Does Conceivability Entail Possibility?

David J. Chalmers

**Department of Philosophy
University of Arizona
Tucson, AZ 85721.**

chalmers@arizona.edu

[[Published in (T. Gendler & J. Hawthorne, eds) *Conceivability and Possibility* (Oxford University Press, 2002), pp.145-200.]]

There is a long tradition in philosophy of using a priori methods to draw conclusions about what is possible and what is necessary, and often in turn to draw conclusions about matters of substantive metaphysics. Arguments like this typically have three steps: first an epistemic claim (about what can be known or conceived), from there to a modal claim (about what is possible or necessary), and from there to a metaphysical claim (about the nature of things in the world).

We find this structure in many different areas of philosophy: in arguments about whether the mental is reducible to the physical (or vice versa), about whether causation and laws are reducible to regularities in nature, about whether knowledge is identical to justified true belief, and so on. Many arguments in these domains first seek to establish an epistemic gap between two phenomena (e.g. that we can know or conceive of one without the other), argue from there to a modal gap (e.g. that it is possible that one could exist without the other), and step from there to a metaphysical gap (e.g. that one is not reducible to the other).

Here, I will mostly be concerned with the second step: the bridge between the epistemic and modal domains. The most popular bridge here is the method of conceivability. One argues that some state of affairs is conceivable, and from there one concludes that this state of affairs is possible. Here, the kind of possibility at issue is *metaphysical* possibility, as opposed to physical possibility, natural possibility, and other sorts of possibility. Metaphysical conclusions turn most directly

on matters of metaphysical possibility: if one domain is reducible to another, the facts about the second should metaphysically necessitate the facts about the first. So it is metaphysical possibility that is relevant in the three-step argument above. And there is at least some plausibility in the idea that conceivability can act as a guide to metaphysical possibility. By contrast, it is very implausible that conceivability entails physical or natural possibility.

For example, it seems conceivable that an object could travel faster than a billion meters per second. This hypothesis is physically and naturally impossible, because it contradicts the laws of physics and the laws of nature. This case may be metaphysically possible, however, since there might well be metaphysically possible worlds with different laws. If we invoke an intuitive conception of a metaphysically possible world as a world that God might have created, if he had so chosen: it seems that God could have created a world in which an object traveled faster than a billion meters per second. So in this case, although conceivability does not mirror natural possibility, it may well mirror metaphysical possibility.

In recent years, conceivability arguments have faced considerable opposition. Many philosophers hold that the step from conceivability to metaphysical possibility has been shown to be invalid, not least due to a number of apparent counterexamples. For example, it is often suggested that complex mathematical falsehoods (such as Goldbach's conjecture or its negation) are conceivable but impossible. It is also widely believed that a posteriori identities provide counterexamples: on this view, it is conceivable but not possible that Hesperus is not Phosphorus, and that water is not H_2O.

To properly assess this matter, we first have to clarify what is meant by 'conceivability'. This term can be understood in many different ways. In some senses of the term, an entailment from conceivability to possibility is out of the question; in other senses, things are not so clear. Here I will isolate three dimensions of difference between notions of conceivability: prima facie vs. ideal conceivability, positive vs. negative conceivability, and primary vs. secondary conceivability. These distinctions are largely independent of each other, so there may be up to eight sorts of conceivability in the vicinity: prima facie

primary positive conceivability, and so on. By making these distinctions, I think at least one plausible and defensible conceivability-possibility thesis can be formulated, free of any clear counterexamples.

As I will be using the term here, conceivability is a property of statements, and the conceivability of a statement is in many cases relative to a speaker or thinker. I think that conceivability is more deeply a property of propositions, but I will not talk that way here, since many philosophers have theoretical views about propositions that can confuse these issues. For a statement S, we will have eight or so ways of disambiguating the claim that S is conceivable for a given subject. I will first give rough characterizations of the various dimensions of difference. Then I will examine various specific notions of conceivability that result, and address the question of the extent to which these notions of conceivability support an entailment from the conceivability of S to the possibility of S. For ease of discussion, I will use sentence symbols such as 'S' loosely, allowing context to disambiguate whether the corresponding sentence is being used or mentioned.

1 Prima Facie Vs. Ideal Conceivability

S is prima facie conceivable for a subject when S is conceivable for that subject on first appearances. That is, after some consideration the subject finds that S passes the tests that are criterial for conceivability. The specific criteria will depend on a substantive notion of conceivability, as outlined in the discussion of the remaining dimensions of conceivability, to remove the apparent circularity. For example, one substantive notion of conceivability (a version of negative conceivability) holds that S is conceivable if no contradiction is detectable in the hypothesis expressed by S. Under this notion, S will be prima facie conceivable for a subject when that subject cannot (after consideration) detect any contradiction in the hypothesis expressed by S.

S is ideally conceivable when S is conceivable on ideal rational reflection. It sometimes happens that S is prima facie conceivable to a subject, but that this prima facie conceivability is undermined by further reflection showing that the tests that are criterial for

conceivability are not in fact passed. In this case, S is not ideally conceivable. Given the substantive notion of (negative) conceivability above, for example, S will be ideally conceivable when ideal rational reflection detects no contradiction in the hypothesis expressed by S; or equivalently, when ~S is not a priori.

An example is provided by any mathematical statement M whose truth-value is currently unknown, but which will later be proved to be true. Here ~M is prima facie conceivable in the sense above (i.e. prima facie negatively conceivable) at least for current subjects, but it is not ideally conceivable, as ideal reflection will rule out ~M a priori.

The notion of ideal rational reflection remains to be clarified. One could try to define ideal conceivability in terms of the capacities of an ideal reasoner — a reasoner free of all contingent cognitive limitations. Using this notion, we could say that S is ideally conceivable if an ideal reasoner would find it to pass the relevant tests (if an ideal reasoner could not rule out the hypothesis expressed by S a priori, for example). A strategy like this is taken by Menzies (1998). One trouble is that it is not obvious that an ideal reasoner is possible or coherent. For example, it may be that for every possible reasoner, there is a more sophisticated possible reasoner. Alternatively, one can dispense with the notion of an ideal reasoner, and simply invoke the notion of undefeatability be better reasoning. Given this notion, we 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. For ideal conceivability, one needs justification that cannot be rationally defeated.

I will not try to give a substantive characterization of what good reasoning consists in, or of what counts as a cognitive limitation to be idealized away from. I suspect that any such attempt would end up being open-ended and incomplete. In general, my approach in this paper is to take certain rational notions as primitive, and to what sort of connection to modal notions emerges. In this case, I am simply appealing to our intuitive grasp of notions of reasoning, and of when

one reasoning process defeats another. I note that the notion of undefeatability here invoked here is also implicit in our concept of knowledge: it is generally held that if one's justification for a belief that P is defeatable by better reasoning, then one does not know that P. So the notion of conceivability is not obviously worse off than the concept of knowledge.

There is also a fairly direct parallel between the idealization present in the notion of ideal conceivability and that present in the familiar notion of apriority. If I cannot know that P independent of experience, but another less limited being could do so, then it is a priori that P. And if I believe that P, but the justification for my belief is defeatable by better reasoning, then it is not a priori that P (unless there is another undefeatable justification). So the notion of apriority idealizes away from cognitive limitations in much the same way as the notion of ideal conceivability. This is not to say that either of these idealizations are perfectly clear, but at least the idealization is a familiar one. And in practice, the idealizations are easy to apply. We will see that there are certain difficult cases on the far end of idealization where things get tricky; but dealing with such cases may allow us to to further clarify the idealization.

There are a couple of things that should be clarified in advance, however. First, it is important that "better reasoning" about conceivability not be defined even in part as reasoning that better tracks possibility. Such a criterion would trivialize the link between ideal conceivability and possibility. Fortunately there is no reason to expect that such a criterion will come into play, at least on most of the substantive notions of conceivability we will be considering. Where conceivability is defined in terms of what is ruled out a priori, for example, we have an entirely independent grounding for the notion. Only if conceivability is directly defined in terms of possibility — perhaps as what a subject judges to be possible — will there be a danger of triviality.

Second, in most cases (with an exception to be discussed later), the reasoning in question is restricted to a priori reasoning, and the further reasoning involved in the idealization will remain within the a priori domain. Sometimes this will be an automatic consequence of a

given notion of conceivability (e.g. the negative notion of conceivability above), and sometimes it can be seen as a stipulation. Either way, this restriction is important if this issue is to shed light on the issue of a priori access to modality.

