Modality, Rationalism, and Conditionals

by

Jason Decker

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the
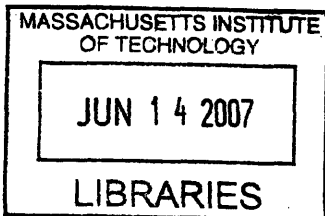
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

October 2006

Author . . . .
Department of Linguistics and Philosophy
October 3, 2006

Certified by. . . . .
Stephen Yablo
Professor
Thesis Supervisor

Accepted by . . . .
Alex Byrne
Chairman, Department Committee on Graduate Students

Modality, Rationalism, and Conditionals

by

Jason Decker

Submitted to the Department of Linguistics and Philosophy
on October 3, 2006, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three interconnected papers on *apriority*, modality, and conditionals. In "Playground Conditionals," I look at three philosophical debates, each of which turns on the epistemic status of a certain kind of conditional—what I call a *playground conditional*. I argue that a close consideration of playground conditionals gives us a better appreciation of what we can do with conditionals and, ultimately, some guidance concerning what to say about the three philosophical debates.

In "Modal Rationalism, Two Dimensionalism, and our Counteractual Sisters", I consider the prospects for modal rationalism in the wake of Kripke's *Naming and Necessity*. Recently there has been a modal rationalist revival, thanks in part to the development of the "two-dimensional" semantic framework. This framework associates two intensions (a *primary intension* and a *secondary intension*) with every sentence. The difficulty comes in finding a definition of primary and secondary intension that would lend the desired support to modal rationalism. After exploring and rejecting some of the proposed definitions in the literature, I sketch an account that can, I think, offer some support to a suitably framed modal rationalism.

Finally, in "A Guide to Modal Guidance," I set about to get clearer on how, exactly, we come to know modal truths. I start by considering two arguments that are designed to show that our access to modal knowledge cannot come from conceivability arguments. I show that these arguments are mistaken. In the process, I attempt to outline a broader and more realistic modal epistemology than one that focuses exclusively on conceivability. I then consider and reject a version of modal rationalism which says that ideal conceivability gives us *a priori* access to modality. Against this, I argue that our modal knowledge is predominantly a posteriori, and that our knowledge of ideal conceivability is *always a posteriori*. In the end, however, I attempt to salvage something that preserves the spirit, if not the letter, of modal rationalism.

Thesis Supervisor: Stephen Yablo
Title: Professor

Acknowledgments

Contents

Chapter 1

PLAYGROUND CONDITIONALS

1.1 Some Contentious Claims

I assume that most of us have some intuitions concerning what *is* and what *is not a priori*. I want to look at some cases where these intuitions clash, and not—it seems to me—because of the admitted fuzziness around the concept of *apriority*, but rather because of a lack of clarity concerning how to understand the claims themselves. Each of the claims in question has the conditional form 'if *p*, then *q*'. I think that looking carefully at a certain class of conditionals—I'll call them *playground conditionals*—gives us a better appreciation of what we can do with conditionals and, ultimately, some guidance concerning what to say about the epistemic status of our contested conditionals.

But lest it be thought that our subject matter is of only marginal philosophical interest, I will begin by situating each of our conditionals in a substantive philosophical debate, where the side one comes down on seems to depend on the epistemic status one assigns to the relevant conditional.

1.1.1 T-sentences in a vat

The first debate concerns Hilary Putnam's anti-skeptical argument in his *Reason, Truth, and History* (1981). Putnam gives us a way of convincing ourselves that we

are not brains in vats, drawing on his causal theory of reference. Anthony Brueckner, in "Brains in a Vat" (1986), reconstructs a version of the argument that he thinks is sound—at least to a point. Here is the conclusion of the sound part of the argument:

(1) My utterances of 'I am a BIV'[1] are false.

It would seem to be a short step from (1) to the desired anti-skeptical conclusion that I am not, in fact, a BIV. All I need do, after all, is invoke the relevant T(ish)-sentence:

(2) My utterances of 'I am a BIV' are true iff I am a BIV.

Together, (1) and (2) entail that I am a not a BIV. Skepticism has thus been refuted. Right?

Well, Brueckner thinks that this argument is too quick. In particular, he thinks that, whether or not (2) is true, we can't *assume* that we know it, at least when we are trying to construct an anti-skeptical argument. The problem is, we can't assume that we know what is being expressed by (2), unless we already know whether we are speaking English or "Vat-English". And to say that we already know that we are speaking English would be to *presuppose* that the skeptic is wrong, not to give some interesting argument against her.

But, even though we don't know what (2) expresses——Brueckner continues——we do know that it is going to come out true, whether I am speaking English or Vat-English. That is, we know that:

(3) My utterances of "My utterances of 'I am a BIV' are true iff I am a BIV" are true.

But this knowledge cannot be hooked up with (1) to get the conclusion that I am not a BIV.

On the face of it, this is a surprising argument. One might have thought that T-sentences, like (2), were going to come out *a priori*.[2] They certainly look like

---

[1]BIV = Brain in a vat.

[2]What does it mean for a sentence (as opposed to, say, a *proposition*) to be *a priori*? This is a question on which I'd like to remain as neutral as possible. I think that those of us who aren't

good candidates for claims that are knowable with justification that is independent of experience.[3]

Evidence: Perhaps it is just laziness on my part, but I have never taken it upon myself to count all the atoms in the universe. In fact, I have never undertaken even a more modest counting project in this area. I wouldn't know how to get started on it. But one thing I am *quite* sure of is that *the sentence 'the number of atoms in the universe is odd' is true iff the number of atoms in the universe is odd*. And I have a similar level of confidence in every other T-sentence. My justification for believing them seems in no way empirical.[4]

So we have two views. On the first view, (2) is not knowable *a priori*. At most, what is knowable *a priori* is (3) (and even that might not qualify for *a priori* status). If you take this view, you will think that the anti-skeptical argument fails to properly establish its conclusion. On the second view, (2) *is* knowable *a priori*. If one takes

---

completely skeptical of the whole idea of *apriority* have a rough and intuitive grasp of what is meant by calling a sentence *a priori*. It is surprisingly hard, however, to give a precise analysis of *apriority* for sentences. One might try something like: A sentence S is knowable *a priori* for an agent A (at t) just in case the proposition that A intends to express with S (at t) is knowable by A with justification that is independent of experience. This definition would obviously need to be filled out with specifications of what is meant by 'independent of', 'experience', and 'the proposition that . . . . . . intends to express with ————', but the general idea is clear enough. See Chalmers 2002b, f.n. 6, for a similar definition of *apriority* for sentences. There are, of course, other ways that one could go. Stephen Yablo has suggested to me (p.c.) that one could even think of 'a priori' as a sentence operator instead of predicate that applies to sentences. I think that this is worth pursuing, though I will not defend it here.

[3]Experiences will be required in order for one to have the relevant concepts, but these experiences might merely *enable* the relevant belief, but not justify it. (Of course we need to be careful here, for sometimes experiences might play both roles—a point which Scott Soames has made in several places.)

[4]Remarks that Ori Simchen makes in Simchen 2003 suggest that he would take T-sentences to express truths that are members of a rather odd class of *contingent analytic a posteriori* truths, and that he would take me to be confusing intuitions of *analyticity* for intuitions of *apriority*. T-sentences like (2) are contingent, he would say, because the string of symbols making up 'I am a BIV' could have been used to express anything, or nothing at all. Their analyticity follows from his demonstrative understanding of the linguistic device of quotation. (2) would be analyzed as something like 'Any of my utterances that sametoken that[the subsentence following 'iff' in the sentence-token I am currently employing] are true iff I am a BIV' (where the bracketed material is to be understood as an accompanying demonstration, not as part of the sentence proper). Simchen claims that the particular kind of self-reference in such sentences guarantees truth. But he is careful to note that the analyticity of (2) does not mean that the claim is also *a priori*. The issue seems to be the speaker's reliance on sense perception and memory to understand the sentence. His reasoning seems to me to ignore the distinction between enabling and justifying experiences. But, in any event, one might worry that any real or apparent *aposteriority* is an artifact of his rather suspect demonstrative understanding of the device of quotation.

this view, it looks like we *do* have a route to the anti-skeptical conclusion. One will either have to admit the conclusion, or show that the argument goes off the rails at some earlier stage.

1.1.2 Male sisters?

On to the second debate. This one will require a little more setup. I hope the reader will bear with me as I try to negotiate between the competing demands of being relevant and avoiding obscurity.

Kripke's *Naming and Necessity* (1972) dealt a heavy blow. It was no longer philosophically acceptable to assume—at least without doing some serious explaining— that epistemic notions like *conceivability* and *apriority* went hand-in-hand with metaphysical notions like *possibility* and *necessity*. To take this line, one now had the formidable task of explaining (or explaining away) Kripke's examples of the *necessary a posteriori* and *contingent a priori*.

One of the main attractions of the view that has come to be called "two-dimensionalism" is that it promises to give an elegant treatment of Kripke's examples while also giving us the resources to reconcile epistemic and modal notions. How does it propose to do this? Let's consider one of Kripke's examples of the *necessary a posteriori*: 'water is H_2O'. Kripke claimed that natural kind terms like 'water' and 'H_2O' function as *rigid designators*—that is, they pick out with respect to *every* possible world state whatever they pick out with respect to the *actual* world state. Given this definition, any identity statement $\ulcorner s$ is $t\urcorner$ involving only rigid designators is going to be either necessarily true or necessarily false. Since science has taught us that the term 'water' picks out the natural kind H_2O (and it's obvious that 'H_2O' does as well), 'water is H_2O' is both true and necessary. However, it does *not* appear to be *a priori*. We seem to be able to conceive of a scenario where water would have turned out *not* to be H_2O, and it is only through a significant, empirical, discovery that we learn that this scenario is impossible. *Necessity*, we are invited to conclude, doesn't always go with *apriority*, and *conceivability* doesn't always go with *possibility*.

Without delving into the gory details, the basic idea behind two-dimensionalism is

14

that every sentence can be associated with two distinct sets of truth-conditions—two functions from possible worlds to truth-values.[5] The key difference between the two functions is in how we think of the possible worlds which serve as their inputs. We can think of these worlds either as counterfactual possibilities (roughly, ways the world could have been), or as counter*actual* possibilities (roughly, ways the world could be). What David Chalmers has called the *primary intension* of a sentence is the function that results when the worlds are considered as counter*actual*. The *secondary intension*, on the other hand, is what results when the worlds are considered as counter*factual*.[6]

When confronted with a case of the necessary *a posteriori* like 'water is H_2O', the two-dimensionalist can say that, although this sentence has a necessary secondary intension (all worlds are mapped onto the value *true* when they are considered as counterfactual), it has a contingent primary intension (there are some worlds which, when considered as counteractual, will get mapped onto the value *false*). When we evaluate a sentence for *necessity* or *contingency*, the sentence's counterfactual profile runs the show; when we evaluate a sentence for *apriority* or *aposteriority*, the sentence's counter*actual* profile takes over (a claim will be *a priori* just in case it has a necessary primary intension). When we conceive of a coherent scenario where water would have turned out to be some non-H_2O chemical (XYZ, for instance[7]), we are thinking of this world as counter*actual*, not counterfactual, so such scenarios do not threaten the *necessary* status of the claim that water is H_2O. Thus, Kripke's examples do nothing to suggest that *apriority* doesn't go along with necessity-of-primary-intension *(1-necessity)*, or that *conceivability* doesn't go along with possibility-of-primary-intension *(1-possibility)*.

---

[5]A parallel claim is made with respect to the subsentential constituents of sentences, but I will focus on sentences, to keep things simple.

[6]Chalmers has written an enormous amount on the topic of primary and secondary intensions. See Chalmers 1996 and 2002a for particularly good introductions to the terrain. He has also written a number of more recent articles, such as Chalmers 2004, which helpfully distinguish between the many ways one can construct a two-dimensional account. Frank Jackson (1998) also has a primary/secondary intension distinction, but he calls primary intensions *A-intensions* and secondary intensions *C-intensions*. (As in all of the best children's books, 'A' is for *actual* and 'C' is for *counterfactual*.)

[7]See Putnam 1996.

So far so good, but how, exactly, do we consider worlds as counterfactual and counteractual for the purpose of determining the primary and secondary intensions of a given sentence? What we would like is a method, for each kind of intension, for thinking about the worlds in the right mode. Happily, there is pretty widespread agreement concerning how to treat a world w as counterfactual. All we have to do is ask whether the following is satisfied:

[SIE][8] If W had been the case, it would have been that S.

(Where 'W' is replaced with a name or complete description of w and 'S' is replaced with the sentence in question.) If SIE is true, then the secondary intension of the sentence maps w onto the value *true*. If SIE is false, then the secondary intension maps w onto *false*. The basic idea is that when we evaluate claims about a world using a "subjunctive" or "counterfactual" conditional, we automatically treat the worlds as counterfactual.[9]

How to treat a world as counter*actual*, however, is a more controversial matter. Chalmers has, at various times, suggested that we could perhaps simply use indicative conditionals. The idea is this: In order to consider a world w as counteractual, for the purpose of evaluating the primary intension of a given sentence, we ask whether the following is satisfied:

[PIE][10] If W is actual, then S.

If PIE is true, the primary intension of the sentence maps the world w onto the value *true*; if PIE is false, the primary intension maps w onto *false*.

Stephen Yablo (2002) gives an argument that suggests that PIE is not going to help the two-dimensionalist reconcile *apriority* and *necessity*, for while some claims are plausibly *a priori*, absolutely *no* claim is going to have a necessary primary intension, when primary intensions are evaluated according to PIE. We can always get worlds

---

[8]SIE = Secondary Intension Evaluator.

[9]I put the labels 'subjunctive' and 'counterfactual' in scare quotes because, for familiar reasons, neither is a very happy way of naming the relevant class of conditionals (see Chapter 1 of Bennett 2003 for details). Unfortunately, I don't have a better label in hand.

[10]PIE = Primary Intension Evaluator.

"into" the primary intension of a sentence by considering worlds characterized by certain deviant shifts in the meanings of our words. The following conditional looks true:

(4) If 'sister' actually refers to male siblings, then sisters are male (not female) siblings.

Using PIE to evaluate the primary intension of 'sisters are female', then, it looks like we are going to have to map the world where 'sister' refers to male siblings onto the value *false*. But 'sisters are female' is surely a paradigm case of the *a priori*—it's *analytic*, after all.

Chalmers is ambivalent about this argument, but he thinks that, whether or not it is sound, it doesn't affect his more settled views concerning primary intensions. He simply replaces PIE with PIE':

[PIE'] It is *a priori* that if W is actual, then S.[11]

And Chalmers thinks that whether or not (4) (the male sisters conditional) is *true*, it is certainly not *a priori*. In order to get from the antecedent to the consequent *a priori*, one would need to be able to go *a priori* from claims about the word 'sister' to claims about sisters. But, it is a "substantive, *a posteriori*, metalinguistic" fact that 'sister' refers to sisters (or anything else for that matter) (2002a, p. 170).

I myself, however, feel a strong pull to the claim that (4) is *a priori*. The antecedent seems to have built-in all of the information one could need with respect to the reference of the word 'sister' in order to be able to draw the conclusion that is expressed by the consequent. At any rate, it is certainly far from obvious that Chalmers is right to say that (4) is *a posteriori*. And the prospects for an account of primary intensions based on PIE' ride on what we should say about this kind of case.

---

[11]This time, the schematic letter 'W' will have to be replaced by what Chalmers calls a "canonical description" of *w*.

1.1.3 Seeking (*a priori*) closure

Our third debate is connected to the second. Thinking about what to say about the debate described in the last section can lead one to a rather striking puzzle: One source of apparently *a priori* claims is so-called *analytic* truths, such as:

(5) Sisters are female siblings.

A second source of arguably *a priori* claims is the R-schema (" '*t*' refers to (all and only) *t*s"). So, for instance, we have:

(6) 'Sister' refers to (all and only) sisters.

Let's grant that (5) and (6) are both *a priori*. The problem is that, together, they entail something that does not look at all *a priori*:

(7) 'Sister' refers to female siblings.

Our knowledge of this is surely substantive and empirical. *A priori* reflection alone cannot lead us to it. Otherwise, it looks like we could learn huge chunks of the English language *a priori*, and this surely isn't a credible claim.

One might reasonably have expected *apriority* to be closed under *a priori* entailment—if *p* and *q* are both *a priori*, and they *a priori* entail *r*, we would expect that *r* would also be knowable *a priori*. But some of us are strongly inclined to think (5) and (6) are both *a priori*, and they certainly have the appearance of *a priori* entailing (7), which manifestly isn't *a priori*.

We have already seen, in the last section, how Chalmers would respond to this kind of puzzle: he would deny that (6) is *a priori*. In a discussion of this puzzle in Yablo 2002, Stephen Yablo takes a different way out. He simply denies that *apriority* is closed under *a priori* entailment (he even sketches a notion of *apriority* for which we would not expect closure to hold). As for myself, I think that *apriority* must be closed under *a priori* entailment. I also think that while (6) is not *a priori* as it stands, a suitably hedged restatement of it probably is.

The reason it is not *a priori* as it stands is that there are plenty of instances of the R-schema that aren't even true (or truth evaluable, for that matter). Consider " 'is' refers to iss" and " 'flimflamaphone' refers to flimflamaphones".

There is, however, a straightforward revision we can make to the R-schema in order to skirt around this last problem. The trick is to state the schema as a conditional, where we load into the antecedent enough to weed out the problem cases. The revised schema will look something like this:[12]

> [R'] If 't' is a (singular count) noun (pluralizable with 's'[13] in the language I am currently speaking), then 't' refers to ts.

I have a very strong intuition that instances of this schema are *a priori*. If this is right, however, we are still left with the closure puzzle, for we can combine the R'-sentence for 'sister',

> (8) If 'sister' is a (singular count) noun (pluralizable with 's' in the language I am currently speaking), then 'sister' refers to sisters,

with (5) ('sisters are female siblings') to get:

> (9) If 'sister' is a (singular count) noun (pluralizable with 's' in the language I am currently speaking), then 'sister' refers to female siblings.

This looks every bit as *a posteriori* as (7). So, if we claim that (8) is *a priori*, we must either deny closure, with Yablo, or find some other way to block the inference to (9).

1.2 The Ways of Indicatives

Let's collect together and rename our contested conditionals. From §1.1.1, we have instances of the T-schema, such as (2). From §1.1.2, we have the troubling 'sister' conditional, (4). Finally, from §1.1.3, we have instances of the hedged R'-schema,

---

[12]Thanks to Selim Berker and the MATTI reading group for helping me state (R') more precisely than I otherwise would have thought to.

[13]That is, pluralizable by affixing the letter 's' to the end of the term.

such as (8). To make things memorable, let's rename these (TBIV), (SISM), and (RSIS), respectively:

(TBIV) My utterances of 'I am a BIV' are true iff I am a BIV.

(SISM) If 'sister' actually refers to male siblings, then sisters are male (not female) siblings.

(RSIS) If 'sister' is a (singular count) noun (pluralizable with 's' in the language I am currently speaking), then 'sister' refers to sisters.[14]

I want to explore what we should say about the epistemic status of these conditionals. In order to do this, though, we need to get clearer on what these conditionals *mean*. I will start by briefly discussing a theory which would understand them all as material conditionals.

But, first, a confession: In what follows I am going to be assuming (without argument) that indicative conditionals do, in fact, have truth conditions, and that our considered intuitions about which ones are true give us some evidence for what those truth conditions can and cannot be. I should say, however, that not everyone agrees that indicative conditionals are truth apt. Dorothy Edgington (1986), Jonathan Bennett (2003), and others have argued that indicatives do not have truth conditions. I think that there are strong arguments against this view. Chapter 4 of William Lycan's *Real Conditionals* (2001), for example, gives a laundry list of grievances against the "no-truth-value thesis". I think his considerations, and others besides, combine to make a compelling case.

1.2.1 Indicatives aren't material conditionals

Assuming that indicatives do, in fact, have truth conditions, what might these truth conditions be? One view is that indicatives conditionals are just material conditionals.[15] Perhaps the world would be a nicer and neater place if they were right. But, unfortunately, the ordinary indicative conditionals we use in everyday conversation

---

[14]In the following, I will often abbreviate this cumbersome construction '(singular count) noun (pluralizable with 's' in the language I am currently speaking)' as 'NOUN'.

[15]This has been defended by H.P. Grice (1989) and Frank Jackson (1987), among others.

are *not* material conditionals. The difficulties with the "equivalence thesis" are familiar, and I won't rehearse them all here. However, I would like to highlight a few issues.

The first place one might be tempted to look for trouble is in the so-called *paradoxes of material implication*. These involve sentences like:

(10) If Mars has six moons, then I have six moons.

(11) If Mars is made of port wine cheese, then George W. Bush is President.

If the equivalence theory is right, (10) has to be counted as *true*, since it has a false antecedent (Mars has only two moons), and (11) must also be counted as true, for it has a true consequent. However, these look like very unhappy conditionals!

Grice (1989) famously argued that the *unhappiness* here is not to be confused with *falsehood*. (10) and (11) are indeed *true*, he argued, but they fail to accord with some pragmatic conversational principles, and for this reason, utterances of them tend to come off badly. For example, (assuming that the equivalence thesis is right), the only reason why I think that (10) is true is because its antecedent is false. But this being the case, I really shouldn't go around saying (10). For one thing, there is something stronger that I believe and take myself to have just as much evidence for, i.e. that Mars does not have six moons. Not only is this logically stronger than (10), it is also *shorter*. An utterance of (10) by me in a conversational context is thus likely to be misleading. We have a general expectation, in most conversational contexts, that our interlocuter is not deliberately saying something that is both more verbose, and also weaker, than what she takes herself to have evidence for. In fact, because the conversational rules are such as they are, when we assert a conditional, we *conversationally implicate* that we do not take ourselves to be in an epistemic position to deny the antecedent.

