in a world is purely a function of shape, and whatever is cassini-shaped has the right sort of shape, then the answer has got to be no. So 'cassinis are oval' is true in all worlds-taken-as-actual, or conceptually necessary.

But, of course, it it is very far from apriori 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. 'Cassinis are oval' is an analytic (conceptually necessary) claim that cannot be know apriori.

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 extensions 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 apriori 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.

Now, 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 counterworld. 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 (advocated by Chalmers) harks back to the indicative account of truth in a counteractual world. For S to hold in w-considered-as-actual is, on that account, for it to be the case that if w, then S. We rejected this account on the ground that it makes every sentence conceptually contingent. (If 'sibling' means triangle, sisters are not siblings.) But, you may say, there is an obvious fix. It should be not merely true but apriori that if w obtains then S. It is not apriori that sisters aren't triangles, so a world where 'sibling' means triangle is not on the new definition a world where sisters fail to be siblings.

S's "epistemic" intension is the set of worlds such that it's a priori that if w obtains, then S. Suppose that conceptual necessity is conceived as necessity of the epistemic intension. What happens to our argument above that conceptual necessity is a kind of analyticity but not a kind of apriority?

The argument would to fall apart. 'cassinis are oval' may have a necessary primary intension, but its epistemic intension is contingent. (It holds apriori, one could argue, that if cassinis don't look egg-shaped, then they're not oval.) But then, if conceptual necessity goes with the epistemic intension, 'cassinis are oval' is conceptually contingent. And so it no longer serves as a counterexample to the idea that whatever is conceptually necessary is apriori.

This assumes, however, that intensions built on apriori indicatives avoid the problems that were raised for intensions built on ordinary indicatives. Are they? What is clearly true is that the example we gave no longer works. But this is only because it is not apriori that sisters fail to be triangles. It is apriori (let's assume) that sisters aren't triangles. And so it is apriori too that if 'sibling' means triangle, then sisters aren't siblings.

I'm not sure it is needed, here is some argument to back the claim up. I believe that if 'sibling' means triangle, then sisters aren't siblings. Suppose that belief is 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 triangle. Call that the actual-meaning-fact. Does it form part of my justification for believing that if 'sibling' means triangle, 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 is compromised. Say I fall under the impression that 'sibling' means triangle. Have I now lost my grounds for thinking that if it means triangle, then sisters aren't siblings? Surely not. My reasons for thinking that if 'sibling' means triangle, 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?[14]

Where does this leave us? If my belief in the conditional is apriori justified, then there is a world not in the epistemic intension of 'sisters have siblings.' The same argument shows that no statement S, however apriori in appearance, has a necessary epistemic intension. I conclude that the apriori-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.

Maybe the problem is not with the aprioritizing as such, but the type of conditional aprioritized. A further option is to call S true in w-considered-as-actual iff it holds a priori that (w obtains ➔ S) -- it holds apriori that it would have turned out that S, had w turned out to be actual. The intensions that result are priory intensions. If conceptual necessity is necessity of the priory intension, maybe the inference to a counterworld can be saved. Certainly it isn't refuted by our original 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.)

Now the modal rationalist runs into a different problem. Not only is apriori truth supposed to go hand in hand with intensional necessity, apriori falsehood is supposed to go hand in hand with intensional impossibility. In the absence of worlds w such that S is true if w, S should be apriori false. Since one can never tell apriori whether cassinis would have turned out to be oval had it turned out that w, 'cassinis are oval' has nothing in its priory intension. The same goes for 'cassinis are not oval.' It goes in fact for just about all sentences whose predicates express response-enabled concepts. If one can't determine apriori 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 apriori, there should be a world not in its priory intension. Why should the modal rationalist want any more? Take the refutation of physicalism. If it is not a priori that "if PHYSICS, then PAIN," then there are worlds not in that conditional's priory intension. And worlds not in that priory intension are worlds physically like ours in which no one feels anything.

This argument assumes that priory intensions are like primary ones in a certain respect, when really the two are different. If PIs are primary intensions, then worlds not in a sentence's PI are worlds in which S is false. If PIs are priory intensions, S need not be false in the omitted worlds.. All we can say is that it fails to be a priori that had it turned out that w, it would have turned out that S. It might still be true that it would have turned out that S! Once again, then, failures of apriority do not deliver worlds in which S is false. And you need worlds like that to reach the conclusion that there could be PHYSICS without PAIN.