2 Positive vs. Negative Conceivability

Negative notions of conceivability hold that S is conceivable when S is not *ruled out*. For example, a popular sense of "conceivable" in common usage holds roughly that S is conceivable when it is not ruled out by what one knows, or by what one believes. I will set this popular usage aside as tangential to our main purposes here: philosophers are usually concerned with senses in which S can be conceivable even when one knows that S is not actually the case. More relevant notions of negative conceivability can be obtained by constraining the ways in which S might be ruled out.

The central sort of negative conceivability holds that S is negatively conceivable when S is not ruled out a priori, or when there is no (apparent) contradiction in S. One can disambiguate the notion depending by applying the distinction between prima facie and ideal conceivability, as above. We can say that S is prima facie negatively conceivable for a subject when that subject, after consideration, cannot rule out S on a priori grounds. And we can say that S is ideally negatively conceivable when it is not a priori that ~S.

One subtlety concerns cases of indeterminacy. For some S (perhaps statements that are not truth-evaluable, or some statements involving vague predicates), it may be a priori that it is indeterminate whether S. If so, it is not a priori that ~S. In such a case, is S negatively conceivable? For various reasons, it seems best to say that it is not: in these cases, the possibility that S is not truly left open. To handle such cases, one can say that S is negatively conceivable when det(S) cannot be ruled out, and that S is ideally negatively conceivable when it is not a priori that ~det(S). Here "det S" expresses the claim that S is determinately the case, and "~det(S)" expresses the claim that S is false or indeterminate. (In other frameworks for dealing with indeterminacy, one can adopt a corresponding definition.) In the case of a priori indeterminacy above, it will be a priori that ~det(S), so S

will not be ideally negatively conceivable.

Positive notions of conceivability require that one can form some sort of positive conception of a situation in which S is the case. One can place the varieties of positive conceivability under the broad rubric of *imagination*: to positively conceive of a situation is to in some sense imagine a specific configuration of objects and properties. It is common to imagine situations in considerable detail, and this imagination is often accompanied by interpretation and reasoning. When one imagines a situation and reasons about it, the object of one's imagination is often revealed as a situation in which S is this case, for some S. When this is the case, we can say that the imagined situation *verifies* S, and that one has *imagined that* S. Overall, we can say that S is positively conceivable when one can imagine that S: that is, when one can imagine a situation that verifies S. (This definition, and the following discussion, is indebted to the discussion of conceivability in Yablo 1993.)

Different notions of conceivability correspond to different notions of imagination. One such notion is tied to *perceptual imagination*. A subject perceptually imagines that S when the subject has a perceptual mental image that represents S as being the case. This happens roughly when the image relevantly resembles a perceptual experience that represents S as being the case (see Gendler/Hawthorne, this volume[?]). For example, one can perceptually imagine that a pig flies by forming a visual image of a flying pig, where this can be understood as an image that relevantly resembles a visual experience as of a flying pig.

Perceptually imagining that P differs from supposing that P, or from entertaining the proposition that P, in that it involves not just an attitude toward P, but toward some specific situation that stands in a certain relationship to P. To perceptually imagine that pigs fly, we form a mental image that represents a specific situation (one with a certain configuration of animals), and we take this to be a situation in which pigs fly. Here, we can say that the imagined situation verifies "pigs fly". More generally, one can say that when one perceptually imagine that P, one perceptually imagines a situation that verifies P. Unlike entertaining or supposing that P, the phenomenology of perceptually

imagining that P has a mediated objectual character, with an attitude toward an intermediate mental object (here, an imagined situation) playing a crucial role. This objectual character (noted by Yablo 1993) is distinctive of positive conceivability.

This objectual character is present in cases of imagination that are no grounded in imagery. There is a sense in which we can imagine situations that do not seem to be potential contents of perceptual experiences. One can imagine situations beyond the scale of perception: e.g. molecules of H_2O, or Germany winning the Second World War. One can imagine situations that are unperceivable in principle: e.g. the existence of an invisible being that leaves no trace on perception. And one can imagine pairs of situations that are perceptually indistinguishable: e.g. the situations postulated by two scientific hypotheses that make the same empirical predictions, or arguably the existence of a conscious being and its zombie twin (an unconscious physically identical duplicate).

In these cases, we do not form a perceptual image that represents S. Nevertheless, we do more than merely suppose that S, or entertain the hypothesis that S. Our relation to S has a mediated objectual character that is analogous to that found in the case of perceptual imaginability. In this case, we have an intuition of (or as of) a *world* in which S, or at least of (or as of) a situation in which S, where a situation is (roughly) a configuration of objects and properties within a world. We might say that in these cases, one can *modally imagine* that P. One modally imagines that P if one modally imagines a world that verifies P, or a situation that verifies P. Modal imagination goes beyond perceptual imagination, for the reasons above, but it shares with perceptual imagination its mediated objectual character.

"Modal imagination" is used here as a label for a certain sort of familiar mental act, and like other such categories, it resists straightforward definition. But its phenomenology is familiar. One has a positive intuition of a certain configuration within a world, and takes that configuration to satisfy a certain description. When one modally imagines H_2O molecules, for example, one imagines a configuration of particles. To modally imagine Germany winning the Second World War, one might imagine a world in which certain German armies win certain

battles and go on to overwhelm Allied forces within Europe. When one reflects on these imagined (parts of) worlds, they reveal themselves as (parts of) worlds in which there are H_2O molecules, or in which Germany won the Second World War.

Just as modally imagining that S goes beyond entertaining the proposition that S, modally imagining a world that verifies S goes beyond entertaining a proposition (even a highly specific proposition) that implies S. If this were all there were to modal imagination, then we could modally imagine any proposition trivially: just take the proposition itself, and conjoin with further propositions if necessary. But there are many propositions that we cannot easily modally imagine: complex unknown mathematical propositions M, for example. In these cases, we have no intuition of a world verifying M, even though we can entertain many specific propositions that imply M. So imagining a world is not merely entertaining a description. Of course it may be that imagining a world involves standing in *some* relation to a detailed description of that world (one presumably uses one's conceptual resources to imagine a world), but if so, this relationship goes beyond mere entertaining or supposing. Rather, it is a relation that is distinctive of modal imagination.

We can say that an imagined situation verifies S when reflection on the situation reveals it as a situation in which S. Understood this way, verification is a broadly epistemic relation, tied to certain rational processes. Importantly, verification is stronger than a mere evidential relation. We have seen that one can imagine situations in which no perceptual evidence is involved, as with the case of the nuclear force above. One can also imagine a situation in which one has strong evidence that S, such that the imagined situation is nevertheless epistemically compatible with ~S: a situation where experimental results point to a certain sort of particle behavior, for example, or where usually reliable witnesses testify that someone committed a crime. In such cases, consideration of the imagined situation alone does not reveal that it as a situation in which S (as opposed to a situation in which there is strong evidence for S), so the imagined situations do not verify S. In this respect, verification of a statement by an imagined situation is broadly analogous to an entailment of one statement by another (a priori entailment, in the central cases): if it is

coherent to suppose that the situation obtains without S being the case, then the situation does not verify S.

Just as imagining a unicorn does not entail the existence of the imagined unicorn, imagining a situation does not entail the existence of the imagined situation, and imagining a world does not entail the existence of the imagined world. Nothing here entails that one should be ontologically committed to situations or worlds at all. Rather, for our purposes these can be regarded as mere intentional objects, useful in characterizing the cognitive or phenomenological structure of modal imagination. It should also be noted that nothing here presupposes that when one imagines a situation or a world, there is a metaphysically possible situation or world that corresponds to the object of one's imagination. Again, these can simply be seen as *apparent* situations or worlds, of the sort represented in an act of imagination. For all that has been said so far, the imagination of situations and worlds may greatly outstrip the bounds of metaphysical possibility.

Indeed, it is arguable that one can modally imagine S when S involves an a priori contradiction. An example may be a case in which one imagines a geometric object with contradictory properties. In cases like this, one imagines a situation in something less than full detail. Another example may be a case when one imagines that a true mathematical claim (Goldbach's conjecture, perhaps) is false, by imagining a situation in which experts announce it to be false. In this sort of case, one might misinterpret the imagined situation as a situation in which S; here, the situation is merely one in which one has evidence for S.