So if I were to utter (10) in a normal conversational context, what I would literally say would be true, but my utterance would carry with it an implicature that is false. Unfortunately, it is easy to confuse *true utterances with false implicatures* with *false*

utterances, and this, Grice says, is where we get our inclination to evaluate (10) as false. A similar pragmatic story is told for (11).

Grice's story, elegant as it is, lacks plausibility for certain conditionals. Notice, for instance, that any statement of the form 'if p, then not p' is going to come out true on the equivalence account whenever the antecedent is false. So the following is true:

(12) If Grice was just a philosophical hack, then Grice *wasn't* just a philosophical hack.

Perhaps we could have learned to live with (10) and (11) being true, but I, for one, refuse to accept (12) as true (no matter how many false implicatures we assign to it to soften the blow).[16] I think that most of us who aren't trying to save a theory think that 'if p, then not p'-conditionals are one and all false.[17]

Here's an equally strange result: If the equivalence thesis is right, then the sentence 'It's not the case that if Grice was just a philosophical hack, then Grice was not just a philosophical hack' entails that Grice *was* just a philosophical hack (since $\neg(P \supset Q)$ entails P). It is hard to see how implicatures could be invoked to save the truth of *this* claim.

---

[16]Jackson (1987) differs from Grice with respect to what kind of implicature he thinks is in play with the "true but unhappy" indicatives. Whereas Grice relies on conversational implicatures (calculable from literal content plus some maxims governing conversation), Jackson thinks it better to go the route of *conventional* implicatures (implicatures that are, in some sense, part of semantic content). I don't have the space to fully flesh out Jackson's view, but let it suffice to say that it too must treat (12) as literally true, and this is a result that I think we should refuse to swallow, even to gain elegance and theoretical economy.

[17]But if that is so, what are we to do with Russell's famous argument against naïve realism (and others like it)?

> Naïve realism leads to physics, and physics, if true, shows that naïve realism is false.
> Therefore naïve realism, if true, is false; therefore it is false. (1950, p. 15)

There are two things we could say here: (1) Russell is a logician. When logicians construct proofs involving conditional claims, they often assume that the conditionals in question are material conditionals, and exploit this assumption throughout the proof. And it is uncontroversial that *material* conditionals of the form 'If p, then not p' can be true. It would probably do no violence to Russell's argument to understand him as meaning to employ material conditionals. (2) Alternatively, one could understand the argument as a *reductio ad absurdum*. He starts with the assumption that naïve realism is true, and derives a claim that expresses something that is not only false, but impossible ('If naïve realism is true, then it is false'). We're invited to conclude that the starting assumption must be false.

22

1.2.2 Indicatives aren't Material Conditionals II

Another interesting difficulty with the equivalence account—one that has received less attention—is that it has a hard time explaining why indicative conditionals appear to be *opaque contexts*, in Quine's sense.[18] In opaque contexts, coreferential substitutions cannot be counted on to preserve truth. Belief attribution contexts (e.g., 'X believes that ......') are paradigm examples. We cannot count on switching 'Declan Patrick Aloysius MacManus' for 'Elvis Costello' in 'Darin believes that Elvis Costello recorded the ground-breaking album *My Aim is True*' and preserving truth (even though Elvis Costello just *is* Declan Patrick Aloysius MacManus). Interestingly, the indicative conditional context, 'If X, then ......' appears to suffer from this same kind of opacity. Consider the following two conditionals:

(13) If Hesperus appearances end up being caused by Mars, then Hesperus is Mars.

(14) If Hesperus appearances end up being caused by Mars, then Venus is Mars.

Now, (13) is uncontroversially true, and (14) is just the result of taking (13), deleting the word 'Hesperus' in the consequent, and replacing it with the coreferential term 'Venus'. But (14) looks, at least to me, to be false. If this intuition is right, then (at least some) 'If X, then ......'-contexts are going to be opaque.

I will have more to say about this in a bit. For now, we just need to appreciate that this does not fit well with the equivalence theory. Combining two observations will make clear why: First, remember that a material conditional is supposed to be truth-functional. Feed it the truth value of the antecedent and the truth value of the consequent and it churns out a resulting truth-value. It doesn't care how we got to the truth values that we used as inputs—it is blind to everything but alethic valence. Importantly, this means that material conditionals are insensitive to any sort of interaction between the antecedent and the consequent.[19] Our sentence (12)

---

[18]"We may speak of a context as *referentially opaque* when, by putting a statement ϕ into that context, we can cause a purely referential occurrence in ϕ to be not purely referential in the whole context" (1966, p. 160).

[19]I'll sharpen up this claim in a bit.

('If Grice was just a philosophical hack, then Grice *wasn't* just a philosophical hack') was a striking illustration of this. If we understand this as a material conditional, we have to ignore the fact that the consequent is the denial of the antecedent. We just check to see if the antecedent is false, or the consequent is true. If either of these is the case, we must call the conditional true; otherwise, we must call it false.

Second, if we take on board Kripke's claim that proper names rigidly designate their referents (as we should), we will want to say that Hesperus is Mars iff Venus is Mars (since Hesperus just *is* Venus). Given this observation, we must conclude that the truth value of the consequent of (13) (taken by itself) is going to be the same as the truth value of the consequent of (14) (taken by itself). And the same is obviously true of their antecedents (since their antecedents are the same).

Combining our two observations, if we understand (13) and (14) as material conditionals, we must either count both as true or count both as false. Material conditionals are functions of truth values, and, given Kripke's observations, we're going to have the same inputs for (13) and (14). There's no room here for opacity.[20]

1.3 Playground Conditionals

As I noted in the last section, indicative conditionals differ importantly from material conditionals in that the truth-value of an indicative is (at least sometimes) sensitive to the relationship between its antecedent and its consequent. Something like the Grice sentence (12) can be true when understood as a material conditional, but given the relationship between its antecedent and consequent, it is bound to be false when understood as an indicative conditional.

There is a certain group of conditionals that make this point even more vivid. Some of us had our first encounter with them on the playground. Angry little Billy said something like, "I hate X, she's a jerk." Then a classmate kindly reminded Billy that today was "opposite day" (where assertions are evaluated as having the truth-

---

[20]The antecedent or consequent of a material conditional might contain—or even *be*—an opaque context, but this doesn't change the fact that, when understood materially, 'If X, then . . . . . . ' can't *itself* be opaque.

conditions that their negations standardly have). Not sure whether to believe this news, Billy responded, "if today is really opposite day, then I love X." Obviously, Billy intended that if the antecedent of his conditional was true, the consequent was to take on a non-standard meaning. Let's call conditionals like this *playground conditionals*. Here's another example: Billy says to Mary, "Yeah, you're real smart, if 'smart' means stupid!"

It will be useful to have a sort of rough, working characterization of playground conditionals in place. Something like the following will suffice for my purposes:

[PG] $\ulcorner A \rightarrow C \urcorner$ is a playground conditional iff the supposition that A is true changes (or constrains) the semantic contribution that C makes to the conditional.

This definition isn't perfect, but I think that it serves to set the right intuitive boundaries.

Now, we cannot understand playground conditionals as material conditionals. Take Billy's first conditional:

(15) If today is opposite day, then I love X.

Let's assume that it is not opposite day. Notice that, understood as a material conditional, (15) will be true if Billy loves X. If the consequent of a material conditional is true, after all, it just doesn't matter what the antecedent says, the material conditional is going to churn out the value *true*. But Billy's utterance is surely *false* if he loves X. Indeed, it is absolutely crucial to understanding Billy's utterance that one realize this. If Billy loves X, he is guilty of lying to us. Everyone on the playground understand this—it would take a philosopher to deny it.

No doubt playground conditionals are a strange bunch. But I think that they show us something important about what we can do (and what we in fact quite often do) with indicative conditionals. In particular, they show us how the antecedent of an indicative conditional can play havoc with the meaning of its consequent. Or, to put things more carefully, they show us that the antecedent of a conditional can change the way the consequent (or certain words in the consequent) contributes to

the overall meaning of the conditional. It's easy to see that 'I love X' functions very differently in (15) than it does in (16):

(16) If it was X who wrote this romantic letter to me, then I love X.

In (15), we have a supposition—that today is opposite day—and then a claim made under that supposition. What claim is made? The claim is roughly this: *it's not the case that the speaker loves X*. This claim is made using a form of words that would, under normal circumstances, be used to make its negation. But since these words fall under the scope of a supposition—one that changes how words are used—all is well. In (16), on the other hand, we have a more conventional antecedent, which does not concern the meanings of words, so the contribution that 'I love X' makes in this context is the standard one.

I will now return to the question of the epistemic status of the contested conditionals that I identified in §1.1. I'll argue that each of them is (or involves) a playground conditional, and then try to motivate the claim that, when they are recognized as such, we see that they are, indeed, *a priori*.[21] I'll also say, just briefly, how this affects each of the philosophical debates that I sketched in §1.1.

1.4 The argument

1.4.1 In skeletal form

Here is, in the most skeletal form, how I think our claims are connected: First, for the sake of argument, assume that *apriority* is closed under *a priori* entailment (I will come back to this assumption at the end of the argument). Now ask about the epistemic status of (SISM).

(SISM) If 'sister' actually refers to male siblings, then sisters are male (not female) siblings.

---

[21]Well, we'll see that the situation with the biconditional (TBIV) is a little more complicated, and that it isn't, strictly speaking, *a priori* But a closely related set of conditionals are both *a priori*.

We saw in §1.1.2 that Chalmers, for his part, takes it to be knowable only *a posteriori*, for he thinks that you can't get *a priori* from claims about the word 'sister' to claims about sisters. However, if (RSIS) were true and *a priori*, then (given closure) we *could* get to (SISM) *a priori*.

(RSIS) If 'sister' is a NOUN, then 'sister' refers to sisters.

So is (RSIS) knowable *a priori*? It looks like it to me, but someone could object that all that is really *a priori* is that *if 'sister' is a* NOUN, *then " 'sister' refers to sisters"* *is true*.[22] But if T-sentences, such as (TBIV), are *a priori*, then this wouldn't be a problem. We could simply disquote to get the desired result. Actually, a T-sentence is stronger than what we need—all that is really needed for disquotational purposes is the left-to-right direction of the relevant T-sentence.

To sum up, then, if we could establish that left-to-right direction of T-sentences like (TBIV) are knowable *a priori*, then we'd be in a position to argue that (RSIS) is *a priori*. And if we could establish that (RSIS) is knowable *a priori*, we would be in a position to argue that (SISM) is also *a priori*.

All of this, however, depends on *apriority* being closed under *a priori* entailment. And we have already seen, in §1.1.3, that this view runs into a serious puzzle, especially if we take the view that claims like (RSIS) are *a priori*. In §1.4.5, I will argue that paying close attention to opacity phenomena in playground conditionals is the key to saving closure from this puzzle.

Now that we have a feel for how the argument is supposed to go, let's dive in. I will spend most of my time on (TBIV), since this is the heart of the argument, and the rest falls into place fairly easily if we can establish the desired result here.

1.4.2 (TBIV)

The waters get deep quickly, for the matter of (TBIV)'s epistemic status is a bit complicated. (TBIV) is a biconditional, so it encodes two conditional claims. Let's

---

[22]We already saw in §1.1.1 that Anthony Brueckner (1986) makes this kind of move against the *apriority* of T-sentences. He also makes the kind of move that I'm suggesting here, which undermines the *apriority* of instances of our hedged R-schema.

27

call the left-to-right direction '(TBIV→)' and the right-to-left direction '(TBIV←)'. So we have:

(TBIV→) If my utterances of 'I am a BIV' are true, then I am a BIV.

(TBIV←) If I am a BIV, then my utterances of 'I am a BIV' are true.

For the next couple of subsections I will focus on (TBIV→). I'll argue that it is a true and *a priori* playground conditional. I will then turn to (TBIV←). This, I will claim, is not knowable *a priori*, and perhaps isn't even true. This is not a problem for the argument I sketched in the last section (since only sentences like (TBIV→) played an essential role there). However, it might look like it *is* a problem for Brueckner's opponent in the debate over whether the Putnam-style anti-skeptical argument can establish its anti-skeptical conclusion (see §1.1.1). The issue there was whether we can get from the claim that my utterances of 'I am a BIV' are false to the claim that I am not a BIV. To bridge that gap, it looks like like we need (TBIV←).[23] I'll have more to say about this in a bit.

(TBIV→)

(TBIV→) is a playground conditional. Its antecedent sets up the supposition that the speaker's utterances of 'I am a BIV' are true. Since the consequent of the conditional is itself a token of 'I am a BIV', it is a consequence of this supposition that the consequent of the conditional must be true. So we know that, on the supposition set up by the antecedent, the consequent is going to be true. Since this reasoning can be done *a priori*, it looks like (TBIV→) is both true and *a priori*.

But we have already seen a form of objection that could be used against this kind of argument. Chalmers and Brueckner would object that my knowledge that the consequent of the conditional is true under the assumption set up by the antecedent isn't enough to let me know that *if my utterances of 'I am a BIV' are true, then I am a BIV*. At most, it allows me to know that the sentence (TBIV→) is true. But, again, knowing that (TBIV→) is true isn't the same as knowing the proposition that (TBIV→) expresses.

---

[23]If it's not already clear why this is what we would need, consider the contrapositive of (TBIV←).

28

Chalmers and Brueckner are clearly right to say that knowing that a sentence is true isn't always sufficient for knowing the proposition that it expresses. X might come to know, for example, on the basis of trustworthy testimony, that the German sentence 'Schnee ist weiß.' expresses a truth—and X might know this while being ignorant of the fact that snow is white. This much is obvious.

But just how this obvious truth connects with the current debate is unclear. In the German example, the fact that X might not be a German speaker is clearly what is doing the bulk of the work in producing our intuition that X can know that the sentence is true without knowing the proposition it expresses. If, on the other hand, X *does* understand German, it is not at all obvious how it can be the case that she knows that 'Schnee ist weiß' is true yet fails to know that snow is white.

The situation that I am in with respect to the sentence (TBIV$\rightarrow$) seems to be quite unlike the situation X is in with respect to the German sentence she takes to be true. My utterances of 'I am a BIV' are made in a language of which I am a life-long native speaker—a language that I take myself to understand. If I am right that I understand the language that I am speaking, what is to stop me from knowing without further ado that if my utterances of 'I am a BIV' are true, then I am a BIV?

Brueckner would reply here that, given the skeptical challenge, I *don't* know which language I am speaking, so I don't know what I mean by the sentence 'I am a BIV'. Given this, knowing that my utterances of 'I am a BIV' are true would *not*, by itself, permit me to conclude that I am a BIV. I can't draw this conclusion because I don't know what conclusion I would thereby be drawing.

But this is a suspicious line of reasoning. Brueckner's claim that I don't know what I mean by my utterances of 'I am a BIV' is supposed to follow from Putnam's *semantic externalism* plus the possibility that the skeptical hypothesis is true. Semantic externalism says that meanings "ain't in the head"; content is partially determined by a speaker or thinker's natural and social environment. Notice that if we were not assuming a kind of semantic externalism, we would have gotten off the boat long before admitting that we don't know what we mean by 'I am a BIV'.

Given that we are assuming semantic externalism, it would seem strange and out

of place to also impose some kind of internalist constraint on *understanding*. If we are going to accept semantic externalism, we should be externalists "all the way down". That is, we should be externalists not just about the content of sentences, but also about the contents of the thoughts that we use these sentences to express. If we embrace this kind of thoroughgoing externalism, it seems that I *do* know what 'I am a BIV' means; it means that I am a BIV. The thought that I express with the sentence 'I am a BIV' is guaranteed to have the same externally individuated content as the sentence itself.[24] Brueckner seems to think that this is not good enough for *knowing the proposition expressed by the sentence*. But it is unclear what more he is demanding here, or if he is right to demand it.

Knowing which proposition is expressed

I want to explore this issue further by considering a similar dialectic that plays out with respect to Kripke's argument for the contingent *a priori*. Kripke's argument goes something like this: Suppose the proper name 'Lee' is introduced into the language to stand for the youngest Chinese spy.[25] Now, as we have said before, it seems plausible that names are rigid designators. Descriptions, however, are *not* rigid designators, for they pick out different things with respect to different worlds. With respect to our world, 'the youngest Chinese spy' picks some some individual *i*. If we now consider a nearby world where *i* is stillborn, there will still be someone there who is *the youngest Chinese spy*, but it will not be *i*. Given these facts, Kripke thinks that the sentence 'Lee is the youngest Chinese spy (if anyone is)' will express a proposition that is *contingent*, but also knowable *a priori*. It is contingent because there are worlds where Lee didn't get into the spying business. It is *a priori* because we know that *our* world isn't one of them. How do we know that? We introduced the name 'Lee' by *stipulating* that it was to refer to whoever the description actually refers to. This guarantees that, at least in our world, 'Lee' and 'the youngest Chinese spy' are coreferential.

---

[24]Similar points are made in Burge 1998a and Gibbons 1996.

[25]This particular example is David Kaplan's (Kaplan 1991).

In Soames Soames 2003 and 2005, Scott Soames indicates that he finds Kripke's argument unconvincing. Assume that I have never had contact with any Chinese spy, but I do know, on general grounds, that China has spies. Against this background, I introduce the name 'Lee' to refer to the youngest Chinese spy. Soames says that if one does this,

> one doesn't know, or believe, of any individual i that i is the youngest Chinese spy. Clearly the mere performance of a linguistic ceremony of using the description to introduce a name can't change this. If, despite this, one's introduction of the name 'Lee' for i is successful, then ['Lee is the youngest Chinese spy (if anyone is)'] will come to express a proposition about i that one doesn't know to be true, and can't come to know except by gaining *de re* knowledge of i through further empirical investigation. (2005, p. 56)

Soames is inclined to think that I don't even *understand* this name 'Lee' (a position that I think is untenable). However, he thinks that, even if I can be said to understand the name, I still don't know that Lee is the youngest Chinese spy (if anyone is), for, in order to *know the singular proposition* expressed by 'Lee is the youngest Chinese spy', one must have some *de re* knowledge of Lee.[26] More generally:

> [DE RE] If a sentence S expresses a singular proposition p involving an individual i, knowing that S is true doesn't, by itself, guarantee that one will know that p. In order to know that p, one must know which proposition S expresses, and this requires, at the very least, having some kind of *de re* knowledge of i.

At various points, Soames suggests that this *de re* knowledge requires 'sufficient contact' with the relevant individual and that sufficient contact involves some kind of perceptual, or empirical, knowledge of, or *acquaintance with* her.

Against this, I suggest that Soames, and other like-minded theorists—Evans (1985) and Pryor, for example—(understandably) take *the normal way* of having *de re* beliefs about everyday physical things for *the only way* of having *de re* beliefs about them. Suppose Al is a very skilled meteorologist. Al and Sue are standing outside having a long discussion concerning whether or not tonight's Red Sox game will

---

[26] James Pryor, in his (as yet) unpublished paper, "Indexicality and Aprioricity" (2004), makes similar claims about sentences with indexicals and demonstratives.

be played as scheduled. Al, taking note of recent temperature and pressure changes (or whatever), knows that, even as they are speaking, there is a large, dark, rain cloud forming behind him. Pointing behind him (without looking), he says, "Let's call that rain cloud 'Bob'. If Bob settles over Fenway Park, the game will be canceled for sure". Does Al know which proposition he has expressed? Surely he does. He isn't picking out the rain cloud in his thought the way most of us would, but I don't see why this bars him from having *de re* beliefs about it.

I agree that *there is a sense* in which Kripke's name stipulator doesn't know exactly which proposition she is expressing. And, to return to the issue at hand, I agree with Brueckner that *there is a sense* in which I don't know exactly which proposition I express with 'I am a BIV'. But *there is a sense* in which most of us don't know exactly which propositions we are expressing most of the time. Prior to Amedeo Avogadro's 1811 article in the *Journal de Physique*, it was not known that water was H_2O (and of course, Avogadro was largely ignored by his peers, so it wasn't until much later that this knowledge was disseminated to the masses). Roll back the historical clock a little further and it wasn't even understood that water is made of hydrogen and oxygen. So, did Shakespeare know *exactly* which proposition he was expressing with the sentence 'water is wet'? How could he? The word 'water' referred to H_2O just as much then as it does now. But Shakespeare didn't know that. For all he knew, he could have turned out to be on Twin Earth, and on Twin Earth, the sentence 'water is wet' expresses a very *different proposition* than it does in the actual world. So, there's *a sense* in which Shakespeare didn't know what proposition he was expressing with the sentence 'water is wet'. Soames or Brueckner might say that Shakespeare could know that water is wet because he could point at a sample of water and say "this is the stuff I'm talking about. It's wet". But if, for all Shakespeare knows, that stuff he's pointing at could be any of countless different chemicals, how does acquaintance coupled with pointing give him special credentials to know which proposition he expressing? He doesn't know exactly what he's acquainted with or what he's pointing at! How is his pointing better than a "mental pointing" through the use of descriptions?

So, with respect to the sentence 'Lee is the youngest Chinese spy', I'm inclined to say the following: There are many different ways to "know which proposition is expressed" by a sentence, for there are lots of ways to pick things out. We can pick them out through a particular sense experience, or a collection thereof, or memories of sense experience—but we can also pick out things through descriptions. What principled reason is there for thinking that only the former can give us the kind of knowledge that we need in order to know singular propositions or know which proposition is expressed by a sentence involving a proper name? Robin Jeshion, in Jeshion 2002, argues—quite convincingly, I think—that, "What distinguishes *de re* thought is its structural or organizational role in thought; acquaintance, and any evidential or epistemic relation, is inessential" (p. 67). *De re* thought, she argues, involves the "mental tokening of a name" (p. 67ff), where names are, "vehicles for thinking about objects in a way that requires no particular mode of presentation of the referent" (p. 64). On this view, *de re* knowledge doesn't require a certain kind of external causal relation to the relevant individual, it simply involves thinking about the individual in a certain way.