I present this as a problem for priory intensions, but epistemic intensions are every bit as vulnerable to it. I don't see 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 this world 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 is true however things turn out.

17. Grasping Meaning

Why expect an analytic (conceptually necessary) sentence to be knowable apriori? 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 w-true, but only how to work out whether S is w-true. If this sort of grasp is possible, then its not being apriori fails to establish the existence of a falsifying world. The sentence might be (like 'cassinis are oval') aposteriori but true in every candidate for actuality.

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 ➔ S, then I don't understand S. To understand, I have to know that w ➔ S. This makes all the difference. If knowledge of the conditionals is necessary for understanding S, then understanding S is sufficient for knowing the conditionals, and so they are known apriori.

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. Which it is according to the modal rationalist.

Why are other forms of understanding ruled out? The closest thing I've found to an explicit discussion is Chalmers's reply to Loar in The Conscious Mind.

Summarizing (that discussion) greatly, Loar thinks that pain is a recognitional concept;[15] c-fiber firings is a theoretical concept; recognitional and theoretical concepts are cognitively distinct; but their distinctness notwithstanding, "it is reasonable to expect a recognitional concept R to "introduce" the same property as a theoretical [concept] P." This means that 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. That is enough to make them apriori inequivalent, so there is nothing to stop them from referring to the same state. Chalmers doubts that all these elements can 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 apriori, then reference to R and P is fixed in different ways and the reference-fixing intensions can come apart in certain conceivable situations (1996, 373).

This is puzzling. Observational concepts (of which recognitional or a subtype) do not have their reference fixed in any epistemically available way, hence not in a different such way than holds for theoretical concepts. This is not special pleading on behalf on the concept of pain; it holds for all observational kinds.

Now, Chalmers appreciates that Loar claims that "recognitional concepts refer "directly," without the aid of reference-fixing properties...." (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" (373).

I don't know how Loar would reply, because I don't know how much tolerance he has for the notion of primary intension. It doesn't matter, because we can concede a substantial primary intension. Such an intension is not at odds with Loar's claim of directness. The directness Loar is talking about is epistemic; one doesn't (and couldn't) infer that the cactus is R from its lower-level properties. The directness Chalmers says he can't have is semantic, 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 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. One might consider it 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 you are oval is not a brute fact about you but depends inter alia on your shape.

But "reference-fixer" can be said in two ways. 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 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.

The claim about concepts like oval is that they lack strong reference fixers. Speakers don't know what qualifies something to be regarded as oval; they don't know any conditions that get the extension right no matter how things turn out. 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 is enough.

So: it would indeed 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 the claim is right. Speakers don't (and often can't) determine extensions apriori 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 and others have made the following anti-physicalist argument. It is not a

priori that "if PHYSICS then PAIN"; so, the primary intension of this sentence cannot contain every world; so, there are worlds physically just like this one in which pain is lacking; so, physicalism is false.

The problem as you might guess is with the inference from not apriori 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 the following:

Pain is (like oval) is a grokking-concept (=observational concept). As a result, 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 things follow. (1) 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 unknown to us. Just as you can't tell whether w lacks ovals except by sampling its shapes, you can't tell whether it lacks pains except by sampling its brain states. (2) Not only is a world like w of no particular help, it isn't needed. Suppose that v is a world just like ours in every respect. The question of whether there is pain in v is the question of whether there's anything there that hurts if sampled in the right sort of first-personal way. Whether something hurts when sampled is not the kind of thing that can be decided from the armchair. If we are trying to explain why PHYSICS doesn't apriori entail PAIN, a world whose zombie-ness can't be apriori ruled out works just as well as a real zombie world would.

18. Evaluative Predicates

Our grasp of a concept is rationalistic if it consists in whole or part of a certain kind of knowledge: knowledge of conditionals of the form w obtains ➔ x, y, z,...are the Cs. What if your conditionals put x,y,z,...into a concept's extension in w, while mine count x , y, z out. 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 figure oval, and this is not because of misinformation or error or oversight on either side, then probably we have different concepts; probably we mean slightly different things by 'oval.' 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 forgetting something or etc., 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 like this is "intolerant of brute disagreement."

Suppose though 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 extensions of 'wrong' (and where the disagreement does not trace back to etc.) 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 "merely semantical" just because you can't see any good way of bring the

two parties into line. A phrase sometimes used for concepts like this is "tolerant of brute disagreement."