To avoid cases like these, one can isolate a notion of *coherent modal imagination*. In this sense, S is positively conceivable when one can *coherently* modally imagine a situation that verifies S. A situation is coherently imagined when it is possible to fill in arbitrary details in the imagined situation such that no contradiction reveals itself. To coherently imagine a situation that verifies S, one must be able to coherently imagine a situation such that reasoning about the imagined situation reveals it as a situation that verifies S. This notion is our core notion of positive conceivability: I will henceforth say that S is

positively conceivable when it is coherently modally imaginable.

One can then introduce prima facie and ideal versions of positive conceivability. S is prima facie positively conceivable when one can modally imagine a situation that one takes to be coherent and that one takes to verify S. S is ideally positively conceivable when S is prima facie positively conceivable and this positive conceivability cannot be undermined on idealized reflection. In effect, we can distinguish prima facie coherence from true coherence, and prima facie verification from true verification, where the "true" notions involve idealization on rational reflection. True coherence requires that arbitrary details can be filled in with no contradiction revealing itself on idealized reflection, whereas prima facie coherence requires merely the appearance of coherence. True verification requires that the imagined situation is revealed as a situation in which S even on idealized reflection, whereas prima facie verification requires merely the appearance that the imagined situation is a situation in which S. Then (invoking the "true" notions) one can say that S is ideally positively conceivable when one could coherently imagine a situation that verifies S.

When S is ideally positively conceivable, it must be possible in principle to flesh out any missing details of an imagined situation that verifies S, such that these details are imagined clearly and distinctly and such that no contradiction is revealed. It must also be the case that arbitrary rational reflection on the imagined situation will not undermine the interpretation of the imagined situation as one in which S is the case. These strictures are demanding, but they are not unreasonable. They are the strictures typically demanded of good thought-experiments, for example.

A typical philosophical thought-experiment starts with prima facie positive conceivability. A subject does not imagine a situation in fine detail: microphysical details are usually left unspecified, for example. Instead, a subject imagines a situation with certain important features specified, notes that a situation of this kind appears to verify S, and judges that the remaining details are not crucial: they can in principle be filled in to yield a full coherent conception of a situation that verifies S. For the thought-experiment to yield the intended conclusion, this prima facie judgment must be be correct, so that S is ideally positively

conceivable. If better reasoning would reveal that the details cannot be coherently filled in, or that the situation does not truly verify S, then the thought-experiment will typically fail in its purpose. If the prima facie judgment is not defeatable in this way, however, then the thought-experiment succeeds, and S is ideally positively conceivable.

Clear cases of prima facie positive conceivability without ideal positive conceivability are surprisingly hard to come by. Possible examples might include the two cases above: imagining an impossible object, and imagining a situation in which mathematicians announce that M (for some false M). In these cases, however, even a moment's reflection is enough to undermine the positive conceivability. In the first case, one can easily detect a contradiction (or the inability to fill in crucial detail). In the second case, reflection reveals the situation as one in which one has evidence that M, but as not necessarily a situation in which M. So these cases will be prima facie positively conceivable under only the most superficial of reasoning processes.

A slightly better example of prima facie without ideal positive conceivability may be the Grim Reaper paradox (Benardete 1964; Hawthorne 2000). There are countably many grim reapers, one for every positive integer. Grim reaper 1 is disposed to kill you with a scythe at 1pm, if and only if you are still alive then (otherwise his scythe remains immobile throughout), taking 30 minutes about it. Grim reaper 2 is disposed to kill you with a scythe at 12:30 pm, if and only if you are still alive then, taking 15 minutes about it. Grim reaper 3 is disposed to kill you with a scythe at 12:15 pm, and so on. You are still alive just before 12pm, you can only die through the motion of a grim reaper's scythe, and once dead you stay dead. On the face of it, this situation seems conceivable — each reaper seems conceivable individually and intrinsically, and it seems reasonable to combine distinct individuals with distinct intrinsic properties into one situation. But a little reflection reveals that the situation as described is contradictory. I cannot survive to any moment past 12pm (a grim reaper would get me first), but I cannot be killed (for grim reaper n to kill me, I must have survived grim reaper n+1, which is impossible). So the description D of the situation is prima facie positively conceivable but not ideally positively conceivable.

Note that the mathematical case is a case in the subject has coherently imagined a situation but in which the imagined situation does not verify S on reflection, while the Grim Reaper and impossible object cases are cases in a situation has not been coherently imagined. Of course in both these cases, the problem is revealed by a little reflection. One might say that in this case (and in the mathematical case above), even if we have prima facie positive conceivability, we do not have secunda facie positive conceivability.

Cases of secunda facie positive conceivability without ideal positive conceivability seem to be extremely thin on the ground. Perhaps the best candidates involve rational but false beliefs in an a priori domain such as mathematics. In general, the details of an imagined situation will be irrelevant to the positive conceivability of a mathematical claim, since reflection suggests that the truth of the mathematical claim is independent of the imagined goings-on in the world. Rather, a mathematical claim will be positively conceivable insofar as there is rational reason to accept that claim; in that case, any imagined situation can be taken to verify the claim. One will have secunda facie positive conceivability without ideal positive conceivability when these reasons stand up to secunda facie scrutiny, but are undermined by ideal reflection. Frege's set of all sets may be such a case: Frege had good a priori reasons for accepting it that survived considerable reflection, but ideal (or at least Russellian) reflection reveals a deep contradiction.

If S is positively conceivable, S is negatively conceivable (in both the prima facie and ideal cases). If one can coherently imagine a situation verifying S, then one cannot rule out that S (though this interacts a little with the primary/secondary distinction below). The reverse is not the case, at least where prima facie conceivability is concerned: many statements are prima facie negatively conceivable without being prima facie positively conceivable. For example, as we saw above, many complex mathematical statements M are such that one cannot rule out M's truth, but one cannot imagine any situation (any part of a world) that verifies S. Something similar goes for statements in other a priori domains. And even in empirical domains, it may be that one cannot rule out M, but one cannot conceive of a situation in which M, due to limited powers of imagination, for example.

Clear cases of ideal negative conceivability without ideal positive conceivability are much harder to find. One might try mathematical statements that are true but not knowable a priori by any possible being. If there were such statements, they and their negations would be ideally negatively conceivable, but probably not ideally positively conceivable. But it is far from clear that there are any such statements. I will return to this matter later.

Positive conceivability, rather than negative conceivability, seems to be what most philosophers have had in mind when discussing conceivability. It is positive conceivability that corresponds to the sort of clear and distinct modal intuition that Descartes had in mind, and which reflects the practice in the method of conceivability as used in contemporary philosophical thought-experiments. When Yablo (1993) dismisses the first Goldbach example as not really being an instance of conceivability, he is in effect saying that negative conceivability is not true conceivability, and there is something to this.

Still, it must be conceded that negative conceivability is at least better defined than positive conceivability. The characterization of positive conceivability that I have given here, invoking the notion of a modally imagining a situation, cannot be considered a reductive definition. At best, it is something of a clarification. Nevertheless, there seems to be a reasonably clear intuitive notion in the vicinity, which most people seem to have a grasp on. It may be that the notion can be given a more rigorous definition, or it may be that it should be taken as primitive; this is one of the central open questions in the area.

The distinction between positive and negative conceivability bears at least some relation to van Cleve's (1983) distinction between strong and weak conceivability. According to van Cleve, S is strongly conceivable for a subject when the subject sees that S is possible; and S is weakly conceivable when the subject does not see that S is impossible. There is an obvious link between one reading of "seeing that S is impossible" and the idea of ruling out the hypothesis that S. And the notion of "seeing that S is possible" can be read as a sort of modal intuition that S of the sort that goes along with modally imagining that S.

I think that it is best not to import the notion of possibility so directly into a definition of conceivability, to avoid the threat of trivializing the link with possibility. In particular, there is a threat that the idealized version of seeing that S is possible will collapse into correctly judging that S is possible, which will be linked trivially to possibility. Still, the idea is closely related to the idea of coherently imagining a world (or a part of a world) that verifies S: both involve a sort of modal appearance. The main advantage of the construal I have given is that it builds in no presupposition that the imagined world is metaphysically possible, or even that it seems metaphysically possible. It builds in some broadly modal elements, in the ideas of imagining a world, of coherence, and of verification. But importantly, the modalities here are cognitive or epistemic, and presuppose no tie to the metaphysical. To imagine a world is simply to engage in a distinctive and familiar sort of mental act; and the notions of coherence and verification are wholly grounded in rational notions. So there is no danger of trivializing the link between positive conceivability and possibility.