With respect to the sentence 'I am a BIV', I am inclined to say the following: *Knowing which proposition is expressed* comes in degrees. I think that I understand what is meant by the sentence " 'boson' refers to bosons", even though I know almost nothing about bosons. Indeed, I think that I *know which proposition this sentence expresses*, even if this knowledge isn't as robust as it could be. Shakespeare knew which proposition he expressed with 'water is wet', even though he didn't know as much about water (or wetness, for that matter) as we do.

The (perhaps very thin) understanding I have of 'I am a BIV' is sufficient for me to know (with some degree of robustness) which proposition is expressed by 'I am a not a BIV'. There are some lineups that I could pick the proposition out of. Put it next to the proposition that $2 + 2 = 4$ and the proposition that *snow is white*, and I will be able to pick it out just fine as the proposition expressed by 'I am not a BIV'. There are other lineups out of which I wouldn't be able to pick the proposition. But this is the situation I usually find myself in with respect to the sentences I understand.

(TBIV←)

So far, then, we have been given no convincing reason for thinking that (TBIV→) isn't *a priori*. (TBIV←), on the other hand, is *not* going to be *a priori*. It says that if I am a BIV then my utterances of 'I am a BIV' are true. Putnam's argument from the causal theory of reference indicates that this isn't even going to be *true*, let alone *a priori*. Assume that I am a BIV. Then when I take myself to be tokening the term 'BIV', what am I referring to? It can't be *brains* in *vats*, for, *ex hypothesi*, there aren't the necessary causal chains connecting my use of the terms 'brain' and 'vat' to brains and vats. It is unclear what I *am* referring to with these terms if I am a brain in a vat—perhaps certain kinds of sense impressions, or features of the computer program that is responsible for stimulating my brain. But on either of these accounts, it is clear that my utterances of 'I am a BIV' are going to be false. Under the supposition that I am a brain in a vat, I am *not* a kind of sense impression, nor am I a feature of a computer program. Other candidates for the reference of my terms (that are consistent with both the hypothesis that I'm a BIV and the causal theory of reference) seem to lead to the same conclusion: If I am a BIV, then my utterances of 'I am a BIV' are *not* true, they're false.

So, if we accept the causal theory of reference, it seems that we are led to conclude that (TBIV←) is false. Again, this doesn't affect the argument that I outlined in §1.4.1. However, it might look like it does settle the debate between Brueckner and the anti-skeptic in favor of Brueckner. Remember that Brueckner agreed with Putnam that whether I am speaking English or Vat-English, my utterances of 'I am a BIV' are false. The anti-skeptic held that this fact, taken together with the relevant T-sentence, was sufficient to allow me to conclude that I am not a BIV. Brueckner argued that it was not. Who is right? Well, we have just seen that it's false that if I am a BIV, then my utterances of 'I am a BIV' are true. If contrapositives are logically equivalent (as we always insist to our undergraduate students), then the following must also be false:

(17) If it's not the case that my utterances of 'I am a BIV' are true, then it's not the case that I am a BIV.

Assuming bivalence, this looks like the very principle that the anti-skeptic needed in order to establish her conclusion.

Luckily for the anti-skeptic, however, indicative conditional contrapositives aren't always logically equivalent.[27] Playground conditionals are good illustrations of this. Consider the following two sentences, as uttered by Billy to his classmate:

(18) If today is opposite day, then I love X.

(19) If I don't love X, then today is not opposite day.

These are not logically equivalent, even though (19) is the contrapositive of (18). Billy accepts (18), but he doesn't accept (19). He knows that he doesn't love X, but he rightly doesn't consider himself to be in a position to *modus ponens* his way to a knowledge that today is not, in fact, opposite day. Mere introspection of his feelings toward X can't possibly put him in a position to know this claim. But it *does* put him in a position to know (18).

The upshot is this: We don't have to reject (17) as false just because we rejected its contrapositive, (TBIV←). Indeed, it should be perfectly obvious that we have no business rejecting (17). If (TBIV→) is a true and *a priori* playground conditional, as I've argued it is, surely (17) is one as well. The antecedent of (17) sets up a supposition that constrains the meaning that the consequent can contribute to the conditional. The supposition is that it's not the case that my utterances of 'I am a BIV' are true. Assuming bivalence, this is equivalent to supposing that my utterances of 'I am a BIV' are false. Since the consequent is an instance of this very sentence type, prefixed with the negation operator, the consequent must then be true. So it looks like (17) is going to be true and *a priori* for the same reasons that (TBIV→) is. This being the case, I think we have to side with the anti-skeptic in maintaining that since we know *a priori* that our utterances of 'I am a BIV' are false, we also know that we are not BIVs.[28]

---

[27]This has been pointed out by many philosophers—Stalnaker (1975), for instance.

[28]However, I should add that I don't think that arriving at this conclusion squashes skepticism—it just makes it extremely hard to state the skeptical claim in a way that has a shot at truth. We know that we are not BIVs, but the fact that we know this should not be taken to entail that we

1.4.3 From (TBIV→) to (RSIS)

From §1.4.2, then, we have (TBIV→), and other left-to-right directions of T-sentences. And now that we have this, it seems that we can get (RSIS) straightaway.

(RSIS) says that *if 'sister' is a* NOUN, *then 'sister' refers to sisters*. Once again we have a playground conditional. The antecedent sets up the supposition that the word 'sister' is a NOUN. But this being the case, we know that it is the sort of thing that can be plugged into the R-schema to get a truth. We know the latter simply by understanding how the word 'refers' works. It is—or at least one might reasonably argue that it is—constitutive of the meaning of 'refers' that instances of the schema " '*t*' refers to *t*s" always come out true, provided that *t* is replaced with a singular count noun that is pluralizable with 's'. So given the supposition that 'sister' is such a noun, and our understanding of how 'refer'-contexts work, we can immediately conclude that " 'sister' refers to sisters" is true. Having established this, all we need to do is disquote, using:

(TSIS) If " 'sister' refers to sisters" is true, then 'sister' refers to sisters.

Notice that this has the same form as (TBIV→). It is, therefore, true and *a priori* for the the same reason as the latter (see §1.4.2).

1.4.4 From (RSIS) to (SISM)

Now that we have (RSIS), it is but a short step to (SISM). All it takes is to combine (RSIS), which said that *if 'sister' is* NOUN, *then 'sister' refers to sisters*, with the following uncontroversial premise:

(UNPR) If 'sister' is a NOUN and it actually refers to male siblings, then the NOUN 'sister' actually refers to male siblings.

---

are what we think we are. For if skeptical hypotheses are live, we are, in an important sense, in the dark about the nature of BIVs and what it takes to not be one of them. I realize that these claims have a cryptic ring to them. Since it isn't my primary aim to adjudicate the debate over skepticism, I will unfortunately have to leave the dialectic in this less than ideal state. For more on the sort of approach to Putnam's argument expressed here, see Ebbs 1992 and Wright 1992.

This claim is roughly of the form 'if p, then p', and all indicative conditionals of this form are true.

Now, let's put (RSIS) and (UNPR) together. Notice that the antecedent of (UNPR) is strictly stronger than that of (RSIS). This means that the supposition that we are directed to make by the antecedent of (UNPR) includes the supposition that we are directed to make by the antecedent of (RSIS). This being the case, we can combine (UNPR) and (RSIS) to get:

(20) If 'sister' is a NOUN and it actually refers to male siblings, then the NOUN 'sister' refers to (all and only) sisters, and it refers to (all and only) male siblings.

The consequent of (20) obviously entails that the sisters just *are* the male siblings. Thus, we can conclude (SISM). And again, the reasoning here has been *a priori* in nature, so we can conclude that we know (SISM) *a priori*.

Recalling the debate that this conditional was important for, it is now clear that Chalmers cannot use his PIE′ to define primary intensions, at least if he wants there to be 1-necessary claims (as he does). Remember that 1-necessity is *necessity of primary intension*, which means that PIE′ must map every world onto the value true. But, if (SISM) is *a priori*, it looks like the primary intension of 'sisters are female', as evaluated by PIE′, assigns the value *false* to the possible world where we use the 'sister' to refer to male siblings. It is clear that we could pull a meaning shifting trick like this for any sentence, so absolutely *no* sentence is going to be assigned a necessary primary intension by PIE′.

This has the following consequence: If two-dimensionalism is going to be useful for bringing *conceivability* and *apriority* back into the arms of *possibility* and *necessity*, some other definition of *primary intension* is going to have to be used. Only time will tell if there is a definition that can do the work the two-dimensionalist wants done. What's clear is that PIE′ isn't it.[29]

---

[29]In the next chapter of this thesis, I will consider this issue in more depth and give a suggestion for how I think primary intensions should be understood.

1.4.5 Closure

It's time to discharge the assumption that *apriority* is closed under *a priori* entailment, or at least to dissolve the puzzle for closure that was raised back in §1.1.3. I want to say that the following two sentences are *a priori*:

> (RSIS) If 'sister' is a NOUN, then 'sister' refers to sisters.

> (5) Sisters are female siblings.

And these appear to *a priori* entail:

> (9) If 'sister' is a NOUN, then 'sister' refers to female siblings.

And this looks *a posteriori*. Since it is not open to me to deny (RSIS), (5), or *a priori* closure, how am I going to avoid calling (9) *a priori*? The problem, I think, is that (RSIS) and (5) fail to be *a priori connectable.*

Well, what does *that* mean? I have already argued that (RSIS) is a playground conditional. Part of why we are able to know it *a priori* is that the antecedent constrains what the consequent can mean. We don't have to worry about the consequent " 'sister' refers to sisters" being like " 'is' refers to iss", which isn't even true, let alone *a priori*. The antecedent makes sure that the consequent is going to be well-formed and true.

However, now that we see why (RSIS) is *a priori*, we can see that the move from (RSIS) and (5) to (9) can't be *a priori*, for we don't know *a priori* that the word 'sisters' as it appears in (RSIS) can be hooked up with the word 'sisters' as it appears in (5). In general, because of opacity phenomena, we must be very careful with any inference involving a playground conditional. The normal rules don't necessarily apply. The following, rather silly, inference makes the point nicely:

> (21) If 'Descarte' is a (singular count) noun (pluralizable with 's' in the language I am currently speaking), then 'Descarte' refers to Descartes.

> (22) Descartes was the author of the *Meditations*.

∴ (23) If 'Descarte' is a (singular count) noun (pluralizable with 's' in the lanugage I am currently speaking), then 'Descarte' refers to the author of the *Meditations*.

We know (21) and we know (22), but they aren't *connectable* (*a priori* or otherwise), so we can't infer (23) (which is a good thing, since (23) doesn't appear to be true). The inference relies on a kind of equivocation.

Likewise, we know (RSIS) and (5), and we even know them *a priori*, but we can't connect them *a priori* to get (9). The point isn't that that the inference is equivocal, the point is that we can't rule out *a priori* that it's equivocal. For all we know *a priori*, the inference could be fallacious in the same way that the Descartes inference is fallacious. What saves it from being fallacious? Quite simply, the truth of the antecedent of the playground conditional. Playground conditionals threaten to play around with the meaning of our words, but, in (RSIS), the threat isn't realized. The assumption made by the antecedent of (RSIS) turns out to true.

So, I want to claim that *connectability* depends on the truth of the antecedent of the playground conditional, and *a priori connectability* depends on *a priori* knowledge of its truth. There are, however, anomalous cases to consider. What, for example, are we to do with the following inference?

(24) If 'dog' means cat, then dogs are cats.

(25) It's not the case that dogs are cats.

∴ (26) It's not the case that 'dog' means cat.

If connectability depends on the truth of the antecedent of the playground conditional, then the premises of this argument clearly will not be connectable. Thus, this argument will be problematic in the same way that the Descartes inference was problematic. But, intuitively, the current argument argument looks fine. It certainly doesn't look like it would keep company with the Descartes argument.

This intuition is wrong. The word 'dog', as it's used in the consequent of (24) is not connectable with 'dog' as it appears in (25). The argument is equivocal and therefore invalid. However, there is a nearby argument that is perfectly fine, and,

I suggest, is what an ordinary speaker would have in mind with an argument like (24)–(26):

(27) If 'dog' means cat, then dogs are cats.

(28) The consequent of (27) is false, when understood according to standard meanings.

∴ (29) Since (27) is itself true, the supposition made by its antecedent must be false.

∴ (30) It's not the case that 'dog' means cat.

When we try to *modus tollens* a true playground conditional, we have already decided not to play along with the supposition made by the antecedent before we arrive at the conclusion that the antecedent is false.[30] The "standard meanings" clause of (28) makes clear that this is what's going on.

The upshot of all of this is that arguments like (RSIS)–(9) cannot be used to undermine closure for *apriority*. Such arguments involve premises that, while *a priori* and *connectable*, are not *a priori connectable*.

1.4.6   Recap of the Argument

Let's end by recapping the whole argument of this section. It went roughly like this (where 'AP $[x]$' is to be read 'It's *a priori* that x'):

(PREM)   AP [If 'sister' is a NOUN, then "'sister' refers to (all and only) sisters" is true]   *(premise)*

(TSIS)   AP [If "'sister' refers to (all and only) sisters" is true, then 'sister' refers to (all and only) sisters]

∴ (RSIS)   AP [If 'sister' is a NOUN, then 'sister' refers to (all and only) sisters.] *(From (P1), (TSIS), and Closure)*

(UNPR)   AP [If 'sister' is a NOUN and it actually refers to male siblings, then the NOUN 'sister' actually refers to male siblings]

∴ (SISM)   AP [If the NOUN 'sister' actually refers to male siblings, sisters are male siblings] *(from (RSIS), (UNPR), and Closure)*

---

[30]Notice that, if what I've said in this section is correct, there is no similar problem with using *modus ponens* on a playground conditional.

1.5   A concluding remark

In light of the pivotal role that playground conditionals play in several philosophical debates, and the constraints that they appear to impose on an analysis of indicative conditionals, I think that we do well to give them careful consideration. Indeed, I think it would be interesting to see what sort of theory of indicative conditionals one would arrive at if one didn't treat playground conditionals as mere fringe cases to be accommodated later on (if at all), but rather took them as central, paradigm, instances of indicative conditionals. Unfortunately, this is a project that will have to be left for another occasion.

Chapter 2

MODAL RATIONALISM, TWO-DIMENSIONALISM, AND OUR COUNTERACTUAL SISTERS

2.1 Days of Yore

The story is well-known. Here's a somewhat romanticized version: Philosophers once thought that *apriority* and *necessity* were happily and forever bound together. Kant even proclaimed that, "necessity and strict universality... are infallible tests for distinguishing pure [*a priori*] from empirical knowledge..."(Kant 1990, p. 3); and it was thus secured as philosophical orthodoxy. Then came Kripke[1]. Armed with simple examples like 'water is H_2O' and 'the length of S at t_0 is one meter', Kripke showed us how the epistemic category of *apriority* could be divorced from the metaphysical category of *necessity*. As the post-Kripkean philosophical world tried to come to grips with the *contingent a priori* and the *necessary a posteriori*, "two-dimensionalism" emerged. Borrowing from recent work in formal modal logic, two-dimensionalism offered a framework for understanding what was going on with Kripke's examples. It taught that every sentence could be associated with not just one, but *two* intensions,

---

[1]See his *Naming and Necessity* (Kripke 1972).

and that the right lesson to have learned from Kripke was that the necessity of one of these intensions did not (and should not be expected to) go along with *apriority*. This left open that the necessity of the other intension *might* go along with *apriority*. Some two-dimensionalists, the *modal rationalists*, embraced this suggestion. Then came the trouble. Modal rationalists, armed with the two-dimensional framework, began to argue for wild metaphysical claims, such as the claim that *physicalism is false*.[2] This bred confusion. Everyone agreed that the two-dimensional apparatus was useful for some purposes, but no one could agree on how the apparatus was best interpreted, or if its best interpretation lent any credence to modal rationalism. These are the issues I wish to explore.

In this paper, I will first give a brief characterization of a version of modal rationalism that has recently been defended by David Chalmers. To properly characterize the view will require us to say a bit about two-dimensionalism, and the distinction between *primary* and *secondary* intensions. As we will see, there is some controversy concerning how, exactly, primary intensions should be understood. I will first consider whether Chalmers' own definition of primary intension offers any support to modal rationalism. I will argue that it does not. I will then consider a second way of understanding primary intensions—one that initially looks like it avoids the problems that accompanied Chalmers' proposal. We will see, however, that it does not. After briefly touching on an additional worry that arises for both understandings of primary intensions, I will suggest a way of characterizing primary intensions that *does*, I think, avoid the difficulties.

2.2   Strong Modal Rationalism

The main thesis I want to look at is a version of what Chalmers calls *strong modal rationalism*. This version of strong modal rationalism claims that a sentence is *a priori* if and only if it is *primarily necessary*, or, in shorthand:

---

[2]See, for example, Chalmers' *The Conscious Mind* (Chalmers 1996).

[SMR] primary necessity $\equiv$ *apriority*[3]

This needs to be unpacked a bit. First, what is meant by 'primary necessity'? Chalmers associates every sentence with two sets of truth conditions, each of which can be thought of as a function from possible worlds to truth-values. The difference between the two functions is in how we think of the possible worlds which serve as the inputs. We can think of them either as counterfactual possibilities (roughly, ways the world could have been), or as counter*actual* possibilities (roughly, ways the world could be).[4] What Chalmers calls the *primary intension* of a sentence is the function that results when the worlds are considered as counter*actual*. The *secondary intension*, on the other hand, is what results when the worlds are considered as counter*factual*.[5]

How, exactly, do we consider worlds as counterfactual and counteractual? There is pretty widespread agreement concerning how to treat a world w as counterfactual (for the purpose of evaluating the secondary intension of some particular sentence). All we have to do is ask whether the following is the case:

[SIE][6] If W had been the case, it would have been that S.

(Where 'W' is replaced with a name or complete description of w and 'S' is replaced with the sentence in question.) If SIE is true, then the secondary intension of the

---

[3]As Lauren Ashwell and Darin Harootunian have reminded me, SMR is potentially stronger than the thesis that Chalmers actually defends under the title "strong modal rationalism". In "Does Conceivability Entail Possibility" (Chalmers 2002a), he defines strong modal rationalism in a way that would only get us the left-to-right direction of SMR. However, it's pretty clear that the only reason he does this is because, as I will note below, he thinks (mistakenly) that the right-to-left direction of SMR is trivially guaranteed by his definition of *primary intension*. In his most recent writing on the subject, Chalmers explicitly defends SMR under the banner "The Core Thesis" (Chalmers 2004).

[4]There are (at least) two ways that one could proceed when defining primary intensions. One could think of them as intensions defined over a certain set of world-like objects, the *epistemic possibilities*, or one could think of them as defined over the same modal space of possible worlds as secondary intensions. Chalmers has, at different times, pursued both of these strategies. The current paper is framed in terms of the second strategy. Primary and secondary intensions diverge (when they diverge) not because they are defined over different objects, but because there are two distinctive ways of thinking about a common domain of objects. There are certain advantages that this strategy has over the strategy that takes primary intensions to be defined over a modal space of epistemic possibilities. This side path is worth exploring, but I won't explore it here.

[5]Frank Jackson (Jackson 1998) calls these A-intensions and C-intensions, respectively ('A' for *actual*; 'C' for *counterfactual*).

[6]SIE= Secondary Intension Evaluator.

sentence maps w onto the value *true*. If SIE is false, then the secondary intension maps w onto *false*.

How to treat a world as counter*actual*, however, is another matter. The word 'counteractual', after all, is a term of art. It is introduced by appealing to certain intuitions about a non-counterfactual-seeming way of thinking about other possible worlds. But just how to cash out these intuitions is controversial. Since our launching point in this paper is Chalmers' modal rationalism, let's start with his proposal.

In order to see how his proposal goes, we will first have to note that he associates possible worlds with "canonical descriptions" of those worlds (Chalmers 2002a, p. 166). Primary intensions are then defined, not in terms of the worlds themselves, but in terms of their canonical descriptions.[7] Here is the basic proposal: In order to consider a world with canonical description D as counteractual, for the purpose of evaluating the primary intension of some sentence S, ask:

[PIE$_c$] Is the material conditional $\ulcorner D \supset S \urcorner$ *a priori*?[8]

If the answer to PIE$_c$ is 'yes', then the primary intension of S maps the world described by D onto *true*; if the answer is 'no', the primary intension maps it onto *false*.

Now that we have a notion of primary intension on the table, we can say what *primary necessity* (a.k.a. *1-necessity*) is. It is *not* a special kind of necessity defined over a special set of worlds; it is just necessity when possible worlds are considered as actual—that is, it's when the primary intension maps every possible world onto *true*.

We have almost finished unpacking SMR. There is one detail that's left: what is meant here by *apriority*? In a paper of this size, it's just not possible to do this

---

[7]There are plenty of things that could be worried about with respect to these canonical descriptions (e.g., in what terms are they formed?), but, for now, let's let it pass. I should note, however, that some indexical information will need to be encoded in the description. It is a familiar point that what we really need for counteractual evaluation is *centered worlds*, which can be thought of as ordered pairs of worlds and "centers", where centers function to fix the perspective from which the world is viewed. A center fixes when 'now' is, where 'here' is, who 'I' is, etc. If we want to move to a formulation of primary intensions in terms of canonical descriptions of worlds, then, the information that the center encodes will have to be part of the description. It is also important to keep in mind that canonical descriptions are to be thought of as large *sentences* that say how things are with the given world; they aren't just big definite descriptions (though they will probably contain big definite descriptions).