A lot of people 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 idea of meaning dictating extension. 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. It flows from the meaning that questions of rightness are contestable in the (rather minimal) sense that someone who disagrees with you can't be charged on that basis alone with meaning something different by 'right.' Concepts like right are not intolerant of brute disagreement, indeed they are intolerant of intolerance.

How do we grasp of meaning of 'right'? Could our grasp be rationalistic? If so then everyone grasping the meaning aright knows 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 would be intolerant of brute disagreement. And the 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.

There is a well-known puzzle about morality. On the one hand we are told that you can't derive an ought from an is. "If N then M," where N is descriptive and Q is evaluative, cannot be known apriori. On the other hand it does seem to be apriori that the evaluative facts are fixed by the descriptive ones. The tension here becomes an outright contradiction if our grasp of evaluative concepts is purely rationalistic.

Suppose with the modal rationalist that if it is not apriori that S, 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: these two worlds taken together constitute a counterexample to the (apriori true) thesis that there can be no moral differences without underlying descriptive differences.

It is true 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 by hypothesis the same in u as in v. So:

(1) it is not a priori that if N then M, and it is not a priori that if N then ¬M
(2) if 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) -- an apriori falsehood.

Where is the mistake? Is there a feature of moral truth that we're overlooking? Or does the argument show there can't be any such thing as moral truth?

I claim that the puzzle has nothing essential to do with morality. Consider the conditionals, 'if something is cassini-shaped then it is oval,' and 'if something is cassini-shaped then it is not oval.' Neither of these is knowable apriori. 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)? As before, though, u and v taken together would seem to constitute a counterexample to the thesis that there can be no differences in respect of ovality without underlying geometrical differences.

Where the ovality argument goes wrong is easy to see. Line (2) is false. You can't get a world where cassinis aren't 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 arguably call for a counterworld. But it isn't, so it doesn't.

Why shouldn't the morality puzzle be blamed 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 purely rationalistic. This suggests that we grasp right in an other than rationalistic way. The same argument appears to show that evaluative concepts in general aren't grasped rationalistically.

19. Theoretical Predicates

Consider finally theoretical predicates: acid, energy, force, mass, species, cause, sum. What can be said about our grasp of these? Do we understand 'acid' by knowing a great lot of conditionals of the form 'had it turned out that w, such and such would have turned out to be the acid'?

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 determines for each possible scenario, but discloses to us, for each of these scenarios, where the energy is to be found. Then if you and I 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 -- the explanation must lie in one of two places. Given that we both know a conditional that resolves our disagreement, it must be that

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

(ii) someone is misconstruing the content of 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 change his/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 and so are talking past each other. None of the three scenarios allows for substantive disputes; someone has made a mistake or else we are arguing over words.

This is almost as hard to accept in the case of theoretical disagreement 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's reasonable to believe, between people not guilty of any certifiable mistakes. The effect of the rationalistic theory of grasp is to eliminate this third category. If the third category is real, then the rationalistic theory is incorrect, or anyway incomplete.

The extension of 'energy' in a world is a function of what the correct scientific theory is. Post-positivistic philosophy holds that to find that theory, one must appeal at some point to considerations of naturalness, simplicity, nonarbitrariness, etc..... in a word, considerations of reasonableness. Being an evaluative concept, though, reasonableness is response-enabled. There are no agreed-on canons of reasonableness; you have to let yourself be led to some extent by your gut.

There are places where Chalmers seems to agree with this. Figuring a concept's extension, he says, may involve more than just grinding out inevitable 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. (58).

I suppose that we do have this ability. We can ask ourselves what the most reasonable thing to say is on various hypotheses about how the world turns out. The problem is to see how that argues for the matter's being apriori. After all, 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 apriori, so why should they be apriori 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 apriori" is a non sequitur. One can stage a simulated confrontation with nature on various hypothesis about how nature turns out. It may not be obvious that searchers after the most reasonable hypothesis are doing this. But they are. Judgments of reasonableness and plausibility are arrived at not by reasoning from premises but exercising a type of sensibility. The alternative is to say that reasonableness admits of an apriori discovery procedure, and that we can get by just applying the relevant rules. That is the dream of an apriori inductive logic, which we are supposed to have outgrown

20. Logical Empiricism and Modal Rationalism

If some of this sounds familiar, that's because it borrows from the Quine/Carnap debate about how to be an empiricist.