3 Primary vs. Secondary Conceivability

This distinction draws its motivation from Kripke's discussion of the necessary a posteriori. In the wake of Kripke's arguments that a posteriori statements such as "Hesperus is Phosphorus" are necessary, it has become a familiar observation that there is a sense in which "Hesperus is not Phosphorus" is conceivable, and a sense in which it is not. The first of these senses corresponds to primary conceivability; the second to secondary conceivability.

We can say that S is *primarily conceivable* (or *epistemically conceivable*) when it is conceivable that S is *actually* the case. We can say that S is *secondarily conceivable* (or *subjunctively conceivable*) when S conceivably *might have been* the case. This corresponds to two different ways of thinking about hypothetical possibilities: epistemically, as ways the world might actually be, and subjunctively, as counterfactual ways the world might have been. I have written more on these distinctions elsewhere, but I will give a short characterization here.

It is simplest to start with the case of positive conceivability. When one

imagines a situation, one can consider it *as actual* (as a way the world might actually be), or one can consider it *as counterfactual* (as a way the world might have been). It is often the case that one will describe a situation differently depending on whether one considers it as actual or as counterfactual. We can say that S is primarily positively conceivable when one can coherently imagine a situation that verifies S when considered as actual, and that S is secondarily positively conceivable when one can coherently imagine a situation that verifies S when considered as counterfactual.

Primary conceivability is grounded in the idea that for all we know a priori, there are many ways the world might be. The oceans might contain H_2O or they might contain XYZ; the evening star and the morning star might be the same or distinct; and so on. We can think of these ways the world might be as *epistemic possibilities*, in a broad sense according to which it is epistemically possible that S if the hypothesis that S is not ruled out a priori. When S is epistemically possible, there are usually a number of imaginable situations such that if they actually obtainn S will be the case. These situations can be taken to verify S, when they are considered as actual.

For example, it is epistemically possible in this sense that Hesperus is not Phosphorus (it is not a priori that Hesperus is Phosphorus). In the background of this epistemic possibility are many specific epistemically possible situations in which the heavenly bodies visible in the morning and evening are distinct. Upon consideration, these epistemically possible situations are revealed as instances of the epistemic possibility that Hesperus is not Phosphorus. There is a clear sense in which these situations *verify* the claim that Hesperus is not Phosphorus: for example, if one hypothetically accepts that such a situation actually obtains, one should rationally conclude that Hesperus is not Phosphorus. This sort of relation among epistemic possibilities plays a central role in our thought.

When we consider situations as actual, we consider and evaluate them in the way that we consider and evaluate epistemic possibilities. That is, we say to ourselves: what if the actual world is really that way? One hypothetically assumes that the situation in question is actual, and considers whether, from that assumption, it follows that S is the case.

If so, then the situation verifies S, when considered as actual. In the case above, for example, the situations in question (considered as actual) verify "Hesperus is not Phosphorus". So "Hesperus is not Phosphorus" is primarily positively conceivable.

(Primary conceivability is related to what Yablo (1993) calls "conceivability_ep", which requires that one can imagine believing something true with one's actual P-thought, but it is not quite the same. One difference is that primary conceivability does not require that a conceived situation contain a P-thought. So it is primarily conceivable that nothing exists, or that no-one thinks — these are not ruled out a priori, and are verified by certain situations considered as actual — but they are not conceivable in Yablo's sense.)

Negative, positive, prima facie, and ideal versions of primary conceivability are easy to formulate. We can say that S is primarily negatively conceivable when it is not ruled out a priori that S is actually the case, or more briefly, if S is not ruled out a priori. Positive primary conceivability, by contrast, requires coherently imagining a situation (considered as actual) that verifies S. Prima facie and ideal versions of these notions can be straightforwardly formulated as in the previous section. Primary positive conceivability implies primary negative conceivability, for both the prima facie and ideal versions, but the reverse is not obviously the case.

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.

Secondary conceivability works quite differently. It is grounded in the idea that we can conceive of many counterfactual ways that the world might have been but is not. When we consider imagined situations as counterfactual, we consider and evaluate them in the way that we consider and evaluate counterfactual possibilities in the subjunctive mode. That is, we acknowledge that the character of the actual world is fixed, and say to ourselves: if the situation *had* obtained, what *would have been* the case? If we judge that had the situation obtained, S would have been the case, then we judge that the situation verifies S

when considered as counterfactual.

Take an imagined situation in which the morning star is distinct from the evening star. Along with Kripke, we can say that if this situation had obtained, it would not have been the case that Hesperus was not Phosphorus. So when this situation is considered as counterfactual, it is revealed not as a situation in which Hesperus is not Phosphorus, but rather as a situation in which at least one of the objects is distinct from both Hesperus and Phosphorus (at least if we take for granted the actual-world knowledge that Hesperus is Phosphorus, and if we accept Kripke's intuitions). The reason is that (if Kripke is right) the application of a term like "Hesperus" to a counterfactual situations depends on whether the actual Hesperus (i.e. the planet Venus) is present within that situation, and of course the actual Hesperus and the actual Phosphorus are one and the same. So when considered as counterfactual, this conceivable situation does not verify "Hesperus is not Phosphorus". More generally (if Kripke is right), there is no coherently imaginable situation, considered as counterfactual, that verifies "Hesperus is not Phosphorus". If so, "Hesperus is not Phosphorus" is not secondarily positively conceivable.

Unlike primary conceivability, secondary conceivability is often a posteriori. It is not secondarily conceivable that Hesperus is not Phosphorus, but one could not know that a priori. To know this, one needs the empirical information that Hesperus is actually Phosphorus. This aposteriority is grounded in the fact that the application of our words to subjunctive counterfactual situations often depends on their reference in the actual world, and the latter cannot usually be known a priori.

There are various ways to formulate prima facie and ideal versions of secondary conceivability. One might say that a subject prima facie secondarily conceives of S when the subject imagines a situation and judges that if that situation had obtained, S would have been the case. One can say that S is ideally secondarily conceivable if S is prima facie secondarily conceivable and if the secondary conceivability is not defeatable by idealized rational reflection and complete empirical knowledge. To avoid trivializing a link between conceivability and possibility here, it is probably best to restrict the empirical knowledge

in question to nonmodal knowledge.

This characterizes positive versions of secondary conceivability. One might say that S is negatively secondarily conceivable when a priori reflection and empirical nonmodal knowledge reveals no incoherence in the hypothesis that S might have been the case. In any case, as secondary conceivability turns on a posteriori considerations, it will not be our central concern, and most of these varieties can be set to one side.

4 Gaps between Conceivability and Possibility

With the distinctions above in play, it relatively easy to classify potential gaps between conceivability and possibility.

> (1) *Prima facie conceivability is an imperfect guide to possibility.*

Given that there is a gap between prima facie and ideal conceivability, it is only to be expected that there is a gap between prima facie conceivability and possibility. Prima facie conceivability judgments are sometimes undermined by continued rational reflection, isolating a contradiction or a misdescription in an apparently conceivable state of affairs. When this happens, then any grounds that the conceivability judgment provided for a claim of possibility will also be undermined.

This gap is widest in the case of prima facie negative conceivability judgments. When such a judgment is not backed by a corresponding prima facie positive conceivability judgment, it provides at best weak evidence for possibility. Mathematical cases, such as the prima facie negative conceivability of both Goldbach's conjecture and its negation, provide an obvious source of gaps here. So likewise does any domain in which one might expect to find deep a priori truths.

Prima facie positive conceivability is a much better guide to possibility, but it is still imperfect. The case where one conceives of mathematicians announcing a proof of Goldbach's conjecture (or its negation) is best seen as a case where a superficial prima facie positive conceivability judgment is undermined by a moment's reflection. Other

cases of prima facie conceivability without possibility may be provided by the Grim Reaper paradox and the case of impossible objects.

Cases of secunda facie positive conceivability, where a prima facie positive conceivability judgment survives a reasonably searching process of rational reflection, are a still stronger guide to possibility. In the great majority of cases with a gap between prima facie and ideal positive conceivability, the prima facie judgment is easily undermined by a little reflection. Gaps between secunda facie positive conceivability and ideal positive conceivability seem to be very rare, although perhaps the Frege case is an example.