[8]See Chalmers 2002a, p. 165.

question justice. I will have to assume that the reader has a general idea of what apriority involves. I should, however, clear up one potentially puzzling feature of Chalmers' use of the *a priori*. In PIE$_c$, and elsewhere, Chalmers talks of sentences—sentence *tokens*, that is—as being *a priori*. What does it mean for a *sentence token* (as opposed to, say, a *proposition*) to be *a priori*? In terms that admittedly require further exposition, but, unfortunately, won't receive any here, to say that a sentence token is *a priori* for agent A is to say that it "expresses a thought which can be conclusively non-experientially justified by A on ideal rational reflection."[9,10]

I don't claim that these remarks make 1-necessity and *apriority* crystal clear. But hopefully the reader now has some sense of what the thesis SMR says. I now want to discuss two worries one could have about the view. We will see that the proper resolution of the first of these worries immediately paves the way to the second. This second worry is not as easily resolved, and is crippling to SMR if it doesn't receive a satisfactory response. I will argue that a satisfactory response has not yet been given. I will then consider whether a move to a different definition of primary intensions can avoid the problem.

---

[9]This formulation is (more or less) the account of the *a priori* Chalmers gives in "The Foundations of Two-Dimensional Semantics" (Chalmers 2004, §3.9) and "On Sense and Intension" (Chalmers 2002b, §4, f.n.).

[10]If two-dimensionalism is right, then token thoughts, like token sentences, are plausibly bearers of two-dimensional content. So, one might get worried here about whether we are understanding the thought which A's sentence token expresses according to its primary or secondary intension. If we are understanding it according to its primary intension, then it looks like Chalmers has defined primary intensions in terms of *apriority* and *apriority* in terms of primary intensions. I don't think we need to get worried about this. Even if thoughts have two dimensions of content, the definition given above for *apriority* of sentence tokens does not need to appeal to the two-dimensional nature of this content. Primary and secondary intensions always agree in the value they assign to the actual world. To say that a sentence expresses a thought which could be "conclusively, non-experientially justified by A on ideal rational reflection" just means that the thought A would express with this sentence token is true, and A could obtain a certain kind of justification for believing it to be true. Perhaps we'll want to appeal to primary intensions down the road, when explaining how or why such a justification is available, but, here, at the outset, we needn't appeal to the two-dimensionality of content (at least as far as I can see).

2.3 First Worry with PIE$_c$: Triviality

2.3.1 The Argument

Now that we have PIE$_c$ on the table, perhaps the reader is beginning to feel a little uneasy with Chalmers' thesis SMR. Uneasy not because it looks like it's in danger of being false, but because it looks like it's in danger of being *trivial*. Indeed, even Chalmers seems to think that one direction is trivial.[11] Let's call the trivial direction SMR$_{rl}$. It says that if a sentence S is *a priori*, it is 1-necessary. Chalmers' definition of primary intension appears to secure this. Consider the following argument:

(1) Assume S is *a priori*.

∴ (2) Any material conditional of the form $\ulcorner x \supset S \urcorner$ is *a priori*.

∴ (3) For all canonical descriptions D, $\ulcorner D \supset S \urcorner$ is *a priori*.

∴ (4) S is 1-necessary.

This argument certainly looks valid. However, I think that there is a compelling reason to think that Chalmers shouldn't have stated PIE$_c$ in terms of material conditionals, and when PIE$_c$ and (1)–(4) are restated in terms of the right kind of conditional, a gap opens up between (1) and (2).

2.3.2 What's wrong with using material conditionals?

I will start by saying why Chalmers shouldn't have stated PIE$_c$ in terms of material conditionals: First of all, remember that part of what primary intensions are supposed to do for us is explain away Kripke's examples of the *necessary a posteriori* and *contingent a priori*. As part of Chalmers' dissolution of the first of these puzzles, he says that although the secondary intension of 'water is XYZ' is necessarily false, it has a contingent primary intension (and that is why we can't know that water isn't XYZ *a priori*). One of the worlds that is supposed to stand witness to the 1-contingency (contingency of primary intension) of 'water is XYZ' is Putnam's Twin Earth (where

---

[11]He's not bothered by this, as he thinks the other direction has some substance.

the stuff in the lakes and streams, etc., is XYZ). Thus, if he is going to employ PIE_c, Chalmers must have it that the material conditional, 'the clear liquid stuff in the lakes and streams, etc., is XYZ $\supset$ water is XYZ' is *a priori*. Well, is this material conditional *a priori*? I think the answer is 'No'.

But this is not obvious. The problem is that we don't usually think in terms of horseshoes.[12] It's very implausible that the ordinary English indicative conditional is the material conditional.[13] This creates a problem, as our intuitions about the *apriority* of material conditionals can easily be distorted by our intuitions about the *apriority* of the conditionals we use in everyday discourse. But there is a way to cut down on this distortion. Since a material conditional $\ulcorner p \supset q \urcorner$ is equivalent to the disjunction $\ulcorner \neg p \vee q \urcorner$, we can simply restate PIE_c in terms of the *apriority* of these equivalent disjunctions:

[PIE'_c] Is $\ulcorner \neg D \vee S \urcorner$ *a priori*?

Now, returning to our example, in order to determine if Twin Earth gets mapped onto the value *true* by the primary intension of 'water is XYZ', we ask if the following disjunction is *a priori* (where 'TWIN EARTH' abbreviates Twin Earth's canonical description):

[TE] It is not the case that TWIN EARTH $\vee$ water is XYZ.

There seem to be only two ways we could come to know *a priori* that a disjunction like TE is true. First, we might know *a priori* that one of the disjuncts is true, and thus deduce, through *a priori* reasoning, that the disjunction must also be true. This won't work here. The second disjunct of TE is necessarily false, so it certainly is not a candidate for the status of *a priori* truth. And, while the first conjunct is at least true, it is *not* knowable *a priori* (barring some significant innovations in armchair

---

[12]Thanks are owed to Eric Swanson and Bernhard Nickel, who did much to help me think more clearly about conditionals and Chalmers' use of them. (Which isn't to say that either one of them would consider my current thinking on these issues to be clear.)

[13]Although this position has been defended by some philosophers—most notably H.P. Grice and Frank Jackson. See Grice's "Logic and Conversation" (Grice 1989) and Jackson's *Conditionals* (Jackson 1987). For some considerations against the equivalence thesis, see §1.2 (in Chapter 1) of this thesis.

chemistry). So the first route to an *a priori* knowledge of the disjunction isn't going to work.

The only other route to such knowledge is via a prior knowledge that some very intimate relation—call it relation R—holds between the two disjuncts. This is, for instance, how we know *a priori* that the following disjunction is true: *either it is hailing burning sulfur in Cambridge or it is not the case that it is hailing burning sulfur in Cambridge.* If I happen to know one or other of the disjuncts, it is clear that this knowledge is thoroughly *a posteriori*. But since I know that the denial of each of the disjuncts logically entails the other disjunct, I can come to know the disjunction itself *a priori*. So, perhaps p and q are R-related iff $\ulcorner \neg p \wedge \neg q \urcorner$ is a contradiction. This, however, would be too stringent a condition on R-relatedness to get Chalmers what he needs, for it hardly seems like 'TWIN EARTH $\wedge$ water is not XYZ' is a logical contradiction; the canonical description abbreviated by 'TWIN EARTH', according to Chalmers, won't even contain the words 'water' and 'XYZ' (the descriptions will be in more fundamental qualitative terms).

Here is another way of cashing out relation R that won't work for Chalmers: p and q are R-related iff any possible world in which p is false is a world in which q is true. Why won't this work? Well, water is actually H_2O, and, if Kripke is right, 'water' is a rigid designator, so it follows that water is H_2O in every possible world. Thus, there are no worlds with respect to which the second disjunct of TE is true. Given the suggested definition of R-relatedness, then, the only way the two disjuncts of TE could be R-related is vacuously—that is, if the first disjunct is true in every world. In that case, it would be vacuously true that every world in which the first disjunct is false is a world with respect to which the second is true. But this kind of vacuous R-relatedness isn't such that we can know *a priori* that it obtains for TE, for we cannot know *a priori* that the first disjunct is true in every world.

What we need, and what this definition of R-relatedness fails to give us, is some play with the meaning of the second disjunct of TE. Sure, it *actually* expresses a necessarily false proposition, but we can imagine ways the actual world could have turned out to be, such that 'water' would have turned out to rigidly designate XYZ

instead of H_2O, and the second disjunct would have turned out to be true. Indeed the negation of the first disjunct of TE is one such way. So we need a notion of R-relatedness between p and q, where the supposition of the truth or falsity of one of them can affect the meaning of the other.

Ordinary English conditionals seem to meet this criterion. Members of a certain class of ordinary indicative conditionals demonstrate this vividly. Some of us had our first encounter with them on the playground. Billy said, "I hate X, she's a jerk." Then a classmate kindly reminded Billy that today is "opposite day" (where utterances are evaluated as having the truth-conditions that their negations standardly have). Not sure whether to believe this news, Billy responded, "if today is really opposite day, then I love X." Obviously, Billy intended that if the antecedent of his conditional was true, the consequent was to take on a non-standard meaning. Call conditionals like this *playground conditionals*.[14] Here is another example: Billy says to Mary, "Yeah, you're real smart, if 'smart' means stupid". Something like these playground conditionals looks like just what we need in order to secure a kind of R-relatedness between our disjuncts p and q that allows the truth or falsity of one of them to influence the meaning of the other. Let's define R-relatedness like this:

[R-RELATEDNESS] p and q are R-related iff ⌜if $\neg p$, then q⌝ is true.

The quasi-quoted conditional on the right-hand side of the biconditional is an ordinary English conditional. And, as playground conditionals show, sometimes the antecedents of these conditionals "mess" with the meanings of the consequents.[15]

Now it might seem that we have a viable way for TE to come out *a priori*. What is required is for us to know *a priori* that relation R holds between the two disjuncts. That is, we must know *a priori* that

If TWIN EARTH, then water is XYZ

(where this conditional is understood as an ordinary English conditional). Perhaps this is *a priori*, but there is a serious problem here. R-relatedness was supposed to

---

[14]For more on playground conditionals, see the first chapter of this thesis.

[15]Yablo, in Yablo 2002, spends some time discussing conditionals like this.

give us a special relationship between the two disjuncts of TE. But the *a priori* truth of the conditional 'if Twin Earth, then water is XYZ' does *not* guarantee that there is a special relationship between these disjuncts. For 'water is XYZ' doesn't mean the same thing in TE as it means in this conditional. In the conditional, its meaning has been affected by the antecedent in such a way that it no longer expresses a necessarily false proposition. But in the disjunction it does express a necessarily false proposition. The R-relatedness of p and some q-look-alike that means something different than q can't possibly help explain how we know TE a priori.

We can now see that it was a mistake to define PIE$_c$ in terms of the *apriority* of *material* conditionals. The horseshoe of the material conditional is like an impenetrable barrier. Antecedents and consequents aren't allowed to commingle and influence each other's meanings. And we need at least some level of this commingling if we are to explain why Twin Earth gets mapped onto *true* by the primary intension of 'water is XYZ'.[16]

2.3.3 Getting PIE$_c$ right, and back to triviality

We can easily solve this particular problem by altogether abandoning material conditionals for the purpose of defining primary intensions. What Chalmers really wants and needs in PIE$_c$ is a kind of conditional that allows the antecedent to affect the meaning of the consequent to some degree. Since, as we have just seen, our regular English indicatives seem to allow this kind of interaction, perhaps we could restate PIE$_c$ like this:

[PIE$_c''$] Is ⌜if D, then S⌝ *a priori?*

---

[16]Could we have avoided this R-relatedness business by restating PIE$_c$ in terms of the *a priori* falsity of ⌜$\neg D \wedge S$⌝? Isn't it *a priori* false that (Twin Earth is actual $\wedge$ water $\neq$ H$_2$O)? It would certainly be strange if this helped, since this way of putting PIE$_c$ is logically equivalent to the disjunctive formulation that was given above. Indeed, I think that the very same logical issues *do* arise in evaluating the negated conjunction. However—and this is the reason I bring up the issue at all—it *does* seem that the negated conjunction way stating PIE$_c$ has more initial intuitive appeal than the logically equivalent disjunctive way. I'm not exactly sure why this is the case, but I suspect that there is some cognitive interference being caused by nearby indicative conditionals. I won't try to cash out that metaphor here.

This gets the right result in the Twin Earth case. Twin Earth gets mapped onto *true* by the primary intension of 'water is XYZ' because it is (at least arguably) *a priori* that if TWIN EARTH, then water is XYZ.

Now we are finally in a position to see why SMR$_{rl}$ can't be shown to be trivial by an argument like (1)–(4). Once (2) and (3) are rewritten in terms of regular English conditionals, as they should be, it looks like the move from (1) to (2) is illegitimate—it involves a potential equivocation. The reason the material conditionals in (2) looked *a priori* to us was because they had what we took to be an *a priori* consequent— and this together with the truth conditions for the material conditional secured the *apriority* of the conditional. But if the consequents of some ordinary conditionals can have non-standard contributions to overall meaning, we can't be sure that they are *a priori* just on the basis of (1). So the inference from (1) to (2) fails.

Chalmers might try to insist here that the antecedents of the kind of conditionals he wants to define PIE$_c$ in terms of only ever mess with the *secondary intensions* of the consequents, never the *primary intensions*. And since *apriority* has to do with primary intension only, the inference from (1) to (2) goes through after all. But, obviously, he can't just *stipulate* that he is using a conditional for which this is true. The conditionals in question are part of how he *defines* primary intensions. It would be a bad thing if the conditionals were then stipulated to have some feature that we must appeal to primary intensions in order to isolate.

Since this section has been rather sprawling, let me end with a summary: I started with a simple argument that appeared to show that the right-to-left direction of SMR is trivial. I then gave an argument that Chalmers' original definition of primary intensions, PIE$_c$, does not serve one of the main purposes it is supposed to serve, for it does not deliver the desired result that the primary intension of 'water is XYZ' maps Twin Earth onto *true*. It fails to do this because when we evaluate a material conditional, we do not (or should not) allow the antecedent of the conditional to affect the meaning of the consequent. We do, however, allow this for ordinary indicative conditionals, as is vividly demonstrated by playground conditionals. A restatement

of PIE_c in terms of these conditionals manages to deliver the right result in the Twin Earth case. And once we restate PIE_c so as to get Chalmers' desired results, we see that the argument for the triviality of SMR_{rl} fails.

2.4   Second Worry with PIE_c: The Sisters Problem

2.4.1   The Problem

In "Coulda, Woulda, Shoulda" (Yablo 2002), Stephen Yablo gives an argument which exploits a potential gap which playground conditionals create between *apriority* and 1-necessity. Take a statement that looks paradigmatically *a priori*, like 'sisters are female siblings'. Yablo argues that, given the way we've just defined primary intensions, even this sentence will be 1-contingent. The argument is meant to generalize to show that there are no 1-necessary claims. So, either *apriority* is also an empty category, in which case SMR is uninteresting, or *apriority* isn't an empty category, in which case SMR is false.

In order to see that our paradigmatically *a priori* claim, 'sisters are female siblings', has a contingent primary intension, all we need to do is show that the primary intension of its negation maps at least one world onto the value *true*. And, according to Yablo, there is such a world. Consider counteractual Badworld, where we use the word 'sister' to refer to male siblings.[17] PIE_c'' instructs us to ask if the conditional "if the word 'sister' *actually* refers to male siblings, then sisters are not female siblings" is *a priori*. It looks like it is. For the conditional "if the word 'sister' *actually* refers to male siblings, then sisters are male siblings" looks *a priori*. Thus, Badworld gets mapped onto *false* by the primary intension of 'sisters are female siblings' and onto *true* by the primary intension of 'sisters are male siblings'. So both sentences have contingent primary intensions, even though the former is paradigmatically *a priori*, and the latter is the negation of something paradigmatically *a priori*. Indeed, it looks like we could pull a trick like this for any putative example of an *a priori* truth.

---

[17]Here, and throughout this section, when I say "'p' refers to blahs", it is to be understood as " 'p' refers to *all and only* blahs".

2.4.2 Chalmers' Response

It's not surprising that Chalmers resists Yablo's argument. In "Does Conceivability Entail Possibility" (Chalmers 2002a), Chalmers argues that the conditional,

[BW] if 'sister' refers to male siblings, then sisters are male siblings

is *not a priori*.[18] To see why he thinks this, let's first reconstruct how he takes Yablo's argument for the *apriority* of BW to go. (Terminological note: 'AP[x]' is read " 'x' is *a priori*"):

(5) AP[if 'sister' refers to male siblings, then 'sister' refers to male siblings]

(6) AP['sister' refers to sisters]

∴ (7) AP[if 'sister' refers to male siblings, then sisters are male siblings]

Chalmers objects to premise (6). He says that it is a "substantive, *a posteriori*, metalinguistic" fact that 'sister' refers to sisters (or anything else for that matter) (Chalmers 2002a, 170). So you can't move *a priori* from claims about the word 'sister' to claims about sisters. To say that BW is *a priori* is to assume that you can.

There are (at least) two serious problems with this response. In the next two subsections, I will discuss each in turn.

2.4.3 First Problem

There is a can of worms begging to be opened here concerning the status of instances of the R-schema:

[R] '*t*' refers to *t*s.

Are R-sentences *a priori*? Well, at first glance, some people are tempted to answer 'yes'. A skeptic might complain, however, that there are plenty of instances of the R-schema that aren't *a priori*, for they aren't even true (or truth evaluable, for

---

[18]Of course, his discussion is in terms of material conditionals, but nothing important turns on this.

that matter). Consider " 'is' refers to iss" and " 'flimflamaphone' refers to flimflamaphones".

But there is a straightforward revision we can make to the R-schema in order to skirt around this problem. The trick is to state the schema as a conditional, where we load into the antecedent enough to weed out the problem cases. The revised schema will look something like this:

> [R'] If 't' is a (singular count) noun (pluralizable with 's' in the language I am currently speaking), then 't' refers to ts[19].

Chalmers might still regard this as *a posteriori*, and claim that the hedging accomplished by the antecedent of the revised schema eliminates only *one* source of the R-schema's *aposteriority*. I have already argued at length, in Chapter 1 (§1.4.2) that instances of the R'-schema are true, and *a priori*, playground conditionals. I won't reproduce that argument here. In order for Chalmers' response to Yablo to work, however, he must say that even instances of the R'-schema are *a posteriori*, for Badworld is a world where the *singular count noun* 'sister' refers to male siblings. I think that this is an implausible line to take.

2.4.4 Second Problem

The second problem with Chalmers' response to Yablo's Sisters Problem is that, given his other theoretical commitments, he just isn't entitled to it. It is his *scrutability* thesis that causes the problem. The thesis goes something like this (Note: '$p \to q$' is short for " 'if p, then q' is *a priori*"):

> [SCRUT] If D is a [canonical] description, then for all S, $(D \to S) \vee (D \to \neg S)$.

In English, this says that for every canonical description D and claim S, either D *a priori* entails S, or it *a priori* entails $\neg S$. The idea is that canonical descriptions settle, *a priori*, everything that can be settled.

---

[19]Thanks to Selim Berker and the MATTI group for helping me state R' more precisely than I otherwise would have thought to.

This thesis plays a crucial role in one of Chalmers' anti-materialist arguments. One can think of the argument as a *reductio*: Assume that materialism is true. If that's so, then the canonical description of our world will consist of a complete microphysical description of the world (P), plus some indexical information (I), plus a 'that's it' clause (T). But, Chalmers argues, such a description does not *a priori* entail the existence of qualitative experiences, for zombie worlds (worlds that are PTI-duplicates of our world but contain no consciousness) are primarily conceivable. This creates a problem, for we know that our world *does*, in fact, contain qualitative experiences. Assuming that SCRUT is true, then, this means that PTI is *not* the canonical description of our world. Thus materialism is false. (Notice that if SCRUT is false, then it might be that PTI *is* the canonical description of our world, and it *a posteriori* necessitates the qualitative truths. SCRUT's role in the argument is to rule out this possibility.)

Returning to the matter at hand, SCRUT cause a problem for Chalmers' response to Yablo's Sisters Problem; for, If SCRUT is true, then the response is a red herring. Chalmers' response boils down to the denial of (6) (the claim that it is *a priori* that 'sister' refers to sisters). But hold on. Although Yablo didn't give us a complete canonical description of Badworld, that was simply due to time and space constraints. We were clearly intended to imagine the rest of the description as filled in. Abbreviate the completed description as 'BADWORLD'. Now, when we ask if the conditional 'if BADWORD, then sisters are male siblings' is *a priori*, it won't do to just say, "no, because the instance of the R-schema linking 'sister' to sisters is *a posteriori*." Either it is true with respect to Badworld that 'sister' refers to sisters, or the denial of this claim is true. This, together with SCRUT, tells us that since 'BADWORLD' is a canonical description, either BADWORLD *a priori* entails *'sister' refers to sisters* or it *a priori* entails *'sister' does not refer to sisters*. Either way, Chalmers is in for trouble.

Assume first that BADWORLD → *'sister' refers to sisters*. Then, a version of Yablo's original argument goes through:

(5′) AP(if BADWORLD, then 'sister' refers to male siblings)

(6') AP(if BADWORLD, then 'sister' refers to sisters)

∴ (7') AP(if BADWORLD, then sisters are male siblings)

Thus, the primary intension of 'sisters are male siblings' maps Badworld onto *true*; and the primary intension of 'sisters are not male siblings' maps Badworld onto *false*. This was precisely the result that threatened to tear *1-necessity* and *apriority* apart. And again, it is easy to see that any claim can be shown to be 1-contingent in this way.

Perhaps, then, BADWORLD → *'sister' does not refer to sisters*. But how on earth would we get that BADWORLD *a priori* entails *that*? If we take Badworld seriously as a counteractual possibility, it's hard to see how we could. BADWORLD implies that 'sister' actually refers to male siblings. If we conclude from this that BADWORLD implies that 'sister' doesn't refer to sisters, we are implicitly saying that counteractual hypotheses are not allowed to tell us what words actually refer to. But that's exactly what BADWORLD purports to do. So it looks like, according to Chalmers, there are certain facts about meaning that we don't allow counteractual worlds to specify, and we'll just ignore them if they try.