There were two dogmas of empiricism, you'll recall. One was the analytic/synthetic distinction. The other was "semantic reductionism" -- the idea that each statement is linked by fixed correspondence-rules with a determinate range of confirming observations. Quine of course 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 and so cannot fail as long as that meaning holds fixed.

The point that matters to <u>us</u> is that the correspondence-rules are apriori; this was thought to follow from their analyticity. It will be simplest though to continue to speak of analyticity, following the positivists in treating this as enough for apriority.

First let's look at some apparent differences between the logical empiricist's picture of meaning and the modal rationalist's. The first had analytic correspondence rules connecting theory to experience. Modal rationalists aren't proposing anything like <u>that</u>. Yes, people have to be able to tell apriori (just by their grasp of meaning) whether S is true in a presented world. Gone though is any thought of that world being presented <u>experientially</u>. There is no case then for a charge of experiential or <u>phenomenalistic</u> 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 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.[16] Often it is said to be a matter of <u>convention</u> which sentences will count as protocols. The important point for us is that Carnap thinks there

are apriori rules connecting theoretical statements with protocols whatever protocol 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 confirmation actually works.

This complaint the rationalist can rightly claim to have answered. He never represents limited information about a world 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 (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).

The problem here is not that S's confirmational status is underdetermined until all the empirical evidence is in; it's 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 T 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 corpi (?) 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 (1951, 49)

This could be taken to mean just that the sought-after theory of confirmation would have to be very 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 (the m-functions), none of them looking apriori 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" (1951, 45). The version of underdetermination he 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.(1951, ???)

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 to trade these sorts of desiderata off against each other. If two scientists judged the tradeoffs differently, at most one could be considered to be speaking correctly, ie., in accordance with the meanings of her words. That of course is not how the science game is played.

The interesting thing is that Carnap agrees it's not how the science game is played. His goal as he usually describes it is not to characterize the true nature of meaning, but to give us tools for making our discursive practice more rational and efficient. He thinks 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... (1936/7, #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'm

saying right now is that not even Carnap believes that it's 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 it's how we operate? The modal rationalist does, or at least that 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 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 to 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 the idea 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. such that S holds in them.

If a "homophonic" grasp of the set of verifying worlds was 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' discussion of physicalism. Consider again the conditional "if PHYSICS then PAIN." It's claimed that the only way for this to be non-apriori 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

explain itself. The reason I don't know apriori that if PHYSICS then PAIN is that I can't tell apriori whether the primary intension of 'if PHYSICS then PAIN' contains all worlds. I can't tell that because I can't tell apriori whether the PHYSICS-worlds are a subset of the PAIN-worlds. If they are a subset, there is no puzzle of why the understander doesn't realize it, because it's a given that PHYSICS worlds are, for all she knows apriori, worlds without PAIN.

How then are worlds presented to the meaning-grasper? He must know what lower-level facts have to obtain in a world if it is to verify S. "If the low-level facts turn out like this, then the high-level facts will be like that" (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 verificationism; (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. 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 into a situation in which S really holds. The surprising thing is that we can do that 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.[17] 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) the resisted hypothesis is not conceptually false (remember essential contestability), and (ii) lots of conceptually false scenarios are not resisted (as readers of Calvino and Borges will attest).

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 x 4's as they pushed gracefully through the mud. Expectations were high; last year's death match blood bath of doom had been a thing of beauty, 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." 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 the death match of doom is a thing of beauty.[18]

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. Circles, squares, octagons -- all the other shapes there were plenty of. Why not the one on their list? Hang on, Sally said. It's staring us in the face. What kind of tree are we lying under? A maple tree. She tore off a leaf. Here was the oval they needed! They ran off to claim their prize." If the author wants it be a maple leaf, that's her prerogative. But the leaf's physical properties having been settled, whether the leaf is oval is not up to her. She can, perhaps, arrange for it not to have the expected maple-y shape. But if it does have the expected shape, then there is not a whole lot the author can do to get us to imagine it as oval.

The claim is that imaginative resistance arises not only with evaluative predicates, but also certain descriptive ones: 'oval,' 'aquiline,' 'jagged,' 'rotund,' 'squishy,' 'rough,' 'lilting.' What do these predicates have in common? The further claim is that P makes for imaginative resistance if, and because, P expresses a response-enabled or "grokking" concept.