In any case, if we are looking for a notion of conceivability such that conceivability tracks possibility perfectly, we must focus on ideal conceivability. In this sense conceivability is not a merely psychological notion; it is a *rational* notion, in much the same way that a priority and rational entailment are rational notions. If there is to be a plausible epistemic/modal bridge, it will be a bridge between the rational and modal domains.

> (2) *Positive conceivability is a better guide to possibility than negative conceivability.*

We have seen that prima facie negative conceivability is a relatively weak guide to possibility. The canonical case here is the prima facie negative conceivability of both Goldbach's conjecture and its negation. These cases are not backed by a corresponding prima facie positive conceivability judgment, except for a very superficial judgment in one case. So at least where prima facie conceivability is concerned, positive conceivability is a much better guide to possibility than negative conceivability. This fits the usual practice in philosophy, where the conceivability judgments that are usually taken as evidence of possibility are almost always positive conceivability judgments. (For just this reason, the Goldbach case was never a very compelling counterexample to this practice.)

With ideal conceivability, things are less clear. Certainly ideal positive conceivability is at least as good a guide to possibility as ideal negative conceivability, since the former entails the latter. What is less clear is

whether there are cases of the latter without the former, and if so, whether those cases correspond to possibilities.

The most obvious potential case here is an extension of the Goldbach case above. If either Goldbach's conjecture or its negation is provable, or otherwise knowable a priori, then only one will be ideally negatively conceivable. But perhaps (as noted earlier) there are some true or false mathematical statements whose truth-value cannot be settled even by ideal rational reflection? If so, we would have cases of ideal negative conceivability without ideal positive conceivability and without possibility. It is not at all clear that cases of this type exist, however. I will discuss this and other potential counterexamples to a link between ideal negative conceivability and possibility later. It seems that there are at least no clear counterexamles, so a link between ideal negative conceivabilty and possibility remains tenable.

Overall, we can say that both ideal positive conceivability and ideal negative conceivability are promising as guides to possibility, but that the former is in a slightly better position to be a perfect guide than the latter, due to its added strength.

> (3) *Primary conceivability is an imperfect guide to secondary possibility.*

The other standard source of gaps between conceivability and possibility arises from Kripkean cases. It is often said that it is conceivable that Hesperus is not Phosphorus, or that water is not H_2O, or that heat is not the motion of molecules, but none of these states of affairs are in fact possible. In these cases we have a posteriori necessities and impossibilities, out of the reach of a priori methods.

There are a couple of things to be said here. Clearly, the main sense in which these states of affairs are conceivable involves primary conceivability. As discussed earlier, the states of affairs in question are not secondarily conceivable. At best, they might be prima facie secondarily conceivable for a subject lacking relevant empirical knowledge. They will not be prima facie secondarily conceivable for a subject with the relevant knowledge, and they will not be ideally secondarily conceivable as that notion is spelled out above.

One might then try to save a conceivability/possibility link by suggesting that ideal secondary conceivability entails possibility. This thesis is not implausible, but it is not helpful for our purposes here. The reason is that secondary conceivability, and especially ideal secondary conceivability, is deeply a posteriori. So even if secondary conceivability is a guide to possibility, it will yield no a priori access to modality.

(Around this point, it seems to me that the otherwise excellent discussions of conceivability and possibility by Menzies (1998), van Cleve (1983), and Yablo (1993) all give up too soon, settling for conceivability/possibility theses that are more attenuated than necessary.)

If we are interested in modal rationalism, we should instead focus on ways in which primary conceivability might still be a guide to possibility. Even if it is conceded that strictly speaking, it is not possible that water is H_2O, it can still be argued that the primary conceivability of "water is not H_2O" is revealing something about metaphysical possibility. When we apparently conceive of a world in which water is not H_2O, we conceive of a situation in which some other substance (XYZ, say) is the clear liquid surrounding us in the oceans and lakes, and so on. And this situation is indeed metaphysically possible — so our act of conceiving has indeed yielded access to a possible world. It is just that in a certain sense we have misdescribed it in calling it a world where water is not H_2O, or a world in which water is XYZ. If Kripke is right, it is in fact a world in which XYZ is watery stuff but not water, and a world in which the only water that exists is H_2O.

Further: there remains a sense in which a world with XYZ in the oceans can be seen as satisfying the statement "water is not H_2O". Here, I will give a very brief version of a story that I have told in more detail elsewhere (e.g. Chalmers 1996, forthcoming b; see also Evans 1977, Davies and Humberstone 1980, and Jackson 1998).

As discussed earlier, there is clearly a broad sense in which it is *epistemically* possible that water is not H_2O, in that the hypothesis is not ruled out a priori. Intuitively, there are ways our world could turn

out such that if they turn out that way, it will turn out that water is not H_2O. And if we consider the XYZ-world as an epistemic possibility — that is, we consider the hypothesis that the world with XYZ in the oceans is *our* world — then this epistemic possibility can be seen as an instance of the epistemic possibility that water is not H_2O. We can rationally say "if our world turns out to have XYZ in the oceans (etc.), it will turn out that water is not H_2O". This might be put as a simple indicative conditional: "if XYZ is in the oceans and lakes (etc.), then water is XYZ". Compare: "if Prince Albert Victor committed those murders, then he is Jack the Ripper". Here, the indicative conditional "if P, then Q" can be evaluated using the Ramsey test: if one hypothetically accepts the belief that P, does one arrive at the conclusion that Q?

All this reflects the fact that we have a systematic way of evaluating and describing epistemic possibilities that differs from our way of evaluating and describing subjunctive counterfactual possibilities. In both cases, we consider and describe worlds, but in the epistemic case, we consider them as actual, whereas in the subjunctive case, we consider them as counterfactual. And these two modes of consideration of a world yield two ways in which a world might be seen to satisfy a sentence. When the XYZ-world is considered as actual, it satisfies "water is XYZ"; when it is considered as counterfactual, it does not.

Given a statement S and a world W, the *primary intension* (or *epistemic intension*) of S returns the truth-value of S in W considered as actual. Three heuristics for evaluating the primary intension of S in W correspond to the three tests mentioned above. One can appeal to direct evaluation of epistemic possibilities: is the epistemic possibility that W is actual an instance of the epistemic possibility that S? One can appeal to indicative conditionals (evaluated by the Ramsey test): if W is the case, is S the case?" Or one can appeal to the "turns out" locution: if W turns out to be actual, will it turn out that S?

Primary intensions can be formally defined in terms of a priori entailments. In particular, we can say that the primary intension of S is true in W if the material conditional "if W is actual, then S" is a priori: that is, if the hypothesis that W is actual and S is not the case can be ruled out a priori. S's primary intension is false in W if the conditional

"if W is actual, then ~S" is a priori; and S's primary intension is indeterminate at W is neither of these conditionals priori. For example, the hypothesis that the XYZ-world is actual and water is H_2O can plausibly be ruled out conclusively by rational reflection alone. If so, the material conditional "if the XYZ-world is actual, then water is not H_2O" is a priori, and the primary intension of "water is H_2O" is false in the XYZ-world. For more on the definition of primary intensions, see the further discussion below.

Primary intensions are grounded in the *epistemic* evaluation of statements in worlds: that is, the evaluation of statements in worlds considered as actual. One can also define the notion of a *secondary* (or *subjunctive*) intension, grounded in the subjunctive evaluation of statements in worlds: that is, the evaluation of statements in worlds considered as counterfactual. The secondary intension of a statement S is the function that maps a world W to the truth-value of S in W considered as counterfactual. These correspond to a much more familiar notion of intension in contemporary philosophy, so I will say less about them here.

To characterize secondary intensions with a heuristic, one can appeal to subjunctive conditionals: if W had obtained, would S have been the case? Or one can appeal directly to intuitions about counterfactual possibilities: is W a counterfactual possibility in which S would have been the case? Heuristics of this sort are frequently invoked by Kripke in his evaluation of possible worlds; and his corresponding influential claims about possibility are almost always grounded in subjunctive claims about what might have been the case. So the intensional notions that arise from Kripke's work are all closely tied to secondary intensions.