One measure of how seriously we're taking a counteractual possibility is this: while we're considering the counteractual possibility, are we willing to assertively utter sentences that will only turn out to be true if the counteractual possibility does not turn out to be the actual world? If the answer is 'yes', let us say that the counteractual possibility is not being considered with *full seriousness*. If we say that BADWORLD → *'sister' does not refer to sisters*, for example, we are clearly not considering counteractual Badworld with full seriousness. For if Badworld *is* the actual world, then 'sister' refers to sisters (it's just that sisters aren't what we thought they were).

But, in certain other cases, it's absolutely crucial for Chalmers that we *do* consider a counteractual world with full seriousness. Chalmers doesn't want us to say that TWIN EARTH → water is H_2O. Why not? After all, the right-hand side seems to express a true and necessary identity. Well, the idea seems to be that *if* counteractual

Twin Earth turns out to be the actual world, then 'water is H_2O' will turn out to be false. And for that reason, we should reject the relevant implication.

Chalmers now owes us an explanation of why some counteractual worlds are to be taken with full seriousness, and others are not. And he must tell us where to draw the line. I think that he'll be unable to do this without overtly (or covertly) appealing to primary intensions in the process.

2.5 PIE_y: A Better Flavor?

One strategy that might be taken at this point, in order to avoid the Sisters Problem and preserve SMR, is to switch to a different definition of *primary intension*. Yablo 2002 suggests an alternative: In order to evaluate the primary intension of a sentence S with respect to a world w, ask,

> [PIE_y] If it had turned out that w was actual, would it have turned out
> that S? (Yablo 2002, p. 454)

If the answer is 'yes', the primary intension of S maps w onto *true*, otherwise it gets mapped onto *false*. Yablo argues that this definition delivers the same results as PIE_c in standard cases, including Twin Earth cases, without falling prey to the Sisters Objection. He thinks that the following conditional seems intuitively *false*:

> [SIS] If it had turned out that in the actual world we used 'sister' to refer
> to male siblings, it would have turned out that sisters were male
> siblings.

What our use of language turns out to be like doesn't affect the gender our sisters turn out to have. Yablo notes that this "would have turned out" conditional seems to protect the consequent from deviant meaning shifts brought about by the antecedent. Thus, PIE_y does *not* lead to the undesirable result that all sentences have contingent primary intensions.

Or does it? Consider the following question: if it had turned out that we actually used the word 'sister' to refer to male siblings, would it have turned out that instances

of the R-schema were false? Of course not. The disquotational schema would have turned out to deliver truths no matter what our words turned out to refer to. All the disquotational schema cares about is whether the same noun is being used and mentioned. But if that's right, we can make the following argument (terminological note: '$x \Rightarrow y$' is read, 'if it had turned out that x, it would have turned out that y'):

(10) 'Sister' refers to male siblings $\Rightarrow$ instances of the R-schema are true.

(11) Instances of the R-schema are true $\rightarrow$ " 'sister' refers to sisters" is true.

(12) " 'Sister' refers to sisters" is true $\rightarrow$ 'sister' refers to sisters.

$\therefore$ (13) 'Sister' refers to male siblings $\Rightarrow$ 'sister' refers to sisters.

(14) 'Sister' refers to male siblings $\Rightarrow$ 'sister' refers to male siblings.

$\therefore$ (SIS) 'Sister' refers to male siblings $\Rightarrow$ sisters are male siblings.

Premise (11) says that it's *a priori* that if instances of the R-schema are true, then the sentence " 'sister' refers to sisters" is true (since it's an instance of the R-schema). I think this premise should be fairly uncontroversial (bracketing general worries about the *a priori*).

Premise (12), on the other hand, might look more tendentious. It says that it's *a priori* that if the sentence " 'sister' refers to sisters" is true then 'sister' refers to sisters. But just as the status of the R-sentences is a matter of debate, so is the status of T-sentences (sentences of the form " 'p' is true iff p"). However, I think that only one direction of this biconditional (right to left) could cause a problem for *apriority*. It's a substantive *a posteriori* metalinguistic fact that the sentence 'snow is white' expresses the proposition that snow is white. It could have expressed the proposition that grass is green or the proposition that Mary is quite contrary, or no proposition at all. So we don't know *a priori* that if snow is white then 'snow is white' is true. But don't we know *a priori* that if 'snow is white' is true, then snow is white? Perhaps some experience is required in order to understand the claim and

to have the component concepts, but once we do, we can see that the claim must be true.[20]

Now, assuming that we've established (10), (11), and (12), what gives us the right to conclude (13)? This inference requires the following very plausible-sounding principle:

$$[\text{P1}] \quad \forall p, q, r[((p \Rightarrow q) \wedge (q \rightarrow r)) \supset (p \Rightarrow r)]$$

P1 says that if *it would have turned out that q, had it turned out that p*, and *q implies r*, then it also *would have turned out that r, had it turned out that p*. The intuition behind P1 is just that if things had turned out such that some state of affairs Q had obtained, then every state of affairs that Q implies would also have obtained. This is certainly an attractive principle.

Premise (14) depends on an even more plausible-sounding principle:

$$[\text{P2}] \quad \forall p[p \Rightarrow p]$$

If '$\Rightarrow$' conditionals didn't support P2, they would seem very suspect indeed, and we would be well advised to avoid placing much theoretical hope in them.

If (10)–(SIS) is a sound argument, then PIE_y does not succeed in escaping the Sisters Problem, for it will instruct us to map at least one world onto the value *true* when evaluating the primary intension of 'sisters are male siblings'. Just as PIE_c'' gives every sentence a contingent primary intension, so will PIE_y.

There is a way, however, that Yablo could respond to this argument.[21] To use P1 in an argument against Yablo is basically to beg the question. Recall the claims (7) and (SIS):

(7) "if 'sister' refers to male siblings then sisters are male siblings" is *a priori*.

(SIS) If it had turned out that in the actual world we used 'sister' to refer to male siblings, it would have turned out that sisters were male siblings.

---

[20] A fuller argument for the *apriority* of the left-to-right direction of T-sentences is given in §1.4.2 of the first Chapter of this thesis.

[21] And indeed has, in personal communication.

61

Yablo claims that (7) is true, but (SIS) is false. However, if P1 and P2 are true, this can't be the case. To see this, consider an instantiation of P1 for some sentences p, q, and r, where p and q are the same. We get: $\ulcorner((p \Rightarrow p) \wedge (p \rightarrow r)) \supset (p \Rightarrow r)\urcorner$. Now combine this with an instantiation of P2 for p. Together they entail: $\ulcorner(p \rightarrow r) \supset (p \Rightarrow r)\urcorner$. This means that if p *a priori* entails r, then if it had turned out that p, it would have turned out that r. It's clear that if this is right, then it can't be that (7) is true and (SIS) is false. So, it's question begging to use P2 in the above argument.

Fair enough. It's *not* question begging, however, to use P2 and (10)-(SIS) as part of a different kind of argument—an argument aimed at showing that our strong intuitions about $\Rightarrow$ conditionals lead us in inconsistent directions. Like Yablo, I have the intuition that (SIS) is false. However, my intuition that P2 is true is just as strong (in fact, it's stronger). I have no idea which intuition to give up, now that I see that I can't consistently keep both. Those of us that have inconsistent intuitions about $\Rightarrow$ conditionals would do well to avoid putting them to work in defining primary intensions—or anything else, for that matter.

2.6 An additional problem for PIE_c'' and PIE_y

In addition to the Sister Problem, which I have argued destroys any possibility of defending a strong modal rationalism understood according to PIE_c'' or PIE_y, there might be a second source of problems for building a strong modal rationalism on either of these primary intension evaluators. I say 'there might be' because whether or not one sees a problem here will depend on whether one considers certain claims to be *a priori*. I think that the claims in question are *a priori*, so I see a problem. However, since one will only see a problem if one shares my intuitions about what sorts of things should turn out to be knowable *a priori*—intuitions that I freely admit are controversial—I'll keep my remarks here brief.

The problem, as I see it, is this: Neither PIE_c'' nor PIE_y captures all cases of the *a priori*, so SMR fails in the right-to-left direction, when understood according to

them.[22]

Chalmers' evaluator, PIE_c, fails for cases of the "Cartesian *A Priori*" like the following:

(15) I exist.

(16) I am thinking.

To see that PIE_c'' will render (15) 1-contingent, first consider the primary of intension of its negation, 'I don't exist', with respect Max Black's world containing just two qualitatively indistinguishable iron spheres (call this world $w_{spheres}$ and its canonical description $D_{spheres}$). To do this, PIE_c'' instructs us to ask if the following is *a priori*: *if $D_{spheres}$, then I don't exist*. This indeed does look *a priori*; the claim that $w_{spheres}$ is the actual world *a priori* entails that I don't exist. So, there is at least one world that the primary intension of 'I don't exist' maps onto *true*. The primary intension of 'I exist' will thus have to map this world onto *false*, so (15) will be 1-contingent, even thought it is *a priori*.

And to get a counteractual world bearing witness to the 1-contingency of (16), all we need to do is consider a world, w_{splat}, where I have recently been crushed by a falling anvil. It does seem *a priori* that if this world is actual, then it is not the case that I am thinking.

I should say that Chalmers would admit and welcome these results. He wants there to be counteractual possibilities where he doesn't exist and where he's not thinking—and the canonical descriptions of these possibilities will *a priori* entail that he doesn't exist and that he's not thinking. In order to preserve SMR, then, he needs to deny that (15) and (16) are *a priori*. He certainly wouldn't be alone in denying this, but I find the view that (15) and (16) are knowable only *a posteriori* to be quite unattractive.

Yablo's PIE_y is going to agree with PIE_c'' on the primary intensions of (15) and (16) at $w_{spheres}$ and w_{splat}. If it had turned out that the universe actually contained only two iron spheres, it would have turned out that I didn't exist; and if it had

---

[22]SMR, remember, says that primary necessity $\equiv$ *apriority*.

turned out that I actually had just met my fate underneath a falling anvil, it would have turned out that I wasn't thinking right now. But PIE_y is also going to run into a problem with cases of the "Kaplanian *A Priori*", such as:

(17) I am here.

This looks plausibly *a priori*, but the primary intension, construed according to PIE_y, is contingent. In order to see this, I don't need to contemplate strange worlds containing only iron spheres or tragic worlds involving my gruesome and untimely death. All I need to do is consider a world where I decided to work this afternoon in the kitchen instead of the living room. If it had turned out that I was actually in the kitchen working, it would have turned out that I wasn't here (for where I am—the place 'here' denotes for me right now—is a place in the living room). Nevertheless, I think I do know *a priori* that I am here.[23]

But, again, all of this in a bit contentious, and we've already seen that SMR, interpreted according to either PIE, runs headlong into the Sister Problem, so let's move on.

2.7 A Proposal

2.7.1 Idiolects, Conditionals, and Stubborn Interpretations

Okay then, is there a way of characterizing primary intensions that avoids the Sister Problem and is friendly to SMR? I think that it is certainly too early in the game for the modal rationalist to admit defeat. In this section I'll offer a proposal for how to construct primary intensions in a modal-rationalism-friendly way.

We have already seen that *a priori* material conditionals, Yablo's $\Rightarrow$ conditionals, and *a priori* indicative conditionals will not work. I think that the solution is to be

---

[23]But how. someone will object, do I know *a priori* that I am in any determinate place right now? Perhaps I am hurtling through space at break-neck speed. I'm not sure what I want to say about this case. The issue can be avoided, however, by retreating to:

(17′) I am not elsewhere.

PIE_y is going to deem this 1-contingent, for if it had turned out that I was in the kitchen, it would have turned out that I was not here, but elsewhere.

found by building off a straightforward indicative conditionals approach.

At first, it looks like such an approach is going to immediately founder on Yablo's Sister Problem. Consider the following primary intension evaluator:

[PIE_{ind}] if w is actual, is it the case that S?

This fails to escape our problem; for the following looks true:

[SIS'] If the word 'sister' actually refers to male siblings, then sisters are male.

If this is right, then PIE_{ind} can't help us define *primary intension* in a way that would lend support to SMR.

I agree that there is a reading of (SIS') on which it is a true playground conditional. But I think that there is a another reading of (SIS') that most people can be brought to hear—one on which it comes out *false*. First, note that most people would assess the following conditional as true:

(18) If the word 'sister' actually refers to male siblings, then what I have meant all along by 'sister' isn't what everyone else has meant.

(18) just says that if 'sister' actually refers to male siblings, then *my idiolect has been diverging from English on the word 'sister'*. I think that this idea sheds some light on what's going on with (SIS'). When we evaluate (SIS') as true, we are letting the antecedent of the conditional bully us (at least temporarily) into giving up our idiolect, and adopting one that is in conformity to the way our language is supposed, by the antecedent, to be.

However, we *could* evaluate these conditionals more stubbornly, couldn't we? That is, we could refuse, even provisionally, to update our idiolect to conform with the way the antecedent supposes our language to be. Evaluated stubbornly, I think (SIS') is *false*. We can even force the stubborn reading by rewriting (SIS') as (SIS''):

[SIS''] If the word 'sister' actually refers to male siblings, then 'sisters are male' is true in my pre-antecedent idiolect.

This is clearly false, for, in my pre-antecedent idiolect—the idiolect I spoke (and know I spoke) before the antecedent of the conditional invited me to start making suppositions—'sisters are male' is *analytically* false.

If this is right, then perhaps a definition of *primary intension* that builds off the question,

[PIE] If w is actual, is S true in my pre-antecedent idiolect?

will avoid the Sister Problem.[24] A little reflection shows that the original two-dimensionalist responses to Kripke still go through when primary intensions are understood according to PIE. For example, the following seems true: if Twin Earth is actual (i.e., the stuff in the lakes and streams, etc., is actually XYZ), then 'water is XYZ' is true, even in my pre-antecedent idiolect. My pre-antecedent idiolect doesn't take a particular stand on the chemical structure of water; although it does determine that 'water' will rigidly designate the stuff, whatever it is, that has the same deep, explanatory, features as the watery stuff around here. If that stuff is actually XYZ, then 'water is XYZ' is true in my pre-antecedent idiolect.

I certainly do not wish to claim that this account is without difficulty. I'll end by discussing three objections. The first objection has to do with my free use of this notion of *idiolect*. The second objection maintains that, even with this free use, the account doesn't manage to avoid the Sister Problem. The third, and final, objection holds that, if everything I have said is right, *a priori* knowledge comes too cheaply. Let's look at each in turn.

2.7.2 Objection 1: Idiolects and Narrow Content

First, one might object that wheeling in idiolects to define primary intensions is tantamount to wheeling in *narrow content*. Narrow content is a type of content that

---

[24]The account I'm suggesting here attempts to ground the notion of treating worlds as *counteractual* in terms of our intuitive evaluations of a certain class of indicative conditionals. In his paper "Indicative and Subjunctive Conditionals" (Weatherson 2001), Brian Weatherson does things the other way around—that is, in giving a semantics for indicative conditionals, he employs the notion of treating worlds as counteractual. I think that the resulting account is quite elegant, but, to my mind, it puts the cart before the horse. I think that we have less of a pre-theoretical grip on counteractual evaluation than we do on indicative conditionals.

putatively accompanies mental states or utterances purely in virtue of the intrinsic properties of the thinker/speaker. (*Wide content*, on the other hand, depends on both the agent and features of the world external to her.) If this charge—the charge of smuggling in narrow content—is right, then a problem arises for my account: Narrow content is just as in need of explication as primary intensions, and one might have thought that primary intensions were going to give us some sort of grip on narrow content, especially given the strong modal rationalist thesis that necessary primary intensions go along with *apriority*. The two notions are, at the very least, partners in crime. We don't want to define primary intensions using narrow content.

I think that this first objection is misguided. Wheeling in idiolects is not the same as wheeling in narrow content. For, if there is a narrow/wide content distinction, it applies to expressions in idiolects just as much as it applies to expressions in shared languages. The temptation to think that all "idiolect content" is narrow derives from the temptation to think that meaning in one's idiolect must be *transparent*—that is, the speaker can't be wrong about what expressions mean in his or her own idiolect. Well, what exactly does *that* mean? Perhaps it is best illustrated with an example:[25] Suppose that Kim has a lot of beliefs that we would attribute as arthritis beliefs. Many of them are true. Kim believes that she has arthritis in her wrists, that it is a painful disease, that it has something to do with tissue inflammation, etc. She also has the false belief that she has arthritis in her thigh. When she tells the doctor of this last belief, the doctor tells her what any decent dictionary could have: arthritis is an inflammation of the joints, and cannot be had in one's thigh. Kim accepts this correction with a little embarrassment. What she has discovered here is that the meaning of the word 'arthritis' in her idiolect doesn't match up with the meaning of the word in her language community. But while she was confused about what 'arthritis' means, there's one thing that she surely had right: she at least knew what 'arthritis' meant in her own idiolect. That is, she knew what *she* meant by 'arthritis'. That, many feel, she could not have been wrong about. I think that this view, which friends of narrow content would embrace, is right.

---

[25]The example comes from Tyler Burge's "Individualism and the Mental" (Burge 1998b).

But this does *not* entail that all idiolect content is narrow. Many expressions in my idiolect will have widely individuated truth conditions. The reason is this: even if I do have a special access to a certain kind of meaning in my own idiolect, it is often the case that these meanings underdetermine truth-conditions—that is, a special role is given to the world to play in determining the full truth-conditional content of the expressions. Putnam's elm/beech example[26] can be commandeered to illustrate what I have in mind. Like Putnam, I can't tell the difference between an elm tree and a beech tree. Nevertheless, let's agree with friends of narrow content that I do know exactly what both 'elm' and 'beech' mean in my idiolect. Oversimplifying, 'elm' means *those deciduous trees that the tree experts would call 'elms'*, and 'beech' means *those deciduous trees that the tree experts would call 'beeches'*. What's important to note is that these idiolect meanings leave some work for the external world to do in determining the extensions of 'elm' and 'beech'.

Likewise I know what the word 'water' means in my idiolect. Again, to over-simplify somewhat, it means something like *the dominant clear liquid flowing in the lakes and streams and out the out of the taps, etc.* or perhaps *the stuff that has been responsible for my "waterish" appearances and any other stuff that shares its deep, underlying, structure*. Again, this kind of meaning leaves an open a slot for the world to fill in order to determine extension. If I have been on Earth all along, then the word 'water', in my idiolect, is going to refer to H_2O; if I've been on Twin Earth all along, then it will refer to XYZ.

The upshot is this: While the notion that expressions have narrow content might be particularly at home with idiolects, because of the plausibility of transparency, expressions in idiolects clearly also have wide content. It's unfair to bar idiolects from discussion just because the notion of narrow content can be applied to them with some plausibility (or with more plausibility than it can be applied to tokened expressions understood as bits of shared natural language). The notion of an idiolect, by itself, does not take sides in the debate between friends and foes of narrow content, and neither, I claim, does my formulation of PIE.

---

[26]See his 'The meaning of 'meaning'' (Putnam 1996).

2.7.3 Objection 2: Return of the Sisters

Second, one might object that adopting PIE to evaluate primary intensions doesn't really solve the Sister Problem. Perhaps it is true that even if 'sister' actually means male sibling, 'sisters are not male' is true in my pre-antecedent idiolect. This solves the original problem, as Yablo stated it for PIE_c, but can't we make a simple revision to the example? Instead of considering a counteractual world where the shared English word 'sister' refers to male siblings, consider a counteractual world where 'sister' actually refers to male siblings *in my pre-antecedent idiolect*. If we want to evaluate the primary intension of the sentence 'sisters are male' with respect to this world, (PIE) instructs us to evaluate:

(19) If 'sister' actually refers to male siblings in my pre-antecedent idiolect, then 'sisters are male' is true in my pre-antecent idiolect.

And, one might argue, this looks true.

I'm not quite sure what to make of this conditional. Part of the problem is that I'm inclined to think that the transparency thesis for idiolects is right. In my idiolect, 'sisters are females, not males' is nothing short of an *analytic* truth. I know with absolute, unshakable, certainty that the word 'sister' means *female sibling* in my idiolect. I find that I'm unable even to *entertain* the notion that it means anything else. I believe that this is also what it means in English, but here I'm open to the outside possibility that I might be wrong. But when it comes to my own idiolect, I won't admit this even as an outside possibility. Because of this inclination, I find (19) very difficult to evaluate. I'm certain that both the antecedent and the consequent are false. If (19) was supposed to be a material conditional, then it would follow that it is trivially true. But it's not supposed to be a material conditional, it's supposed to be a regular, "garden variety" conditional. How do we evaluate such conditionals for truth when we find that we are unable to entertain the antecedent? I find it difficult to get solid intuitions in these cases.

However, for those who are tempted to evaluate (19) as true, I'd like to make it less tempting. One way this might be accomplished is by trying to get clearer

about the role that the term 'pre-antecedent' is playing in the consequent of the conditional. The antecedent of (19) has set up a certain supposition about how things are in the actual world. The consequent then makes a claim that is supposed to hold under that supposition. But there is a device in the consequent, this word 'pre-antecedent', that, I believe, manages to smuggle some linguistic "stuff" out of the scope of the supposition. The consequent makes a claim about what's true *in the idiolect I spoke before I started making suppositions.* Even though the relevant supposition was *about* my pre-supposition idiolect, it still makes sense to ask what is true (on this supposition) *in the idiolect as it stood before the supposition was made.*

There are other words in the English language that work roughly like I'm suggesting 'pre-antecedent' works. Take the word 'actual', for instance. This word sometimes functions as a device for "scope defiance"—it can be used to smuggle certain linguistic items out of the scope of a supposition, or hypothesis, or even a fictional context. For example, consider the following sentence:

(20) If you grow an inch, then you will be taller than your *actual* height.