Why should resistance and response-enabledness be connected in this way? Start with the concepts. A distinctive feature of grokking concepts is that their extension in a counteractual situation depends on how that situation (or type of situation) strikes us. Their extension depends on how the situation really strikes us; how it is represented as striking us 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.

22. (Conceptually) Contingent Apriori

I have called a lot of claims apriori. But not much has been done to explicate the notion. (The focus has been on conceptual necessity.) I am not sure it is possible to explain apriority with the materials at hand. But I'll try in the next few sessions to clarify things as far as I can.

Kripke proposes 'water contains hydrogen' as an example of an aposteriori necessity -- an aposteriori metaphysical necessity. 'cassinis are oval' has been promoted here as an example of an aposteriori conceptual necessity. Our aposteriori conceptual necessities are the counterpart in our system to Kripke's aposteriori metaphysical ones.

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

I said above that 'a figure is oval iff it looks egg-shaped.' was apriori, 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 it have turned out the world was oval-free? The answer seems clear. How we turn out to see things is irrelevant to how things turn out to be shaped. Oval figures would have turned out not to look egg-shaped.

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 'oval' would have turned out to have in that circumstance is the meaning it has actually. And it seems clear that it wouldn't have been. 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.

So: 'a figure is oval iff it looks egg-shaped' is an example of an apriori but conceptually contingent truth. It could have turned out that we not prone to see ovals as egg-shaped, because, e.g., we were not prone to see anything as egg-shaped. And, approaching it from the other end, it could have turned out that it was circles and squares that looked to us egg-shaped. And circles and squares are not 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. The puzzle evaporates if we 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 apriori than one in which '=' means nonidentity can stop 'Hesperus=Hesperus' from being apriori.

23. Apriority vs. Conceptual Necessity

I said that it could have turned out that 'oval' didn't 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' didn't 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 it could turn out that 'oval' doesn't 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 for that, letting BLAH be what Crazy Eddie said, it could have turned out that BLAH. The scenario in which it turns out that BLAH could be a scenario in which BLAH 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.) There is no danger then 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 will assume that "it could turn out that..." is an intensional context, that is, one treating synonyms alike. Since 'sister' is synonymous with 'female sibling,' and it could turn out (though it won't!) that 'sister' doesn't mean female sibling, it could turn out that 'sister' doesn't mean sister. The reason it sounds funny to say it is that the statement suggests something false, viz. 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's a use of "it could turn out that S" on which it means that it is not apriori that ¬S. In that sense of the phrase, it really <u>couldn't</u> have turned out that 'sister' didn't mean sister. For we know apriori that 'sister' means sister. If it doesn't sound as bad to say that it could turn out that 'sister' doesn't mean female sibling, that's because we don't know apriori that it does mean female sibling.

Compared to conceptual necessity, apriority is an elusive notion. One reason has already been noted. If it is apriori that 'sister' means sister, but not that it means female sibling, then "it is apriori that..." is not an intensional context; it cares about the difference

between synonyms. ("It could have turned out that... (epistemic sense)" is therefore not intensional either.)

Stranger even than the failure of intensionality is the following. The apriori truths are often claimed to be closed under (obvious) logical consequence. This can't be right, if the textbook explanations of apriority are even roughly correct. It is apriori that S, the textbooks tell us, if one can know that S is true just on the basis of one's grasp of S's meaning. Suppose I know A and A → B just on the basis of my grasp of their meanings, and then I infer B. If this is my only basis for believing B, then I do not know it apriori. For my belief is based in part on my grasp of A's meaning, and A is a different sentence than B.

The failure of logical closure helps us resolve a puzzle. There are many things I know apriori. 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 apriori that 'S' is true iff S.

However I rarely if ever know apriori that a sentence is true. For a sentence's truth depends on its meaning, and my knowledge of meaning is generally aposteriori. I have to learn what a sentence means, and my views on the topic are rationally defeasible under the impact of further evidence. (This applies even to sentences of my own idiolect - even, if such there be, to private language sentences.)

The question is, why can't I arrive at apriori knowledge that 'sisters are siblings' is true from my apriori knowledge that sisters are sibling and my apriori knowledge that if sisters are sibling then 'sisters are siblings' is true?

The problem is not that I can't infer that 'sisters are siblings' is true, by modus ponens, from things I know apriori. I can. 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. My understanding 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 -- either in public language or in my personal idiolect -- 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 apriori that 'S' is true.