A paradigmatic example involves the subjunctive evaluation of a statement such as "water is XYZ" at the XYZ-world, a world that is similar to our own except that the watery liquid in the oceans and lakes is XYZ. If Kripke and Putnam are correct, then if the watery stuff in the oceans and lakes had been XYZ, it would nevertheless not have been the case that water was XYZ: at best, XYZ would have been watery. Corresponding, W does not seem to represent a counterfactual possibility in which water is XYZ. So the secondary intension of "water

is XYZ" is false at the XYZ-world.

We can then say that S is *primarily possible* (or 1-possible) if its primary intension is true in some possible world (i.e. if S is true in some world considered as actual). S is *secondarily possible* (or 2-possible) if its secondary intension is true in some possible world (i.e. if S is true in some world considered as counterfactual). Primary and secondary necessity can be defined analogously.

Secondary possibility and necessity correspond to the standard conception of what it is for a statement to be metaphysically possible or necessary. For example, "water is H_2O" is plausibly 2-necessary, and "water is XYZ" 2-impossible, reflecting their metaphysical necessity and impossibility (as standardly understood) of respectively. On this understanding, we can say that a statement is metaphysically necessary iff it has a necessary secondary intension.

Primary possibility and necessity correspond much more closely to epistemic notions such as apriority. It is clear that when S is a priori, it will have a necessary primary intension, so it will be 1-necessary. Whether the reverse entailment (from 1-necessity to apriority) holds is one of the central issues in this paper, but for now we can note that at least the clearest cases of 1-necessary statements are all plausibly a priori: witness "2+2=4", or "Hesperus, if it exists, is visible in the evening" (1-necessary and a priori), as opposed to "tables exist" and "water is H_2O" (1-contingent, and a posteriori).

The existence of primary and secondary intensions suggests that expression tokens have a complex semantic value that involves both intensions. These intensions will play important roles when the expression is embedded in different contexts. In constructions such as "it might have been the case that S" and subjunctive conditionals, S's secondary intension will be relevant. In constructions such as "it is a priori that S" and indicative conditionals, S's primary intension will be relevant. Both intensions are part of the content of S in both contexts: it is just that the different contexts exploit different aspects of S's content. The propositional content of S might be understood in a number of different ways, but if one holds that the apriority and necessity of S is a function of the proposition that S expresses, then the

proposition expressed by S will be reducible to neither its primary intension nor its secondary intension, but will rather be something that involves at least the structure of both.

We can now see how primary conceivability can act as a guide to possibility. When we find it conceivable that water is not H_2O, there is no possible world that satisfies "water is not H_2O" when the world is considered as *counterfactual*, but there is a possible world that satisfies "water is not H_2O" when the world is considered as *actual*. Put differently, the secondary intension of "water is H_2O" is true in no world, but the primary intension is true in some (centered) worlds. The XYZ-world, and other centered worlds that we might conceive of when we conceive that water is not H_2O, all satisfy the primary intension of "water is H_2O".

We can put this by saying that primary conceivability is an imperfect guide to secondary possibility, but is a much better guide to primary possibility. In all the Kripkean cases in which S is primarily conceivable, S is also primarily possible (or at least Kripke's discussion gives no reason to deny this). There is a (centered) possible world satisfying the primary intension of "Hesperus is not Phosphorus" (e.g. a world where heavenly bodies visible from the center in the morning and the evening are distinct), of "heat is not the motion of molecules" (e.g. a world where something else causes heat sensations), and so on. These worlds are all first-class metaphysical possibilities.

So Kripke's examples are entirely compatible with the thesis that conceivability is a guide to possibility. We just need to make sure that the relevant notions are aligned: primary conceivability is a guide to primary possibility, and secondary conceivability is a guide to secondary possibility. This is no surprise: it would be odd to expect conceivability of a situation considered as actual to be a guide to possibility of a world considered as counterfactual, or vice versa! So we are still left with significant a priori access to the space of possible worlds.

5 Sideline: On Defining Primary Intensions

Primary intensions are intensions that capture the distinctive way a statement is used to describe and evaluate epistemic possibilities. The primary intension of a statement could be defined in various ways, but the most useful definition is that in terms of a priori entailments: the primary intension of S is true at W if the material conditional "if W is actual, then S" is a priori. I elaborate and defend this conception of a primary intension in other work; here I will make a few observations about the definition of primary intensions and about their properties. This material can be skipped by those who are not interested in the fine details of the two-dimensional framework.

(i) For a world to be considered as actual, it must be a *centered* world: a world marked with a specified individual and time. The reason is that an epistemic possibility is not completely determined until one's "viewpoint" is specified. For example, an objective description of the world will not allow me to settle the question of whether I am in Australia or in the US, but a "centered" specification will do this. The hypothesis that a centered world W is actual, for me, will include the hypothesis that I am the being marked at the center and that now is the time marked at the center. A primary intension can then be seen as a function from centered world to truth-values. The primary intension of "I am a philosopher", for example, will be true at those centered worlds in which the subject at the center is a philosopher.

(ii) The evaluation of a conditional involving "If W is actual..." requires a *canonical description* of W. We can say that the primary intension of S is true at W if the material conditional "if D, then S" is a priori, where D is a canonical description of W. The notion of a canonical description can be elaborated in various different ways, which are too complex to discuss in detail here. One needs to isolate a semantically neutral vocabulary in which worlds can be described, and to require a certain sort of complete description within this vocabulary. On the first point, a semantically neutral expression might be seen intuitively as one which behaves the same way in epistemic and subjunctive evaluation, so that it is not susceptible to Twin-Earth thought experiments" (supplemented by indexicals such as "I" and "now" to handle centering). On the second point, one might require a complete description to be ontologically complete, or qualitatively complete, or epistemically complete, in the terms from later in this paper. If the theses of this paper are correct,

these different notions of completeness are coextensive. If the theses of this paper are incorrect, these notions may come apart, yielding different primary intensions. In that case, it is probably best to require epistemic completeness in the definition.

(iii) Primary intensions are defined here as functions over (centered) possible worlds. One can also define a closely related intension as a function over an independently characterized *epistemic space* of maximal epistemic possibilities. Epistemic space is not defined in terms of metaphysical possible worlds, but rather in terms of epistemic notions such as apriority: maximal epistemic possibilities correspond roughly to maximally specific a priori consistent hypotheses concerning the actual world. One can define an intension over this space much as one defines a primary intension. In other work (e.g. Chalmers forthcoming a,b), I have called this an *epistemic intension*.

What is the relationship between the two notions? This relationship turns on the relationship between epistemic space and the space of centered possible worlds, which in turn is closely tied to the relationship between ideal negative conceivability and primary possibility. If this paper's theses are correct, there is a direct correspondence between the two spaces, so that primary intensions as defined here and epistemic intensions are almost identical. If this paper's theses are incorrect, then the definitions come apart: there will be maximal epistemic possibilities that correspond to no centered possible worlds, so the intensions will be defined over different spaces.

For many purposes, especially within the epistemic domain, the notion of an epistemic intension is more fundamental. For example, necessity of epistemic intension is constitutively tied to apriority and other epistemic notions, independently of any views about metaphysical possibility. So epistemic intensions can be used for epistemic purposes regardless of one's further views. For present purposes, however, the link between the epistemic and metaphysical domain is the central focus, so I focus here on primary intensions understood as functions over metaphysically possible worlds. If what I say in this paper is correct, the two intensions ultimately collapse into one.

(iv) The primary intension of some terms can vary between speakers.

For example, Leverrier might use "Neptune" to pick out whatever causes certain orbital perturbations within a world, whereas a friend might use it to pick out (roughly) whatever Leverrier refers to with the name, irrespective of any perturbing role. If so, their primary intensions will vary accordingly. All this reflects the fact that certain conditionals of the form "if W is actual, then Neptune is such-and-such" are a priori for Leverrier but not for his friend. This happens not because of any difference in their rational capacities (which we are idealizing away from), but because of differences in the inferential roles associated with the term. Something similar can happen with most names and natural kind terms.

It follows that primary intensions are not candidates for linguistic meaning, the sort of meaning common to all tokens of an expression type, at least where names and natural kind terms are concerned. (See Chalmers forthcoming a.) To accommodate this phenomenon, primary intensions should be associated in the first instance with expression tokens (or perhaps with types as used on occasions), not with expression types. We can define primary intensions more precisely by saying that the primary intension of a statement token S (used by a speaker) is true in W if the material conditional "if W, then S" is a priori for the speaker. Here, a sentence T will be a priori for a speaker if the belief (or the hypothesis) that the speaker expresses with T could be conclusively justified, on ideal rational reflection, with justification independent of experience. On this account, different material conditionals will be a priori for Leverrier and his friend, so their primary intensions for "Neptune" will differ accordingly.