There is one reading of (20) on which it is clearly absurd. No matter how much you grow, you'll never be taller than you are. But there is a second reading of this sentence, which can be glossed roughly as follows:

(20′) There's a certain height x that you actually are, and if you grow an inch, you'll be taller than x.

On this second reading, the word 'actual' brings some material outside the scope of the supposition that the antecedent has set up. So, the word 'actual' seems, at least sometimes, to work roughly the way I'm suggesting 'pre-antecedent' works.

More work would need to be done to flesh out this suggestion and transform the impressionistic remarks I've made here into a full-fledged defense. But that is a task that I will not take up here. Instead, I'll turn to the third objection.

2.7.4 Objection 3: Overgeneration of *A Priori* Knowledge

In my response to the first objection, I confessed my attraction to a transparency thesis for idiolects. There is a sense in which meaning in one's own idiolect is known in an immediate and *a priori* way. Doesn't this commit me to saying that I know the following two claims *a priori?*

> (26) 'bachelor' refers to bachelors.

> (27) Bachelors are unmarried males.

And doesn't this, in turn, commit me to saying that I also know the following *a priori?*

> (28) 'Bachelor' refers to unmarried males.

Well, sort of. But this doesn't commit me to thinking that English is learnable *a priori.*

In the marketplace of ordinary, every day, language use, I speak in my idiolect, intending (in standard situations) to communicate the idiolect meanings of the sentences I use. This would be a bizarre practice if I didn't believe that my idiolect matched up relatively well with those of other English speakers. I do believe this, and for the most part, all goes well for this belief. When I discover that I'm wrong in some case—that my idiolect diverges from those of other members of my language community on some point—I adjust it accordingly. If I think that arthritis is any kind of tissue inflammation, and my doctor tells me that it's only inflammation of the joints, I update my idiolect. From now on, when I talk about arthritis, I mean to be talking about an inflammation *of the joints.* (Unless, of course, I forget that the doctor corrected me, or become otherwise confused.)

Questions of *apriority*, I want to claim, are sensitive to who is doing the asking, and how they are doing it (i.e., what language are they using). If I am rehearsing the argument (26)–(28) to myself, in my own idiolect, I think that (26) and (27) are *a*

priori and do *a priori* entail (28). But, luckily, (28) is also something that I know—and know *a priori*—to be true in my own idiolect. This is, of course, just what we would expect if the transparency thesis for idiolects is right.

So, there is no problem if I understand the argument according to my own idiolect. If, however, I am being asked to assess the status of these claims by a fellow English speaker, understanding them according to our common English meanings, I will have to say that I do *not* know (26) *a priori*, for I don't even know *a priori* that 'bachelor' is an English word, let alone that it refers to bachelors. In this case, then, the argument doesn't get off the ground. So we do not yet have a reason to think that my account over-generates *a priori* knowledge.

Chapter 3

A GUIDE TO MODAL GUIDANCE

3.1 Introduction

Complaints are in the air about conceivability arguments. Some are directed at particular conceivability arguments; others are more general in nature, casting a shadow of suspicion over the whole philosophical—and folk—practice of drawing modal conclusions from premises concerning conceivability. In this paper, I defend the use of conceivability arguments. I start by looking at the charge that conceivability arguments get bogged down in an infinite justificatory regress. I consider two versions of this complaint and argue that both versions can be answered in a relatively straightforward way. Showing where these regress arguments go wrong, however, is interesting, as it helps us to get into focus some principles that will need to be included in any satisfactory modal epistemology. After dispensing with the regress challenge, I will bring these principles to bear on a second complaint that one encounters in discussions of modal epistemology, this one aimed more narrowly at arguments from *ideal* conceivability. The problem here is supposed to be that such arguments are *useless*, for ideal conceivability always falls outside the epistemic reach of non-ideal reasoners such as ourselves. I will argue that this pessimistic view is mistaken. However, the picture that will emerge will also lead me to reject an (overly optimistic) *modal rationalism*, which says that ideal conceivability gives us *a priori* access to modal truth. I will suggest that our access to modal truth is, in general, thoroughly *a posteriori* in

nature. In the end, however, I will suggest a possible avenue for satisfying the spirit, if not the letter, of the optimistic view. The idea is to move from the notion of *a priori* access to a more permissive notion of *access from the armchair*.

3.2 The infinite regress argument

3.2.1 Where not to look for a regress

I have heard it said that conceivability arguments in general—and David Chalmers' conceivability argument against materialism in particular—are doomed, for they lead to an infinite justificatory regress. The idea is simple: Folks use conceivability arguments because they think, not implausibly, that conceivability is a guide—our only real guide— to modal truths concerning possibility. But conceivability is itself a modal notion. Just taking it at face value, 'ϕ is conceivable' can be analyzed as 'It's possible to conceive of ϕ'. Since conceivability evidence is our guide to possibility, it looks like, in order to justify our belief that ϕ is conceivable, we will need a second conceivability argument starting from the premise 'It's conceivable that one could conceive of ϕ' (or 'conceiving of someone conceiving of ϕ is conceivable'). And now we have *another* modal claim that must be established by consulting conceivability evidence. We could go on like this for as long as you like. Attempting to justify possibility claims with conceivability arguments will always lead one into an *infinite regress*, and this renders such arguments completely useless. They just can't get off the ground.

Let's lay out the argument a little more precisely. It goes something like this:

(1) Our evidence for claims of the form 'possibly ϕ' comes from conceivability arguments.

(2) Conceivability arguments involve a premise of the form 'ψ is conceivable'.

(3) Claims of the form 'ψ is conceivable' are analyzable as claims of the form 'possibly someone conceives of ψ'

From (1)–(3), we can conclude that:

(4) Conceivability arguments involve a premise our evidence for which comes from a conceivability argument.

And finally, from (1), (4), and something like:

(5) If arguments of type T always involve a premise our evidence for which comes from another argument of type T, then justifying any claim with an argument of type T leads to an infinite justificatory regress.

we can get that:

(6) Justifying possibility claims always involves an infinite regress of conceivability arguments.[1]

There are two things that should be noted here. First, (1) doesn't say anything about how much weight to assign to conceivability evidence. There is a range of views that could be adopted, from taking conceivability to *entail* possibility, to taking conceivability to be an easily defeated, non-demonstrative, guide to possibility. Second, the argument does not presuppose any particular theoretical account of conceivability. The only constraint it places on conceivability is encoded in (3), which says that 'ψ is conceivable' is to be understood as 'it is possible to conceive of ψ'.[2] If the regress argument works, it works for *any* such account.

Although (3) seems plausible enough, there are a few notions of conceivability for which it will not be satisfied. Chalmers, in Chalmers 2002a and elsewhere, makes a distinction between *negative* conceivability and *positive* conceivability (pp. 149-156). Something is *negatively conceivable* iff it is *not a priori not the case*. *Positive conceivability*, on the other hand, requires something more: One must be able to "form

---

[1]This conclusion should not be read, "Justification for (believing) possibility claims always comes from an infinite regress of conceivability arguments". On some epistemological views, what *justifies* a given belief can come apart from the *evidence* the believer could marshal in defense of the belief. Indeed, I might not be in a position to adduce *any* non question-begging evidence for a belief, and for all that, it might still be justified. This, for example, is James Pryor's (2000) take on certain basic perceptual beliefs. In any event, as I am using the phrase 'justifying possibility claims' in (6), it has to do with marshaling evidence or giving reasons.

[2]As Stephen Yablo has pointed out to me, this constraint could be questioned. Just as it would be a mistake to analyze *desirability* in terms of what is *possible to desire*, one might argue that it is likewise a mistake to analyze *conceivability* in terms of what is *possible to conceive*. I'll set this interesting point to the side and assume that the analytical move is much less problematic in the case of *conceivability* than in the case of *desirability*.

some sort of positive conception" of a situation that is correctly described by the claim in question (p. 150). Positive conceivability, and most other conceivability notions, will satisfy premise (3) of the regress argument, but *negative conceivability*, and a few others, will not. However, notice that even Chalmers' *negative* conceivability is characterized modally. A claim S is said to be negatively conceivable just in case $\neg S$ is not *a priori* —that is, just in case it is *not possible* to know with *a priori* justification that $\neg S$. Given this characterization, it is easy to see that, with a little bit of effort, we could produce a regress argument very similar to (1)–(6) for negative conceivability. I won't test my readers' patience with the details.

You might be wondering: Does anybody think that conceivability arguments are really undermined by this kind of regress? Anecdotal evidence indicates that the answer is 'yes'; this is an argument that occurs to many people when they are confronted with the idea that conceivability is our guide to possibility.[3] And, in any event, I think it is interesting and instructive to see why there really is no regress problem.

There are two general strategies that one could adopt to show this. The first is to grant that there is a regress in the vicinity, but to deny that this is a problem. We cause the regress by trying to answer too many questions. It might be true that, when asked to provide non question-begging *evidence* for a modal claim, the best we can do is to appeal to some conceivability considerations. But that does not mean that our modal beliefs are always *justified* by inferences from conceivability. Perhaps some of our modal beliefs aren't inferred from *anything*—they are justified *immediately*.

I think that this first strategy is interesting, and certainly worth exploring, but I would like to set it to the side for the moment and pursue a second strategy. Whereas the first strategy granted that there might be a regress problem (if one consented to

---

[3]Andrew Bailey has pursued a regress objection against conceivability arguments in a series of talks, and in an unpublished paper (Bailey ms) which can be found on his website (http://www.uoguelph.ca/~abailey/papers/index2.html). In the paper, Bailey focuses on arguments from *ideal* conceivability (see §3.2.2 of the current paper for a definition of ideal conceivability), but the general form of his argument is the same as the one I have given here (see the part of Bailey's paper where he sketches the abstract form of the argument). After discussing how to respond to the simplest instantiation of this argument form, I'll argue that no regress problem emerges when we move to ideal conceivability.

play the "provide evidence for my modal beliefs" game), and then tried to show why this should not trouble us, the second strategy takes a more straightforward approach, denying that there is any such regress.

The second strategy says: Contrary to what you might have heard, no one *really* thinks that conceivability is our *only* guide to possibility. It's just that our other main guide to possibility is so obvious, so mundane, so dull and uncontroversial, that it doesn't get much press. The other guide I have in mind is just good old *actuality*.[4] Here's a principle that is beyond reproach:

[ACTUALITY] Actuality is a *perfect* guide to possibility.

If it is the case that *p*, then, obviously, it is also *possible* that *p*. So, to the extent that I have evidence for believing any proposition *p*, I also have evidence for believing *possibly p*. This is a rich, if not always terribly interesting, source of modal knowledge.

An example: I recently had to purchase a new pair of reading glasses, after having devoted the full force of my body weight to the destruction of the old pair. Hearing of this, my optician suggested that I purchase some frames made of a material called 'flex titanium' (and that I start watching where I step). In the course of making his sale, he made an interesting claim: He said that the frames are *unbreakable*. Naturally, I snatched them up and bought them straightaway. Now, it *is* true that you can bend these frames almost in two, and they will pop back to their original shape, but when my optician said

(UB) These frames are unbreakable

was he right? Notice that this is a *modal* claim. It amounts to something like 'it is not possible that these frames break,' or, to be more charitable to my optician, perhaps, 'it is not possible that the laws of nature remain as they are and these frames break (under normal temperature conditions, etc.)'.

---

[4]It's not as though this source of modal knowledge hasn't gotten *any* press. The first footnote of Yablo 1993 reads, "Sometimes, of course, [telling that a proposition is possible] is easy. If a proposition *p* is true, and known to be, then its possibility can be inferred from *p* itself. The problem is to find grounds for thinking a proposition possible which is not known to be true, most obviously because it is false."

Suppose my optician was wrong, and it *is* possible to break the frames. No one thinks that the only way I could come to have this modal knowledge is by going through some kind of conceivability argument.[5] If I take the frames, bend them, and they snap in two, then I immediately learn two things—one non-modal, and one modal. The non-modal thing I learn is that these frames broke. The modal thing I learn is that it is *possible* to break them.

Or suppose that someone tells me that it's not possible to solve a certain Sudoku puzzle (because it has no solution). If I then sit down and solve the puzzle, I learn that it *is* possible to solve the puzzle. And, again, no conceivability argument is necessary here. Conceivability is just *one* of our guides to possibility, it is not the only one.

Once we remember this, it is easy to see that the infinite regress argument, (1)-(6), doesn't go through. Whatever rhetorical force it has depends on amphiboly—the argument trades off a structural ambiguity in the first premise. (1) could mean either (i), that possibility claims *can* be argued for with conceivability arguments, or (ii), that conceivability arguments are our *only* way of arguing for possibility claims. If we understand the premise along the lines of (i), it seems true, but it does not license the inferences that are made in the regress argument. Perhaps conceivability claims *can* be established with conceivability arguments, but it is consistent with this to claim that they can also be established through *actual episodes of conceiving*. If we understand (1) along the lines of (ii), on the other hand, it licenses the inferences all right, but the premise itself is obviously *false*. Either way, then, there is no threat of regress.

So, once again, we have at least two guides to possibility:

[CONCEIVABILITY] Conceivability is a guide to possibility (and inconceivability is a guide to impossibility).

[ACTUALITY] Actuality is a perfect guide to possibility.

And, so far, we don't have a regress problem.

---

[5]Though, in my opinion, this could, in the right circumstances, be *a* way of coming to have it.

3.2.2 Where to look for a regress

But perhaps we can cook up a slightly more sophisticated version of the regress argument. This version focuses on conceivability arguments that employ the notion of *ideal conceivability*—arguments like David Chalmers' zombie argument against materialism. Chalmers distinguishes between *prima facie* conceivability and *ideal conceivability* (2002a, pp. 147-49). The *prima facie* conceivability of a claim S (for some particular agent A) just involves A finding that, at least on first appearances, it is possible for S to "[pass] the tests that are criterial for conceivability" (p. 147).[6] Chalmers thinks that, while *prima facie* conceivability is a rough-and-ready guide to possibility, *ideal conceivability* is a *perfect* guide. He says,

> [W]e can say that S is ideally conceivable when there is a possible subject for whom S is *prima facie* conceivable, with justification that is undefeatable by better reasoning. The idea is that when *prima facie* conceivability falls short of ideal conceivability, then the claim that the relevant tests are passed will either be unjustified, or the justification will be defeatable by further reasoning. (p. 148)

Conceivability, as I have already remarked, is a modal notion. Ideal conceivability, as Chalmers defines is, is thus *doubly* modal:

IDEAL CONCEIVABILITY:

(a) It is *possible* to *prima facie* conceive of S

(b) in a way that is *not possible* to undermine with better reasoning.

I claimed earlier that, using ACTUALITY, we can argue to the conclusion that S is conceivable just on the basis of our actually having conceived of a verifying scenario. But when it is ideal conceivability that is at issue, things are more complicated. In order for S to be ideally conceivable, both modal clauses will need to be satisfied. The question is: Can we justify a belief that they are all satisfied without involving ourselves in an infinite regress of ideal conceivability arguments?

---

[6]The relevant tests will depend on exactly what notion of conceivability is in play. The details don't matter for present purposes.

Well, there is certainly a straightforward way that I convince myself that clause (a) is satisfied. For I, myself, might be the kind of agent that is required—i.e., a possible agent for whom S is *prima facie* conceivable. And I could know this simply on the basis of an actual episode of conceiving. ACTUALITY can give us at least this much.

But what about clause (b)? This is a bit more worrisome. ACTUALITY is of no use here, for clause (b) is an *impossibility* claim, not a *possibility* claim, and ACTUALITY gives us no guidance concerning impossibility. Here is an ACTUALITY-like principle that would do the trick, but for its patent falsity:

[NOT-A-GUIDE] Nonactuality entails impossibility.

Since ACTUALITY is not going to help us, how about CONCEIVABILITY? It doesn't look like this can help either, for CONCEIVABILITY can only help us justify impossibility claims if we already have some independent grip on a subclass of the impossibility claims. Here's why: What we are trying to establish is that it is *not possible* to defeat-with-better-reasoning my justification for thinking that I have conceived of S. If we are going to use CONCEIVABILITY to establish this, the relevant conjunct will be the one that says that *inconceivability is a guide to impossibility.* In order to use this guide, however, we will first need some reason for thinking that it is inconceivable that my justification for thinking that I have conceived of S is defeated by better reasoning. But what reason do I have for believing this last impossibility claim? Again, my evidence cannot come from ACTUALITY, and CONCEIVABILITY is only going to help if I have an independent grip on a subclass of the inconceivable. In the absence of such an independent grip, justifying impossibility claims *is* going to involve us in an infinite justificatory regress.

If there is an infinite regress problem, here's where to find it.

3.2.3 Why there isn't really a regress problem

The way out of this regress is, in its basic outline, the same as the way out of the first regress (see §3.2.1). We got out of the first regress by noticing that conceivability is not our only guide to possibility—we also have *actuality*. The way out of the present regress is to notice that there is (at least) a *third* guide to modal truth.

It is relatively easy to independently motivate the need for such a guide: The present "regress" is not caused by any special feature of the notion of ideal conceivability. All that we needed to get the regress going was an impossibility claim—*any* impossibility claim would have sufficed. ACTUALITY is going to be insufficient for justifying any such claim, and trying to justify it with CONCEIVABILITY is going to involve taking on board a premise which is itself an impossibility claim. When we try to justify the latter premise, we will run into the same problem that we had with the original impossibility claim. The wheels of regress have been set in motion.

But we *are* sometimes justified in believing impossibility claims. (For now, let's leave it an open question whether this means that we are sometimes in a position to adduce *evidence* or *reasons* for such beliefs.[7]) It is impossible that $2 + 2 = 5$, it is impossible for there to be female bachelors, it is impossible that the laws of physics will stay the same and I will enter a black hole unharmed. I am fairly certain of each of those claims. I think I could also be justified in coming to believe that my flex titanium frames are unbreakable (in normal conditions, etc.), that a certain Sudoku puzzle is unsolvable, that such and such a justification is undefeatable by better reasoning, and so on. Unless we are going to be modal skeptics about all impossibility claims (and I don't see any reason why we should be), we are going to have to allow that we are sometimes justified in believing these sorts of things. Since we cannot produce any (non question-begging, non regress-involving) evidence for such claims on the assumption that our only guides to modal truth are CONCEIVABILITY and ACTUALITY, we must draw one of two conclusions:

> (I) We are sometimes justified in believing impossibility claims, even

---

[7]See f.n. 1.

though we are not able to provide—nor are we even in possession of—non question-begging evidence for their truth.

(II) We have (at least) a third source of evidence for modal truth.

Either way, we have a solution to the regress problem.

To clarify the issue between (I) and (II), it is helpful to distinguish between two ways in which something could be a *guide* to modal truth. The word 'guide' is ambiguous between a *causal* reading and an *evidential* reading. According to the evidential reading, a guide is a *source of evidence*, in the sense of being (or entailing) a body of propositions from which another body of propositions can reliably be inferred. Propositions concerning scores on the LSAT exam, for instance, are supposed to be an evidential guide to propositions concerning academic potential with respect to law school. According to the causal reading, on the other hand, a guide is *something that actively leads one to a certain goal*. Think, for example, of seeing eye dogs and tour guides. When we call a seeing eye dog a 'guide', we certainly don't mean that it stands in some sort of evidential relation to a body of propositions, we simply mean that the dog's job is to (physically) help one navigate around the world.

CONCEIVABILITY and ACTUALITY are plausibly evidential guides to modal truth.[8] (II) claims that there is (at least) a third evidential guide to modal truth. (I), on the other hand, claims that evidential guides are not always necessary for justified modal beliefs. Nevertheless, it seems plausible that even (I) implies the need for some additional *causal* guide to modal truth (perhaps a faculty of some sort). If these modal beliefs are indeed *justified*, they presumably are causally sponsored by *something*, and this something, whatever it may be, is at least a causal guide to modal truth.

Either way, the regress problem disappears. If the third guide is inferential, then we have a way of grounding a subset of the impossibility claims that does not appeal to CONCEIVABILITY or ACTUALITY considerations, and there is thus no general threat that we will need a regress of ideal conceivability arguments to establish modal claims. On the other hand, if the third guide is a guide only in the causal sense, then, while we might run into a regress if we continue to play the "provide evidence for our modal

---

[8]This is how they are used in conceivability arguments, for instance.

beliefs" game, there is a principled reason to opt out of the game at a certain point: Some of our modal beliefs are not hostage to the availability of evidence for their justification.[9]

At any rate, I think we are going to need to let in two more guides—one of which is plausibly a guide in the causal sense (though this matter is unclear), and the other a guide in the evidential sense. First, the causal guide: *modal intuition*.[10] Here is how George Bealer characterizes intuition:

> By intuitions here, we mean *seemings*: for you to have an intuition that A is just for it to *seem* to you that A. Of course, this kind of seeming is *intellectual*, not experiential—sensory, introspective, imaginative. (2002, p. 73)

Modal intuition, then, is just intuition constrained to matters modal.

Modal intuition seems to be the source of our justification for believing certain modal claims. Our justification for thinking that it's impossible that $2 + 2 = 5$ is just that it *seems* that $2 + 2 = 4$, and that things could not have been otherwise. And likewise, our justification for thinking that it's impossible to be a female bachelor is that it *seems* that these concepts necessarily exclude each other. The idea that claims like these are typically known through conceivability arguments was never very plausible. It is certainly true that we cannot conceive of a situation where $2 + 2 = 5$ or where there are female bachelors, but no one really makes an honest attempt to conceive of these things. We can just *see* at the outset that the enterprise is doomed.[11]

So, we will need to admit modal intuition as a third distinct guide to the modal facts:

---

[9]This is just to repeat the first strategy for responding to the original regress problem (see §3.2.1, page 76).