24. Apriority

What can we say about apriority to explain these puzzling features? I don't have a precise definition to offer, but the following seems to point us in the right direction.

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 apriori (for me) that S iff for some G

(a) that 'S' is G is part of my grasp-constituting information, and
(b) it is conceptually impossible to be G without being true.

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

How can it be apriori that 'sister' means sister yet not apriori that 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 strictly figure in my grasp. Suppose the belief changed in response to some outré counterexample; that would be a change in what I thought sisters were, but not a change in what I meant by 'sister.'

Why are the apriori 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 apriori, I need to know it on the basis of the information whereby I grasp B.

How can an apriori 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 itself is that 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.

Can you give an example?

I am newly arrived in the royal court. A helpful minister explains that "the King" is to be understood so that "the King is the guy wearing the crown, giving all the orders, etc." comes out true. I come as a result to know apriori that the King is the guy wearing the crown, etc. 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 it could have turned out that it was an impostor Richerd who was walking around etc. Would the King have turned out to be Richerd in that circumstance?

I've 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. It is conceptually contingent that the King = the order-giver. Still, I know it apriori.

Why are some conceptually necessary truths not apriori?

Sometimes the information that a typical 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. I should perhaps know that things looking egg-shaped are to be counted oval. But that doesn't enable me to work out that cassinis are oval until I've laid my eyes on one.

You say there is a world in which 'sisters are siblings' turns out not to be true, but there are no worlds in which sisters turn out not be siblings. Shouldn't there then be a world in which it turns out that 'sisters are siblings' is untrue despite that sisters are siblings? Why isn't this a world in which the T-biconditional fails?

It is a world in which 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 such a thing could not happen except by 'sisters are siblings' turning out to mean something different from what you and I know it does mean. A world where it turns out to mean something different is a world where my grasp-making information fails. Such a world is irrelevant

to the issue of whether my information entails the truth of the biconditional.

Why do people think that if it is not apriori that S, there is bound to be a counterworld: a world which taken as actual makes S false?

Suppose S is apriori but there is not a counterworld, that is, every candidate for actuality is an S-world. Surely the fact just mentioned is part of the information whereby we understand S.[19] But that fact entails that 'S' is true. So we possess grasp-making information entailing that S is true; in other words we know S apriori after all. BUT: the sentence beginning "surely" assumes that we grasp S rationalistically. People find the inference plausible because they're assuming that grasp of meaning has got to be rationalistic; and given that assumption, it is pretty plausible.

Bibliography

1. Almog, J. , J. Perry, & H. Wettstein, 1989. Themes from Kaplan (New York: Oxford)

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

3. Block, N. 1991. "What Narrow Content is Not," in B. Loewer & G. Rey, ed. Meaning in Mind: Fodor and His Critics (Oxford: Blackwell)

4. Carnap, R. 1936/7. "Testability and Meaning," Philosophy of Science, Vol. 3, 419-471, and Vol. 4, 1-40

5. Chalmers, D. 1996. The Conscious Mind (New York: Oxford University Press)

6. Chalmers, D. ms1. "The Components of Content"

7. Chalmers, D. ms2. "The Tyranny of the Subjunctive"

8. Chalmers, D. 2001. "Does Conceivability Entail Possibility?" (this volume)

9. Davies, M. and L. Humberstone. 1980. "Two Notions of Necessity," Journal of Philosophical Logic 38, 1-30

10. Evans, G. 1979. "Reference and Contingency," The Monist 62, 161-189

11. Jackson, F. 1994 "Armchair Metaphysics," in M. Michael & J. O'Leary-Hawthorne, ed. Philosophy in Mind (Dordrecht: Kluwer)