Note that the notion of apriority (whether speaker-relative or speaker-independent) requires the same sort of rational idealization as that present in the notion of ideal conceivability. I have defended the claim that relevant conditionals are a priori elsewhere (see also the discussion of scrutability later in this paper). If someone is skeptical about this, or skeptical about the very notion of apriority, it may nevertheless remain plausible that the material conditionals in question have *some* distinctive epistemic status which can be used to define a corresponding notion of primary intension.

(v) To evaluate the primary intension of S in W, it is not required that W

contain a token of S. The heuristics and definition above give no special role for such a token, even when it is present. On another approach, one could define the *contextual intension* of S as a function defined across worlds containing a token of S at the center, returning the truth-value of that token. Contextual intensions are closely related to Stalnaker's *diagonal proposition* (1978), which is also defined in terms of the semantic values of a token in different contexts. These notions differ in fundamental respects from the current notion of a primary intension, which is grounded in the epistemic domain. Contextual intensions turn on the context-dependence of a statement's extension, while primary intensions turn on the use of a statement in evaluating epistemic possibilities.

To see some differences, note that the contextual intension of statements such as "language exists" will plausibly be nowhere false, but the primary intension of "language exists" will be false in many (language-free) centered worlds. This reflects the (broad) epistemic possibility of such worlds: it is not a priori that language exists. Something similar applies to "nothing exists" (whose primary intension is true of an empty world) and many claims about thinkers and about language. The contextual intension also requires an account of what it takes for a token to count as an instance of S's type, raising problems (pointed out by Block and Stalnaker 1999), that tend to break the link between contextual intensions and epistemic notions. If we individuate S's type orthographically, "bachelors are unmarried" has a contingent contextual intension; if we individuate by familiar sorts of semantic content, "water is XYZ" has a necessarily false contextual intension; if we individuate by "narrow content" or some such, then we need an independent account of that sort of content. This issue does not arise for primary intensions. The effect is that primary intensions are much more directly connected to the epistemic domain than are contextual intensions.

This distinction is useful in assessing the relationship between the conceivability-possibility theses I am putting forward here and a related thesis put forward by Kripke (1980). Kripke suggests that when a necessary claim (such as 'heat is the notion of molecules') is 'apparently contingent', then in a qualitatively identical evidential situation, a qualitatively identical statement might have been false.

Translated into the existing framework, this appears to come roughly to the thesis that when S is primarily conceivable for a subject, there is a world at which a certain sort of contextual intension of S is true. This contextual intension is defined at worlds whose center contains a subject in a qualitatively identical evidential situation as the original subject, uttering a qualitatively identical statement. At this world, the contextual intension returns the truth-value of the statement uttered at the center.

This thesis roughly parallels the thesis I have offered here, except with a sort of contextual intension in place of an epistemic intension. It is not entirely clear how to understand the notions of 'qualitatively identical' here, but however the notion is understood, the thesis appears to be false. The reasons are closely related to the considerations about 'Language exists' and the like, above. To invoke cases parallel to the case of heat, we can let 'Bill' be a term introduced to rigidly designate whatever color quale is in the center of my visual field now, or let 'L' be a term introduced to rigidly designate the number 1 if there are languages, and 0 if there are no languages. Then 'Bill=blue' and 'L>0' are a posteriori necessities, associated with the usual sense of apparent contingency. But they fail Kripke's test: they have a necessary contextual intension (of the relevant kind), and all qualitatively identical statements uttered in identical evidential situations will be true. The same problem applies to an adaptation of Kripke's thesis by Bealer (1996), suggesting that a posteriori necessities do not arise with expressions using 'semantically stable' terms — terms whose meaning does not vary across qualitatively identical epistemic situations — so that a priori modal intuitions using these expressions are reliable. Terms such as 'Bill' and 'L' are semantically stable by Bealer's definition, but still yield a posteriori necessities.

In contrast, the thesis I have offered handles these cases straightforwardly. 'Bill=blue' and 'L=1' have a contingent primary intension: the first is false at a world where the subject at the center is has no blue experiences, and the second is false at a world without language. So the inference from primary conceivability of these statement's negations to their primary possibility goes through straightforwardly. This suggests that epistemically defined notions are

more fundamental than contextually defined notions here. (For much more on this matter, see Chalmers forthcoming b.)

(vi) Yablo (this volume) considers various ways in which the epistemic evaluation of statements in worlds might be defined. He rejects both the indicative conditional heuristic and the "turns out" heuristic, on the grounds that they give the wrong result in certain metalinguistic cases. For example, `tail are wings' should be false in a world (considered as actual) where `tail' is used to refer to wings. But Yablo suggests that the indicative conditional "if `tail' refers to wings, then tails are wings' is intuitively correct, as is "if it turns out that `tail' refers to wings, then it will turn out that tails are wings".

I think these judgments of intuitive correctness are unclear, and I think there is at least a reasonable reading of the locutions on which the conditionals in questions are incorrect. (Compare the reasonable: "if `tail' refers to wings, 'tail' does not refer to tails"). But even if Yablo were right about this, it would show only that the heuristics are imperfect, giving the wrong results in special cases. The problem does not arise for the definition I have given here, in terms of a priori material conditionals.

To see this, note that the claim that 'tail' refers to tails is not a priori, but represents substantive a posteriori metalinguistic knowledge. It is a posteriori that the orthographic string 'tail' means anything at all, and it is a posteriori that it means what it does. So there is no a priori entailment from claims involving "'tail'" to corresponding claims involving 'tail'. And in particular, there is no a priori entailment from "'tail' refers to wings" to 'tails are wings'. More generally, there are plausibly no substantive a priori connections between claims about the orthographic string 'tail' and claims about tails, since any inferential connections between these claims rest on a posteriori metalinguistic knowledge. If so, the way that 'tail' is used in such a world will be irrelevant in evaluating the primary intension of statements such as 'tails are wings' in that world. In particular, there is no danger that 'tails are wings' will be true in a world (considered as actual) in which 'tail' refers to wings.

Someone might suggest that there is a semantic concept of "'tail'" that

builds in semantic constraints as well as orthographic constraints, so that it is a priori that 'tail' refers to tails. But then the worlds Yablo considers, in which the orthographic string 'tail' refers to wings, will not be worlds in which 'tail' (construed semantically) refers to wings, so there is no danger that they will be worlds (considered as actual) in which tails are wings. Either way, the usage of the orthographic string 'tail' in a world will be irrelevant in evaluating the primary intension of 'tails are wings' in that world.

Yablo himself endorses a mixed view, on which it is a priori that 'sister' refers to sisters (like the semantic view), but on which it is not a priori that 'sister' refers to female siblings (like the orthographic view), even though it is a priori that sisters are female siblings. A mixed view like this cannot be accommodated in the current framework: the framework requires that apriority is preserved under a priori entailment, but Yablo's view violates this. (For the relevant A, B, C, it is a priori that A, that B, that if A and B then C, but not that C). But it can plausibly be argued that this violation of closure is reason enough to reject Yablo's mixed view. (The view seems to be grounded in an idiosyncratic conception of a priori knowledge, on which a priori knowledge that S depends on metalinguistic knowledge concerning 'S'. I think that such a view should clearly be rejected.) In any case, it seems that any residual problems here arise from Yablo's somewhat idiosyncratic view of these metalinguistic cases, and not from the cases themselves.

For his own view of the epistemic evaluation of statements in worlds, Yablo endorses a "could have turned out" heuristic: "If it had turned out that W, would it have turned out that S?". Although I have occasionally used this heuristic myself in earlier work, I am less comfortable with it than with the "turns out" heuristic, as the subjunctive conditional here can easily be read non-epistemically, and it is too close to the subjunctive "If it had been that W, it would have been that S" for comfort. (I am also worried that Yablo's paper has the wrong title — most uses of "coulda", "woulda", and "shoulda" go with secondary intensions (!), as characterized below.) Still, it may be that there is at least a reading of this locution that gives approximately correct results in most cases.