[10]George Bealer (2002) would object to my keeping CONCEIVABILITY around after I have let in intuition. Unfortunately, it would take me too far off topic to discuss why Bealer thinks this, and why I think he is mistaken.

[11]It is an open—and very interesting—question *how* we do this. Different accounts of intuition offer different explanations. These accounts also disagree on whether or not intuition is a source of *a priori* knowledge. I would like to address these issues in future work, but, for present purposes, I want to remain neutral on them. All I want to claim here is that we *do* have these seemingly spontaneous modal seemings, and they *do* seem to be a source of modal knowledge.

[MODAL INTUITION] Modal intuition is a guide to modal truth.

But is this a guide in evidential sense, or in the causal sense? What one wants to say about this will likely depend to a large extent on one's prior epistemological commitments. Some will say that there is a *faculty* of modal intuition, which is responsible for spontaneous modal "seemings" in the same way that perceptual faculties are responsible for spontaneous perceptual seemings (e.g., there seeming to be something heavy here).[12] Modal intuition could thus be seen as a causal source of (or guide to) justified, modal beliefs. Alternatively (or additionally) one could hold that modal intuitions are an inferential guide to the modal facts. Modal intuitions are *evidence* for modal truths, in the sense that the modal truths can reliably be inferred from the fact that one has had the relevant modal intuition. Now, one would obviously want to tell some story about why this is the case. To repeat, however, if modal intuition is a guide to impossibility in *either* of our senses, then we have a way to adequately respond to the general problem caused by the second regress argument.

It is implausible, though, that modal intuition is our *only* guide to impossibility. There are impossibility claims that we can be justified in believing, but not on the basis of unaided modal intuition. If we want to avoid a local regress problem for these claims, we will need yet another guide to modal truth. For example, I am inclined to think that I could come to have a justified modal belief that my flex titanium frames are unbreakable, or that a certain Sudoku puzzle is unsolvable, without the help of unaided modal intuition. Whether or not it is possible to break the glasses (under normal temperature conditions, etc.) or solve the Sudoku puzzle is not something concerning which I am entitled to have the kind of unaided modal intuitions that I can have with respect to simple mathematical, logical, and conceptual truths.[13] Hard

---

[12]One (by no means conclusive) consideration in favor of this view is that we appear to be susceptible to *modal illusions*, and illusions are usually (if not always) linked with perceptual faculties. For instance, many people are initially inclined to spontaneously judge that, *necessarily*, for any bunch of things you have, there exists a set the members of which are just those things. It seems to me that this spontaneous inclination lingers, even after we have been taught by Russell to mistrust it.

[13]Assume that I don't have in hand a logical proof that the Sudoku puzzle is unsolvable. If I did have in hand such a proof, my justification for taking the Sudoku to be unsolvable would be based on modal intuitions. I simply want to claim that one *could* be justified in believing the puzzle to be unsolvable in the absence of such a proof.

empirical work needs to be done here. I will need to exert some effort toward breaking the frames and solving the puzzle, perhaps getting stronger, cleverer colleagues involved. Only after engaging in these good-faith efforts will I be entitled to conclude that it's not possible to break the frames and it's not possible to solve the puzzle. But after engaging in such efforts, it doesn't seem right to chalk my modal conclusions up to *modal intuitions*. They looks more like the results of *inferences to the best explanation*. The best explanation for my—and my colleagues'—repeated failed attempts at breaking the glasses is that they cannot be broken. The best explanation for my—and my colleagues'—repeated failed attempts at solving the Sudoku puzzle is that it cannot be solved.

So there is a *prima facie* case for allowing the following fourth guide to modal truth:

[IBE] Inference to the best explanation is a guide to impossibility.

Or to put it in a more catchy (but perhaps a bit misleading) slogan form:

[IBE] The right kind of non-actuality is a guide to impossibility.[14]

There are many cloudy issues surrounding *inference to the best explanation* (hereafter *IBE*). It's not perfectly clear how the structure of these arguments should be understood, or how much stock should be placed in them. Some philosophers[15] think that that IBE is a completely worthless form of argument, but this position is *wildly* at odds with common sense and practice.

I will resist delving into the many interesting philosophical issues surrounding IBE, but perhaps I should at least make a first-pass at characterizing the general structure of the argument form. Charles Sanders Peirce, who seems to have been the first philosopher to give inference to the best explanation—or, as he called it, *abduction*—serious attention, characterized it as having the following form: "The surprising fact, C is observed; But if A were true, C would a matter of course; Hence,

---

[14]Compare with NOT-A-GUIDE from above.

[15]See chapter 6 of van Fraassen 1989, for example.

there is reason to suspect that A is true."[16] This formulation might involve some unnecessary features—or exclude necessary features—of IBE, but it is a decent first shot. We often *do* reason, at least roughly, along these lines.

Consider how I could justify a belief in the modal claim that my glasses frames are unbreakable. First, suppose that I put forward the sort of good-faith effort at breaking them that we discussed earlier, and they refuse to break. This is a surprising fact! Most of my other glasses frames break very quickly, without my even *trying* to break them. If these flex titanium frames were unbreakable (in normal temperature conditions, etc.) it would be a matter of course that my frame-breaking efforts, and those of my colleagues, would constantly be frustrated. If it is not possible to break them, *of course* our efforts are going to end badly. So this gives us at least *some* (defeasible) justification for believing the modal claim that the frames are unbreakable.

Importantly, a similar story could be told concerning my justification for believing the (b) clause of some *ideal conceivability* claim. Repeated, failed, attempts at "breaking" my justification for believing a claim to be conceivable can justify me in taking the justification to be "unbreakable".[17]

3.2.4 A quick summary

To conclude, we have seen that, (i), there is no general regress problem for conceivability arguments, (ii), there is no general regress problem for impossibility claims, and (iii), as a consequence of (i) and (ii), there is no general regress problem for Chalmers-style *ideal conceivability* arguments. Along the way, we noted that there are *at least* four separate guides to modal truth:

[CONCEIVABILITY] Conceivability is a guide to possibility (and inconceivability is a guide to impossibility).

[ACTUALITY] Actuality is a perfect guide to possibility.

[MODAL INTUITION] Modal intuition is a guide to modal truth.

---

[16]See Peirce's *Collected Papers* (1960, 5.188), reprinted in Peirce 1995, p. 151.

[17]I'll return to this last point later.

[IBE] The right kind of non-actuality is a guide to impossibility.

Perhaps there are more besides, but I suspect that most other candidates could be seen as special cases of (or combinations of) the above guides.[18] But this already provides a much more realistic survey of the modal epistemic terrain than one usually gets in discussions of this topic.

3.3 The uselessness objection

David Chalmers thinks that ideal conceivability *entails* possibility. In recent years he has spent quite a bit of time and effort defending this claim. What I want to do in this section is, first, to get a little clearer on what Chalmers means by the claim, and then to consider an objection. The objection is roughly this: Whether or not ideal conceivability entails possibility, it is completely *useless* to those of us who don't happen to be ideal reasoners (which, of course, is all of us). Responding to this objection will help us get clearer on what role, if any, ideal conceivability should be playing in our modal epistemology.

3.3.1 Chalmers' entailment thesis

The first task, then, is to get clearer on what Chalmers means when he says that ideal conceivability entails possibility. We have already met two distinctions that Chalmers makes when talking about conceivability notions. First, there is the distinction between *negative* and *positive* notions of conceivability (see §3.2.1). Call this distinction *valence*. Second, there is the distinction between *prima facie* and *ideal* notions (see §3.2.2); call this *resilience*. Actually, Chalmers (2002a) has an intermediate

---

[18]For example, it might be that some of our modal knowledge is arrived at through *induction*, but induction seems (to me) to be a special case of inference to the best explanation (see Bonjour 1998, pp. 206-16). *Testimony* is also a possible way of coming to have some modal knowledge, but the justification received here will be parasitic on the testifier's justification (and it had better not be testimonies all the way down). Or, again, we might conclude that a given claim is impossible upon deriving a contradiction from it; but such an argument relies crucially on a modal intuition of the impossibility of contradictions.

option for resilience: Between *prima facie* and *ideal* conceivability, there is *secunda facie* conceivability. As we will see, this third option is—or should be—doing some important work in Chalmers' modal epistemology. For a claim S to be *secunda facie* conceivable, there must be a possible agent A who (a) *prima facie* conceives of S, and (b) has engaged in a good-faith effort at rationally defeating his or her justification for thinking that S passes the criterial tests, and has failed. Or, as Chalmers puts it, we have *secunda facie* conceivability when a "*prima facie*...conceivability judgment survives a reasonably searching process of rational reflection" (p. 160).

So what kind of conceivability is it that is supposed to entail possibility? With respect to resilience, we have already said that it is going to be a kind of *ideal* conceivability. As for valence, Chalmers thinks that there is both a positive and a negative notion of conceivability that has a decent shot at entailing (a certain kind of) possibility. Since the positive notion entails, but is not entailed by, the negative notion, Chalmers is most confident in the version of the entailment thesis that involves *positive* conceivability.

So far, then, the thesis is that *ideal positive* conceivability entails possibility. But this will not quite do, for it fails to protect against potential Kripko-Putnamian coun-terexamples. For example, it might seem that, for the chemically illiterate, it is ideally positively conceivable that water is XYZ (not H_2O). But 'water is not H_2O' does *not* express a possibility, for 'water' and 'H_2O' are *rigid designators*, picking out in *every* possible world whatever they pick out in the *actual* world. Since, in the actual world, 'water' and 'H_2O' both pick out H_2O, they do so in every possible world. Chalmers solves this problem by distinguishing between *primary* and *secondary* notions of con-ceivability and possibility. Call this distinction *semantic level*. A full understanding of semantic level would require us to delve into Chalmers' "two-dimensionalism", which is a view about how semantic contents should be assigned to sentences and thoughts. For our purposes, however, the details of this view are more distracting than illuminating. I mention it here only to counteract the appearance that Chalmers is defending a thesis that Kripke laid to rest in 1970.

Now that we have the distinctions of *valence, resilience*, and *semantic level* on the

table (or at least partially on the table), we are able to state the entailment thesis precisely:

[CP] Ideal primary positive conceivability *entails* primary possibility (and ideal primary positive inconceivability *entails* primary impossibility).[19]

Chalmers thinks that there are no compelling counterexamples to this thesis, and that there are, in fact, good positive reasons to think that it is true. But not everyone has agreed. Some have even questioned whether the thesis would be of much interest, even if it were true. Let's turn to this last worry.

3.3.2 The objection

It would be one thing if Chalmers put forward (CP) as an idle speculation about the connection between two highly abstract and theoretical notions. But Chalmers (1996, chapter 4) actually has the audacity to use (CP) as a premise in an inference the conclusion of which is as controversial as anything currently available in the philosophical marketplace:[20]

(1) Zombie worlds[21] are IDEALLY conceivable.

(CP) IDEAL conceivability entails primary possibility.

∴ (2) Zombie worlds are primarily possible.

(3) If zombie worlds are primarily possible, then physicalism is false.

∴ (NP) Physicalism is false.

---

[19]Since, in the following, it will always be *positive* and *primary* conceivability that is at issue, I will abbreviate 'ideal positive primary conceivability' as 'IDEAL conceivability', '*prima facie* positive primary conceivability' as 'PRIMA FACIE conceivability', etc.

[20]The following rendering of the argument ignores certain complications that Chalmers addresses in his official version of the argument. We needn't concern ourselves with these complications.

[21]Worlds that are (micro)physical duplicates of our world, but which lack consciousness.

According to the uselessness objection, this argument doesn't get past the first premise. Even if zombie worlds are IDEALLY conceivable, we can never *know* that they are. So who knows if the argument is sound? In any event, it is useless.

Why does the objector think that we can never know whether (1) is true? The idea is that we are not perfect reasoners, so we can never be sure that our justification for thinking that a certain claim passes the criterial tests for conceivability is undefeatable by better reasoning. Proper epistemic humility requires that we always remain open to the possibility that better reasoning would defeat our justification. This being the case, it is hard to see how ideal conceivability could ever be of any *use* to us.

3.3.3 Secunda Facie conceivability to the rescue

To respond to this, we should first admit that ideal conceivability has been allowed to hog too much of the spotlight. But, second, it should be noted that Chalmers does have at his disposal the resources that are needed to answer the objection. All he needs to do is to take some of the focus off ideal conceivability and redirect it to a notion that is already in his conceptual repertoire: SECUNDA FACIE conceivability.

SECUNDA FACIE conceivability is less epistemically problematic than IDEAL conceivability. As we saw in §3.3.1, a claim S is SECUNDA FACIE conceivable just in case:

> SECUNDA FACIE CONCEIVABILITY:
>
> (a) It is *possible* for some agent A to *prima facie* conceive of S
>
> (b) and A's justification for thinking that S passes the tests that are criterial for conceivability can survive a reasonably searching, good-faith effort aimed at defeating it.

Coming to have knowledge that a given claim is SECUNDA FACIE conceivable is no problem at all; all we need is ACTUALITY-type evidence. Through an actual episode of conceiving, I can come to know that clause (a) is satisfied. Then, by actually engaging in a good faith effort at rationally defeating my justification for thinking that the criterial tests are satisfied (perhaps getting other colleagues involved), I can come to know that clause (b) is also satisfied. *I'm* the possible agent.

Chalmers might be right that IDEAL conceivability entails primary possibility, but for the very reason that the uselessness objection points to, it is not, in general, IDEAL conceivability that gives actual reasoners their direct modal guidance. In fact, most of our intelligence about IDEAL conceivability comes directly from SECUNDA FACIE conceivings, plus IBE. Our evidence for thinking that our justification is undefeatable by better reasoning is that we've tried as hard as we can to defeat it, we've gotten other colleagues involved in the effort, and, at the end of the day, it has remained undefeated. Perhaps the best explanation for all of this is that the the justification is undefeatable.

If this is right, our objector is mistaken when she says that we non-ideal reasoners can't be justified in believing claims like 'zombie worlds are IDEALLY conceivable'. We *can* be justified in believing them—our justification comes, in the usual case, from a SECUNDA FACIE conceiving of the relevant claim plus IBE.

This, however, should not be allowed to obscure the fact that SECUNDA FACIE conceivability is also a good *direct* guide to possibility. Even if it turns out that Chalmers' entailment thesis is mistaken, we will still want to allow that SECUNDA FACIE conceivability is the source of a great deal of our modal knowledge. But, given that SECUNDA FACIE conceivings are our main evidence for IDEAL conceivability, and they also provide direct access to the modal facts, why does Chalmers frame the zombie argument in terms of IDEAL conceivability? Wouldn't the following be a much more perspicuous way of arguing?

(1′) Zombie worlds are SECUNDA FACIE conceivable.

(CP′) SECUNDA FACIE conceivability is a really good guide to primary possibility.

These premises give us a high degree of warrant for thinking that:

(2) Zombie worlds are primarily possible.

Now, this inference from (1′) and (CP′) to (2) is clearly not deductively valid. But it is also clear that the premises *do* give us a good—though defeasible—reason for believing the conclusion. All that's left now is to combine this result with

(3) If zombie worlds are primarily possible, then physicalism is false

to get the anti-physicalist conclusion:

(NP) Physicalism is false.

This second inference, from (2) and (3) to (NP), unlike the first, *is* deductively valid.

One bad reason for Chalmers to prefer his formulation to mine is that both of his sub-inferences are deductively valid. It might look like his version of the zombie argument therefore gives us a stronger reason for thinking that physicalism is false. But the apparent gain is illusory. There's a tradeoff: The main premises of Chalmers' argument, (1) and (CP), are stronger than the corresponding premises of my argument, (1′) and (CP′), and they should be assigned a lower credence. Chalmers uses these stronger premises to deductively establish the same result that I use my weaker premises to highly probabilify. So, on either approach, what we are entitled to believe at the end of the day is that physicalism is *probably* false.

It should be stressed that, on either of these approaches, SECUNDA FACIE conceivability is what's doing most of the hard work in giving us our justification for believing the conclusion of the argument. Since it makes the important role of SECUNDA FACIE conceivability transparent, I prefer the second formulation of the zombie argument to Chalmers'.

3.3.4 The role of ideal conceivability

Should we then eliminate all talk of IDEAL conceivability? Is there any work left for it to do? I think the answers are: No, we should not eliminate talk of IDEAL conceivability because, yes, there *is* work left for it to do. While SECUNDA FACIE conceivability is an *extremely* good guide to primary possibility, it is implausible that SECUNDA FACIE conceivability *entails* primary possibility; Chalmers is right to admit that, though cases of SECUNDA FACIE conceivability without primary possibility are "extremely thin on the ground" (2002a, p. 155), they probably can be found.[22]

---

[22]Chalmers suggests that, for Frege, and many others, the idea of a set of all sets might be *secunda facie* conceivable, but this notion is revealed to be self-contradictory upon ideal ("or at least Russellian") reflection.

This appears to open the door for the physicalist to say, "Sure, SECUNDA FACIE conceivability is a really good guide to possibility. But in some cases, it leads us astray in a way that we could never detect with more work and reflection, and this is what is happening in the zombie case. I'll grant you that zombie worlds are SECUNDA FACIE conceivable, and I'll grant you that no amount of extra work would help us defeat our justification for thinking that zombie scenarios meet the criterial tests for conceivability, but still, such worlds are *not* primarily possible. If we insist on relying on SECUNDA FACIE conceivability evidence here, we're going to be led into a brute modal error."

This is where IDEAL conceivability becomes important. If (CP) is right—if IDEAL conceivability *entails* primary possibility—then the threat of brute modal error (arising from conceivability evidence) is eliminated. When a claim is SECUNDA FACIE conceivable, but not possible, we could, at least in principle, bring what's conceivable back into line with what's possible through further reflection. There is better reasoning out there that would defeat our justification for thinking the claim in question passes the criterial tests for conceivability—we just need to find it. The physicalist can no longer claim that the zombie theorist is relying on a source of evidence that sometimes leads to brute errors—errors that don't arise from a misuse of the evidence in question and that can't be corrected by collecting more evidence of the same kind. If the entailment thesis is right, then if zombie worlds are not primarily possible, Chalmers' justification for thinking that zombie worlds pass the tests for conceivability is defeatable by better reasoning. The burden is now on the physicalist to produce this better reasoning. We don't get to declare ourselves the winners of an argument if all we have done is *asserted* that there is better reasoning available that would defeat our opponent's claims—we need to actually *produce* the better reasoning.

Thus, *if* Chalmers' entailment thesis is right, then brute modal error (arising from conceivability evidence) is impossible, and this will constrain the way the physicalist can respond to the zombie argument. In the terms of Chalmers 1996 and 1999, they will have to defend "Type-A", not "Type-B" materialism.[23] I should note, however,

---

[23]Type-A materialists hold that the phenomenal facts are necessitated *a priori* by the microphys-

that the 'if' that I started this paragraph with is a big 'if'. In order for Chalmers to exert this extra pressure on the physicalist, he will need to convince us that IDEAL conceivability is a *perfect* guide to primary possibility. As I have already noted, Chalmers has recently spent a great deal of time and effort defending this claim. For our purposes, however, it is enough to see the role that IDEAL conceivability will be allowed to play in our modal epistemology, if everything else pans out.

3.4   Conceivability and Apriority

3.4.1   SECUNDA FACIE conceivability and apriority

A surprising result of the picture that has been emerging here is that the fact that something is primarily positively ideally conceivable is *not* normally known *a priori*. Since Chalmers has defended the opposite view[24], let's consider whether his own views on the *a priori* can sustain the *apriority* of IDEAL conceivability. Chalmers (2004, §3.9) lists five distinguishing features of the notion of *apriority* that he feels is relevant to discussions of modal epistemology: It is (i) *token relative*, (ii) *mode-of-presentation sensitive*, (iii) *idealized*, (iv) *non-introspective*, and (v) *conclusive*. It is the last two features that cause a problem in the current context. *Non-introspectiveness* means that the relevant notion of *apriority* includes introspection under the broad heading of "experience". Thus, to be justified apart from experience means to be justified apart from introspection. *Conclusiveness* means that "*a priori* justification must meet the sort of conclusive standard associated with proof and analysis, rather than the weaker standard associated with induction and abduction."

However, on the picture that emerged in §3.3, IDEAL conceivability claims are usually inferred from the SECUNDA FACIE conceivability of the claim, plus an IBE-style inference to the best explanation. Any justification that's conferred through

---

ical facts; Type-B materialists hold that the relevant necessitation is *a posteriori*. See page 473 and 474 of Chalmers 1999.

[24] "Primary conceivability is always an a priori matter. We consider specific ways the world might be, in such a way that the true character of the actual world is irrelevant. In doing so, empirical knowledge can be suspended, and only a priori reasoning is required" (2002a, p. 158).

this kind of inference looks like it is going to be both introspective and inconclusive, and thus not *a priori* in Chalmers' sense.

First, the fact that something is SECUNDA FACIE conceivable is usually known on the basis of introspective experience. When I know, of a claim S, that it is SECUNDA FACIE conceivable, this knowledge usually comes from an ACTUALITY inference from my actually having modally imagined a scenario w which appears to be coherent, and which appears to verify S. These modal imaginings are actual mental *events*. The fact that such an event has taken place looks like the sort of thing that I could know about only through *a posteriori* introspection. How could I know that a contingent, datable, event took place, unless I, somehow or other, had the occurrence of this event reflected in my experience (broadly construed)?

Someone will object that this sort of reasoning will render even paradigmatically *a priori* claims *a posteriori*. Take a mathematical equation that I've never encountered before: $431 + 2172 = 2603$. Can I know *a priori* that this equation is true? Well, I can take the numbers 431 and 2172 and add them up in my head. But surely this "adding up in my head" is a datable, mental event. And, if what I said above is right, I can't know *a priori*, of such an event, that it has occurred; I know about it through introspection. Doesn't this mean that I can't know *a priori* that $431 + 2172 = 2603$? If so, then something must have gone wrong, for simple mathematical claims are paradigmatically *a priori*.