12. Jackson, F. 1998. From Metaphysics to Ethics. A Defense of Conceptual Analysis (Oxford: Clarendon Press)

13. Kripke, S. Naming and Necessity (Cambridge: Harvard)

14. Loar, B.1990. "Phenomenal States" Philosophical Perspectives 4, 81-108

15. McGinn, C. 1999. Knowledge and Reality (Oxford: Clarendon)

16. Moran, R. 1989. "Seeing and Believing: Metaphor, Image, and Force," Critical Inquiry 16, 87-112

17. O'Grady, P. 1999. Philosophy and Phenomenological Research. 49, 1015-1027.

18. Quine, W.V. 1951. "Two dogmas of empiricism," Philosophical Review 60, reprinted in Quine 1961

19. Quine, W.V. 1960. Word & Object (Cambridge: MIT Press)

20. Quine, W.V. 1961. From a Logical Point of View, 2nd edition (New York: Harper & Row)

21. Segerberg, K. 1972. "Two Dimensional Modal Logic," Journal of Philosophical Logic 2, 77-96

22. Stalnaker, R. 1972. "Assertion," in P. Cole, ed., Syntax and Semantics 9 (New York: Academic Press)

23. Stalnaker, R. 1991. "How to Do Semantics for the Language of Thought," in B. Loewer & G. Rey, ed. Meaning in Mind: Fodor and His Critics (Oxford: Blackwell)

24. Szabo-Gendler, T. 2000. "The Puzzle of Imaginative Resistance" Journal of Philosophy 97, 55-81.

25. Walton, K. 1994. "Morals in Fiction and Fictional Morality" Proceedings of the Aristotelian Society, suppl vol. 68, 27-50

26. White, S. 1982. "Partial Character and the Language of Thought" Pacific Philosophical Quarterly 63, 347-365

27. Yablo, S. 1993. "Is Conceivability A Guide to Possibility?" Philosophy & Phenomenological Research 53, 1-42

---

[1] See in this connection Chalmers, "Tyranny of the Subjunctive."

[2] At least three ideas were involved. (1) Instead of moving S over to w, bring w back to S. (2) To accomplish (1), evaluate S "on the hypothesis" that w actually obtains. (3) To accomplish (2), look at the indicative conditional "if w actually obtains, then S." (1) and (2) are perhaps hinted at in Evans 1979 and Davies and Humberstone 1980, but not made fully explicit until Chalmers, ms1. For (3) see Chalmers 1996 and (especially) ms2. Also of interest is Jackson 1994.

[3][3]A third option is to leave S and w where they are, and treat "true if" as a transworld primitive. This is one possible reading of Chalmers' remark in "The Components of Content" 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]" (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 pretheoretical grip on the second is that much weaker).

[4] Chalmers, "The Tyranny of the Subjunctive."

[5] Indicative conditionals are conditionals with antecedent and consequent in the indicative mood. Philosophers have proposed various theories of these conditionals. One, defended by Grice, is that they are "material" or truth-table conditionals. Another, defended by Adams, 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 apriori.) The objection in the text applies regardless. However the indicative is interpreted, A's apriori entailing C suffices for the apriority of "if A then C." The conditional "if w obtains (in which horses are wingless and 'tail' means wing), then horses do not have tails" has A apriori entailing C, so the conditional is apriori. (The version of this that bears on Chalmers' proposal: we know apriori that ¬(horses lack wings & 'wing' means tail & horses have tails).)

[6] 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.

[7] One option is to say that yellowness and the sensation of it are identified together by means of a gigantic Ramsey-type theoretical definition. I will ignore that possibility.

[8] 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 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 is saying something similar: reference is fixed experientally, not descriptively by a condition alluding inter alia to a certain sort of experience.

[9] 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)

[10] 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. If I am wrong, the reader can substitute a different example. (Other predicates arguably in this category: "jagged," "pungent," "rotund.")

[11] Peacocke 1989.

[12] I am not claiming that this is what the bad-judge observers would say. They might well regard themselves as having learned something new about which figures are really oval. The claim is only that this is what we should say about the reactions of as-if actual observers who don't think Figure One looks right. Their reactions might be a good guide to ovality as they understand the term, but they are a terrible guide to ovality as we understand it.

[13] The class of "Cassinian ovals," although not all are really oval.

[14] 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 triangle, then sisters are siblings. The question is whether my justification is aposteriori, 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. 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 based on the premise that ¬A. Since I do know that if "sibling" means triangle, then

sisters are siblings, my belief is not based on the premise that "sibling" does not mean triangle. But that is just to say that my belief is not based on the on the actual-meaning-fact. If it is not based on that, then it is not based on any empirical evidence. And if it is not based on any empirical evidence, then it is not apriori.

[15] This section is sloppy about recognitional vs. observational, not to mention Loar-recognitional vs. recognitional in our sense.

[16] Ayer1959, 237.

[17] For more on imaginative resistance, see the papers by Gendler, Moran, and Walton.

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

[19][19] This is a bit of an exaggeration, since knowing of each w that w ➔ S is not yet knowing that w ➔ S for all w.