To respond to this, we will need to appeal to a distinction between *experiences that justify a belief that p* and *experiences that are "enabling conditions" for a belief that p*. In discussions of *apriority*, it is often noted that the standard definition of *apriority* ('knowable apart from experience') isn't meant to exclude experiences that merely enable us to *entertain* the proposition that p—it is just meant to exclude those that enter into one's *justification* for believing that p. For example, one needs to have a variety of experiences in order to develop the concepts *female*, *fox*, and *vixen*, but these experiences in no way contribute to my *justification* for believing that vixens are female foxes. They *enable* the belief, but they don't *justify* it.

Okay, but how does this help us with the mathematical case, for my mental episode

95

of computation was certainly not required for me to be able to entertain the relevant proposition. True enough, but it's plausible that *experiences required for concept possession* aren't the only kinds of enabling experiences. In the mathematical case, it looks like the computation *enables* me to have a mathematical intuition. The truth of the mathematical claim needn't be—nor is it *in fact*—*inferred* from the occurrence of a certain experience of mental computation. The mental computation simply gets me into a position where I can *see* that the equation is true. Indeed, it is epistemically fortunate for me that I do not need to infer the truth of the mathematical claim from the occurrence of the mental computation, for it is not *at all* clear how such an inference would go. One might try to run an inference to the best explanation, but this would rely on the dubious premise that the best explanation for the mental event in question is that $431 + 2172 = 2603$.

But whatever one thinks of this "enabling conditions" move in the mathematical case, it is pretty clear that it's not going to work in the present case. My introspective judgment that a certain mental event of conceiving has taken place is *precisely* what justifies me in believing that such a conceiving is possible. In the mathematical case, the occurrence of the mental event of computation didn't *entail* that $431 + 2172 = 2603$ (which isn't about mental events). Perhaps it could serve as *evidence* for the mathematical claim, but the evidential connection would have to be indirect. But, in the present case, the occurrence of the mental event of conceiving straightforwardly *entails* that such a conceiving is possible (for actually p entails possibly p). It is this inferential connection that provides my justification for taking S to be conceivable. So the introspected mental episodes are playing the *justificatory* role for my belief, and the resulting belief is thus known only *a posteriori*.

Indeed, this isn't the only piece of *a posteriori* data that is usually impressed into service in order to justify a SECUNDA FACIE conceivability claim. An introspected event of modal imagination, by itself, only justifies me in taking S to be PRIMA FACIE conceivable. Then comes the hard mental work. I must subject this *prima facie* conceiving to careful scrutiny, looking for subtle, hidden incoherence, and confirming that the imagined scenario, if coherent, *does*, in fact, verify the claim in question (as

opposed to some closely related claim[25]). If my attempts at defeating my *prima facie* justification continually fail, then the claim will have earned the title 'SECUNDA FACIE conceivable'. The thing to note here is that this trying and failing consists in a series of mental events. My knowledge of such events is, again, attained only *a posteriori*.

Thus, in the usual case, both clauses of a SECUNDA FACIE conceivability claim are justified *a posteriori*, for their justification comes through introspected experience. So the first problem is that IDEAL conceivability claims are inferred from SECUNDA FACIE conceivability claims, and our justification for believing these SECUNDA FACIE conceivability claims lacks feature (iv) of Chalmers-*apriority*.

The second problem has to do with the nature of the inference from SECUNDA FACIE conceivability to IDEAL conceivability. The inference is *abductive*—it is an inference to the best explanation. Thus, the justification conferred cannot be *conclusive* in Chalmers' sense. So, in addition to lacking feature (iv), our justification for believing IDEAL conceivability claims also lacks feature (v). Such claims are not, then—at least in the usual case—known *a priori*.

This is a unexpected result. Chalmers led us to believe that IDEAL conceivability gives us *a priori* access to the modal facts. Now it is coming out that our knowledge of IDEAL conceivability is almost always justified *a posteriori*! What gives?

First, notice that, throughout my argument so far, I have repeated phrases like 'in the usual case', 'almost always', 'hardly ever', etc. But Chalmers can happily admit that our justification for believing IDEAL conceivability claims is almost always *a posteriori*. All he needs is that it is *possible* to get the justification *a priori*. It is no argument against a claim's being *a priori* knowable that it is not usually—or even *is never*—known *a priori*. Take some absurdly complicated mathematical problem. Due to my (formidable) memory and computational limitations, I am unlikely to know the result *a priori*. I'm probably going to need to punch the relevant numbers into a calculator, or write a computer program to generate the answer, or ask my wife, etc., all of which are ways of coming to know the result *a posteriori*. And perhaps all other actually existing beings share in my limitations enough so that they too would

---

[25]See Yablo 2000 and Chalmers 1996.

need to resort to *a posteriori* methods to answer the problem. None of that entails that the solution to the problem is know*able* only *a posteriori*. It might very well be knowable *a priori* for a being blessed with a larger memory bank and a more powerful cognitive processor. And *a priori* knowability is usually thought of as an *idealized* notion—a claim is *a priori* knowable iff it is known *a priori* by any possible being.[26]

So, the question is: Is it *possible* to know SECUNDA FACIE conceivability claims *a priori*? Is it possible to know, for some claim S, that there is a possible agent A who *prima facie* primarily positively conceives of S, and who has engaged in a good-faith, "reasonably searching", effort at rationally defeating his or her justification for thinking that S passes the tests that are criterial for conceivability, and has failed? Let's consult our modal guides.

Could I come to know *a priori* that a claim is SECUNDA FACIE conceivable with CONCEIVABILITY? Well, perhaps it is conceivable that there is an agent who *prima facie* conceives of S and who has engaged in, and failed in, a good-faith effort to rationally defeat her justification for thinking S passes the criterial tests. But, what is my justification for believing *this* modal claim? If we are going to avoid the kind of regress we discussed in §3.2, it can't be CONCEIVABILITY-type reasons at every level. One of the other modal guides will have to be involved at some point. What about ACTUALITY? For reasons that we have already seen, this can give us only *a posteriori* justification for accepting a SECUNDA FACIE conceivability claim. And we have also seen that IBE is never an *a priori* matter, according to Chalmers' definition of *apriority*.

We are left, then, with MODAL INTUITION. If SECUNDA FACIE conceivability claims are going to be knowable *a priori*, it will have to be because it is possible to have a modal intuition that such an agent is possible.[27] Perhaps we *could* have this kind of modal intuition. Without knowing more about where modal intuitions come from—and what governs their generation—it is hard to rule this out. It *might* be

---

[26]There are problems with how far we should be allowed to idealize. Let's ignore them.

[27]Assuming, that is, that our four modal guides exhaust the methods for gaining modal knowledge. I think they *do* have a good shot at doing this, but I certainly haven't given any argument to that effect. Even if there are other guides, however, it seems quite unlikely that they'll give us *a priori* access to modal truth.

that one can have a modal intuition that there is a possible agent A who *prima facie* conceives of S and who has engaged in a good-faith effort at rationally defeating her justification for thinking that S passes the criterial tests for conceivability. But this would be a weird intuition indeed! I can't say that I have ever enjoyed a modal seeming of this kind. But, since all Chalmers needs is that SECUNDA FACIE conceivability claims are *a priori* know*able* (not that they are usually, or even *ever*, known *a priori*), there is at least *hope* that some SECUNDA FACIE conceivability claims are knowable *a priori*.

3.4.2 IDEAL conceivability and apriority

But even if this hope is fulfilled (which I doubt it will be), the *a priori* knowledge so gained cannot help us come to know with *a priori* justification that the corresponding IDEAL CONCEIVABILITY claim is also satisfied.

Remember what the IDEAL conceivability of S involves:

(a) It is *possible* to *prima facie* conceive of S

(b) in a way that is *not possible* to undermine with better reasoning.

Let's concede, for the sake of argument, that the corresponding SECUNDA FACIE conceivability claim is knowable *a priori*, so we can help ourselves to the *apriority* of (a). Still, that leaves the hardest part yet to be done. How could we know *a priori* that it is impossible to undermine A's conceiving of S with better reasoning? We have already seen that the usual route from SECUNDA FACIE conceivability to IDEAL CONCEIVABILITY is via an inference to the best explanation, but this, by definition, is *not* Chalmers-*a priori*. And, as we saw in §3.2, CONCEIVABILITY is only going to help us if we have an independent grip on what's impossible to conceive. Finally, ACTUALITY considerations can only help establish impossibility claims when combined with a Chalmers-*a posteriori* IBE. We are left, once again, with MODAL INTUITION. Could such intuitions give us *a priori* justification for believing that A's justification is undefeatable by better reasoning?

99

Could we have a modal intuition that there is a possible agent who *prima facie* conceives of S (for some S) and whose justification for thinking that S passes the criterial tests is undefeatable by better reasoning? I think the answer is 'no'. Perhaps we sometimes enjoy modal intuitions to the effect that a certain bit of justification is undefeatable by better reasoning. But this kind of intuition is always directed at a *particular* bit of justification. If there's not a particular bit of justification in question, I cannot have an intuition that *it* is undefeatable by better reasoning. Such intuitions, if reliable, have got to be about something in particular.

Recall that, if my justification for believing the whole IDEAL conceivability claim is going to be *a priori*, my justification for thinking that there *is* this possible agent A who *prima facie* conceives of S must also be based on a one-off, unaided, modal intuition. It cannot be based on the fact that I take myself to be an agent who has *prima facie* conceived of S (this would be the usual, *a posteriori*, way of justifying the relevant conceivability claim). Given the method I must use to come to this knowledge, it doesn't look like I can now have additional intuitions about A's justification, and whether or not it is defeatable. I know that A must have modally imagined a scenario that she takes to verify S, but I shouldn't have any idea what this scenario is. And without having any idea what the relevant scenario is, it doesn't seem that I can have any intuitions about whether A's justification is undefeatable by better reasoning. I just don't know enough about A's justification to be able to narrow down modal space in this way.

If I *do* know enough about A's justification to have these additional modal intuitions about it, then it looks like I have cheated. To the extent that I know why this possible agent A thinks S passes the tests that are criterial for conceivability, I have the same reasons for thinking that S passes the tests. Imagining what A would have to modally imagine in order to have justification for thinking that S is positively conceivable itself involves engaging in a modal imagining. My conclusion that there is a possible agent A who conceives of S now looks like it is justified by an actual event of modal imagining, not a one-off modal intuition.

If I am careful not to have any idea what S's justification is, it looks like I shouldn't

have any modal intuitions about whether or not this justification is defeatable by better reasoning. If I find myself with such intuitions, I should ignore them—they don't seem like they could possibly be reliable.

The upshot is this: If the (a) clause of an IDEAL conceivability claim is justified by a one-off modal intuition, then it doesn't appear that we could ever be in a position to have a reliable one-off modal intuition to justify the (b) clause as well. One clause or the other is going to need to be justified by something other than unaided modal intuition. And, we've already argued that the other options available for modal guidance are only going to give us *a posteriori* justification. We can conclude, then, that our knowledge of IDEAL conceivability is always *a posteriori*—and that there is no way of getting it *a priori*.

3.5 Salvaging Modal Rationalism

The foregoing would certainly seem to be destructive to *modal rationalism*, which holds that IDEAL conceivability gives us *a priori* access to modal truth.[28] Modal rationalism, at least for our purposes, can be understood as consisting of two claims:[29]

> MODAL RATIONALISM:
>
> (CA) IDEAL conceivability is knowable *a priori*.
>
> (CP) IDEAL conceivability entails primary possibility.

Combining these two claims gives us something that deserves the title 'modal rationalism'. If IDEAL conceivability is knowable *a priori* and it entails primary possibility,

---

[28]Although Chalmers doesn't explicitly include the idea that IDEAL conceivability gives us *a priori* access to modal truth in any of his official definitions of modal rationalism (see the next footnote), the claim is clearly essential to his project. See Chalmers 2002a, where Chalmers sets out to defend "a long tradition in philosophy of using a priori methods to draw conclusions about what is possible and what is necessary" (p. 145).

[29]Chalmers (2002a) discusses three modal rationalist theses: (1) *weak modal rationalism* (ideal primary positive conceivability entails primary possibility), *strong modal rationalism* (ideal primary *negative* conceivability entails primary possibility), and *pure modal rationalism* (ideal primary positive conceivability ≡ ideal primary negative conceivability ≡ primary possibility). My thesis (CP) is just Chalmers' "weak modal rationalism". I prefer not to call this a form of modal rationalism until it is explicitly combined with (CA), which says that IDEAL conceivability is an *a priori* affair.

then primary possibility is sometimes knowable *a priori*.[30]

If the argument of §3.4.2 is right, then (CA) is false, and thus MODAL RATIO-
NALISM is false. Even Chalmers' highly constrained notion of *conceivability* fails to
give us *a priori* access to modal truth. This might seem to cause a serious problem
for Chalmers' zombie argument. If we can't know *a priori* that zombie worlds are
IDEALLY conceivable, then we can't know *a priori* that they are primarily possible, so
we can't know *a priori* that physicalism is false. This seems to leave open the possi-
bility that an empirical, scientific, investigation might reveal that there is something
wrong with zombie worlds—we thought we could conceive of them, and we couldn't
have known *a priori* that we were wrong, but it turns out that we *were* wrong, as our
empirical investigation has shown.

This is not the right lesson to have taken home. It is true that IDEAL conceiv-
ability claims are only knowable *a posteriori*, but it is also true that the kind of *a
posteriori* evidence that enters into the picture is of a very special kind; it is the kind
of evidence which can, in principle, be gathered from the comfort of the armchair.
We might say that, although IDEAL conceivability claims are not *a priori*, they are
at least *armchairori*, where S is armchairori iff S is *knowable from the armchair*,
independently of any outside empirical investigation.

Now, this "definition" of *armchairority* is *horribly* unclear. What is meant by
these phrases 'knowable from the armchair' and 'independently of outside empirical
investigation'? On one reading of 'knowable from the armchair', *everything* is going
to qualify. I can read physics textbooks, talk to experts, look through microscopes,
orchestrate grand experiments, etc., all while ensconced comfortably in my armchair.
Indeed, I could, at least in principle, talk to God from my armchair, and he could
just *tell* me anything I want to know. These are the sorts of doxastic practices that
are supposed to be excluded by the 'independently of outside empirical investigation'
clause. Perhaps it is best to imagine that my armchair is located in room that is

---

[30]Notice the 'sometimes'. (CA) and (CP) do not *insure* that primary possibility is always an *a
priori* affair. It might turn out that a very small fraction of the primarily possible claims are IDEALLY
conceivable. To insure that primary possibility is always *a priori* accessible, we would need to add
the converse of (CP) to our theses.

pitch black, silent, odorless, and so on. The point here is to exclude what we would intuitively think of as experiences of the "outside" world, without also excluding experiences of our own "interior" mental worlds. Traditional notions of *apriority* fail to distinguish between the two, since both are types of *experience*.[31]

None of this makes this notion of armchairori knowledge perfectly clear, but that is okay. All that's important is to notice that there is conceptual room for a notion that is close to *apriority*, but which operates with a more narrow conception of what counts as an experience—a conception which only encompasses what we would intuitively consider "outer" sensory experiences.

The point is this: Perhaps SECUNDA FACIE and IDEAL conceivability are not the sorts of things that we can discern *a priori*, but, so far, I have given no reason for thinking that they are not discernible armchairori. I cannot know *a priori* that certain datable mental events have taken place, but these are also not the sorts of things that I need to conduct an outer, empirical, investigation to uncover. I can have knowledge of them "from the armchair". So, even if (CA) is untenable, perhaps the parallel thesis involving armchairori knowledge *is* tenable. This, I think, would be an interesting avenue of investigation. And, combined with (CP), it certainly seems to capture the spirit, if not the letter, of modal rationalism.

---

[31]But see Boghossian (1996, f.n. 4), who says that he "[has] always found it natural to regard *a priori* knowledge as encompassing both knowledge that is based on no experience as well as knowledge that is based purely on *inner* experience " I think that this is probably a highly idiosyncratic use of the term. Our knowledge of tickles, headaches, and blind rage is usually thought of as *a posteriori*. Incidently, Chalmers has some good reasons for resisting this broader notion of *apriority*. Primary intensions, for Chalmers, represent the "*a priori* aspect" of meaning, and sentences with necessary primary intensions are supposed to represent conceptual truths. However, if we read Chalmers' definition of *primary intension* as invoking Boghossian-*apriority*, it turns out that 'I have a headache now' has a necessary primary intension (given that I now have a headache). But it is certainly not a *conceptual* truth—of any kind—that I have a headache. A similar point is made in Yablo 2002, p. 459.

Bibliography

Bailey, Andrew. m.s. "The Unsoundness of Arguments From Conceivability (draft)." Found on his website (http://www.uoguelph.ca/~abailey/papers/index2.html).

Bealer, George. 2002. "Modal Epistemology and the Rationalist Renaissance." In *Conceivability and Possibility*, edited by Gendler and Hawthorne. Oxford: Oxford University Press.

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Boghossian, Paul. 1996. "Analyticity Reconsidered." *Noûs* 30:360–392.

Bonjour, Laurence. 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Brueckner, Anthony. 1986. "Brains in a Vat." *The Journal of Philosophy* 83:148–167.

Burge, Tyler. 1998a. "Individualism and Self-Knowledge." In *Externalism and Self-Knowledge*, edited by Peter Ludlow and Norah Martin, 111–128. CSLI.

— — —. 1998b. "Individualism and the Mental." In *Externalism and Self-Knowledge*, edited by Peter Ludlow and Norah Martin, 21–83. CSLI.

Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.

— — —. 1999. "Materialism and the Metaphysics of Modality." *Philosophy and Phenomenological Research* 59:473–93.

— — —. 2002a. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, edited by Gendler and Hawthorne, 145–200. Oxford: Oxford University Press.

————. 2002b. "On Sense and Intension." *Philosophical Perspectives* 16: Language and Mind:135–82.

————. (forthcoming) 2004. "The Foundations of Two–Dimensional Semantics." In *Two–Dimensional Semantics: Foundations and Applications*, edited by M. Garcia-Carpintero and J. Macia. New York: Oxford University Press.

Ebbs, Gary. 1992. "Skepticism, Objectivity, and Brains in Vats." *Pacific Philosophical Quarterly* 73:239–266.

Edgington, Dorothy. 1986. "Do Conditionals Have Truth–Conditions?" *Critica* XVIII (52): 3–20.

Evans, Gareth. 1985. *Varieties of Reference*. Oxford University Press.

Fraassen, Bas C. Van. 1989. *Laws and Symmetry*. Oxford: Clarendon Press.

Gibbons, John. 1996. "Externalism and Knowledge of Content." *The Philosophical Review* 105 (3): 287–310 (Jul.).

Grice, Paul. 1989. Chapter 2 of *Studies in the Ways of Words*, 24–42. Harvard University Press.

Jackson, Frank. 1987. *Conditionals*. Oxford: Basil Blackwell.

————. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford.

Jeshion, Robin. 2002. "Acquaintanceless De Re Belief." In *Meaning and Truth: Essays in Philosophical Semantics*, edited by ORourke Campbell and Shier., 53–78. New York: Seven Bridges Press.

Kant, Immanuel. 1990. *The Critique of Pure Reason*. Amherst, NY: J. M. D. Meiklejohn, trans, Prometheus Books.

Kaplan, David. 1991. "Demonstratives." In *Themes From Kaplan*, edited by J. Almog et al., 481–563. Oxford: Oxford.

Kripke, Saul. 1972. *Naming and Necessity*. Cambridge: Harvard University Press.

Lycan, William. 2001. *Real Conditionals*. Oxford: Oxford University Press.

Peirce, Charles S. 1960. "Lectures on Pragmatism." In *Collected Papers of Charles Sanders Peirce*, edited by Hartshorne and Weiss, Volume V, Second Printing, 13–131. Cambridge, MA: The Belknap Press.

————. 1995. "Abduction amd Induction." In *Philosophical Writings of Peirce*, edited by Justus Buchler, 150–156. New York: Dover.

Pryor, James. 2000. "The Skeptic and the Dogmatist." *Noûs* 34 (4): 517–549.

————. 2004, July. "Indexicality and Aprioricity." (unpublished, can be found on his website).

Putnam, Hilary. 1981. *Reason, Truth, and History*. New York: Cambridge University Press.

————. 1996. "The Meaning of 'Meaning'." In *Readings in Language and Mind*, edited by Heimir Geirsson and Michael Losonsky, 157–198. Blackwell.

Quine, W.V.O. 1966. "Three Grades of Modal Involvement." In *The Ways of Paradox and Other Essays*, 158–176. Cambridge: Harvard University Press.

Russell, Betrand. 1950. *An Inquirty into Meaning and Truth*. Allen and Unwin.

Simchen, Ori. 2003. "Meaningfulness and Contingent Analyticity." *Nous* 37:278–302.

Soames, Scott. 2003. *Philosophical Analysis in the Twentieth Century*. Volume 2: The Age of Meaning. Princeton: Princeton University Press.

————. 2005. *Reference and Description*. Princeton: Princeton University Press.

Stalnaker, Robert. 1975. "A Theory of Conditionals (1968)." In *Causation and Conditionals*, 165–179. London: Oxford University Press.

Weatherson, Brian. 2001. "Indicative and Subjunctive Conditionals." *The Philosophical Quarterly* 51 (203): 200–216 (April).

Wright, Crispin. 1992. "On Putnam's Proof that We Are Not Brains-in-a-Vat." *Proceedings of the Aristotelian Society* 92:67–94.

Yablo, Stephen. 1993. "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53:1–42.

————. 2000. "Textbook Kripkeanism and the Open Texture of Concepts." *Pacific Philosophical Quarterly* 81:98–122.

————. 2002. "Coulda, Woulda, Shoulda." In *Conceivability and Possibility*, edited by Gendler and Hawthorne, 441–492. Oxford: Oxford University Press.