6 Conceivability/Possibility Theses

To summarize: if any variety of a priori conceivability entails possibility, it must be a variety of ideal primary conceivability, and the variety of possibility that is entailed must be primary possibility. And positive conceivability is always at least a good a guide to possibility as negative conceivability. This leaves us with the following as the most plausible entailment between conceivability and possibility:

> (1) *Ideal primary positive conceivability entails primary possibility.*

That is, if S is ideally primarily positively conceivable, then there is some metaphysically possible centered world satisfying S's primary intension (or that satisfies S when considered as actual).

We have also left open the status of the following:

> (2) *Ideal primary negative conceivability entails primary possibility.*

For completeness, I note that the following two theses also remain plausible, although neither suffices for a thoroughgoing modal rationalism.

> (3) *Secunda facie primary positive conceivability is an extremely good guide to primary possibility.*

> (4) *Ideal secondary (positive/negative) conceivability entails secondary possibility.*

We have seen that the first thesis is compatible with the standard clear counterexamples to a link between conceivability and possibility, as is the second (although there are some unclear counterexamples that may threaten the second). So if there are any counterexamples to these two theses, they must come from a different source, and their existence will gain no support from the standard cases.

For my part, I think that thesis (1) is almost certainly true, and that thesis (2) is very likely true. In most of the rest of the paper, I will

discuss what counterexamples to these theses would involve, and give a quick sketch of reasons to think the theses true.

Note that because the conceivability and possibility of a statement is speaker relative, the conceivability-possibility theses above must be interpreted in a speaker-relative way: if S is conceivable for a speaker, S is possible for that speaker. There are two main sources of speaker-relativity here: variation in cognitive capacity, which affects prima facie conceivability, and variation in primary and secondary intensions of terms, which affects primary conceivability/possibility and secondary conceivability/possibility respectively.

For the central theses (1) and (2), only the second sort of variation is relevant. This variation manifests itself in phenomena such as the following: "Neptune does not perturb the orbit of Uranus" may be ideally primarily conceivable for Leverrier's friends but not for Leverrier himself; it will also be primarily possible for Leverrier's friends, but not for Leverrier himself. In a similar way, "I am not David Chalmers" may be ideally secondarily conceivable for you but not for me; it is also secondarily possible for you but not for me. Because the variation here affects conceivability and possibility equally, it does not threaten an inference from conceivability to possibility. It suggests at most that in cases where this variation is present, the inference must be speaker-relative.

One might worry that because the notion of ideal conceivability itself involves the notion of possibility (for example, in claims about what some possible being could conceive, or about what is defeatable), there is a danger of circularity. There are a few different issues here. First, one might worry that this rules out a reduction of possibility to conceivability. But I am not trying to give such a reduction, but am simply investigating the connection between the two notions. Second, one might worry that circularity will make the conceivability-possibility thesis trivial. But the notion of possibility enters into the definition of conceivability in such a roundabout way that the thesis clearly remains substantive. Finally, one might worry that defining conceivability in terms of possibility renders conceivability toothless as an epistemic guide to possibility, and so defeats modal rationalism. But this is not so: modal rationalism holds that modality is a priori *accessible*, and so

invokes the notion of possibility in a precisely parallel manner. If ideal conceivability tracks possibility, then modal facts are rationally accessible, as required.

(If one wanted to give a reductive account of possiblity in terms of conceivability, there is one strategy that is worth trying. Instead of invoking possible beings in the definition of conceivability, one could invoke conceivable beings. There might then be a sort of bootstrapping definition. First, the notion of conceivability would be grounded in our own prima facie conceivings. Second, we can conceive of beings who are better reasoners than us, with fewer cognitive limitations. Third, those beings could presumably conceive of better reasoners still. And so on. It is not out of the question that this process might lead to some sort of limit or fixed point. If so, one might obtain a recursive (not noncircular) definition of possibility in terms of conceivability. I cannot pretend that this matter is entirely clear, however.)

Does this account leave room for modal error? If theses (1) and (2) are correct, then modal errors arising from conceivability judgments will stem either from the difference between prima facie and ideal conceivability or from the difference between primary and secondary conceivability (and possibility). These modal errors will fall into one or more of the following classes:

(i) Prima facie negative conceivability judgments can go wrong in cases where a "deep" a priori contradiction is not revealed by prima facie reasoning.

(ii) Prima facie positive conceivability judgments can go wrong when (a) an imagined situation that is taken to verify P does not in fact verify S, upon rational reflection; or when (b) an imagined situation is not coherently imagined, because of the failure to notice a deep contradiction, or because of the inability to fill in crucial details.

(iii) Primary conceivability judgments can go wrong if a subject mistakenly expects them to be a guide to secondary possibility.

(iv) Prima facie secondary conceivability judgments can go wrong as a guide to secondary possibility when a subject is misinformed about relevant nonmodal empirical facts, and perhaps when an incautious

subject is merely ignorant of those facts.

7 From Negative to Positive Conceivability

In the remainder of the paper, I will focus on the status of the central conceivability-possibility theses (1) and (2), discussing what is required in order for them to be true, what form counterexamples must take, whether there are any plausible counterexamples, and what might ground the truth of the theses. Given space limitations, this discussion will only scratch the surface, but I hope to convey at least a broad view of the terrain. In discussing these theses, we can restrict our attention mostly to ideal primary positive conceivability and ideal primary negative conceivability, and to primary possibility. "Ideal primary" should be understood throughout, in references to positive and negative conceivability, and "possibility" should always be read as primary possibility. The speaker-relativity of the relevant claims should also be understood throughout.

I will first address the question of whether (ideal primary) negative conceivability entails (primary) possibility. Here I will factor out the question, addressed later, of whether positive conceivability entails possibility, and address the question of whether negative conceivability entails positive conceivability.

I call the (empty or nonempty) class of statements that are negatively conceivable but not positively conceivable the *twilight zone*. Potential members of the twilight zone come from two sources: inscrutabilities and open inconceivabilities.

8 Inscrutabilities

The class of inscrutabilities can be introduced by considering an attractive thesis about truth and reference.

> **SCRUTABILITY OF TRUTH AND REFERENCE:** Once we know how the world is qualitatively, we are in a position to know what our terms refer to and whether our statements are true.

Take the case of reference first. Often we do not know what our terms refer to, but this knowledge is usually grounded in some qualitative ignorance of the way the world is. Given enough qualitative information (typically information about physical and mental states, although more on this later), we are in a position to know what our terms refer to. This reflects common practice in the theory of reference: in thinking about reference of a term in actual and hypothetical situations (considered as actual), it suffices to give a complete enough qualitative description of relevant features of those situations. From here, reference can be determined.

There are a few difficulties with the thesis of scrutability of reference. The first is that it is not entirely obvious what it means to "know what a term refers to". Presumably this is to be able to give some sort of alternative description of the referent; but just which alternative descriptions qualify? The second is that there may be a degree of indeterminacy in the reference of our terms over and above what is present in the truth-values of our statements. Examples might include terms like "number" and "symphony", or Quine's examples of mass-nouns and count-nouns in Japanese (see Benacerraf 1965, Horgan 1986, and Quine 1961 respectively). In these cases, it seems that there are multiple ways to assign referents to our terms, each of which capture our intuitions about the truth-values of our statements, insofar as these truth-values are determinate. In these cases it is not at all clear that there is a fact of the matter between these assignments of reference.

For both of these reasons it is easier to focus on the scrutability of truth. On the first issue, there is no analogous problem making sense of what it is to know the truth-value of a statement. On the second issue, almost all of the indeterminacies discussed above drop out when it comes to the truth-values of statements. (The exception may be such statements as "the number two is a set of sets", and the like, but now we are at least down to an isolated problem in the metaphysical domain, as opposed to a problem that arises with every use of the word "two".) And most of the intuitive backing behind the scrutability of reference (e.g. that given enough qualitative information, we can know who Jack the Ripper is) is reflected in the scrutability of truth (e.g. that given enough qualitative information, we can know whether Jack the

Ripper was Prince Albert Victor).

The scrutability of truth can be formulated somewhat more precisely as follows:

> **SCRUTABILITY OF TRUTH** (second pass): If D is a complete qualitative description of the world, then for all T, T ⊃ (D implies T).

Here and elsewhere, 'A ⊃ B' is a material conditional, and A implies B iff 'A ⊃ B' is a priori. So the scrutability thesis says, in effect, that all truths are derivable through a priori reasoning from a complete qualitative description of the world. The thesis can also be weakened somewhat to hold that all truths are derivable from a complete *enough* qualitative description, thus avoiding the need to invoke a description of the whole world for every truth, but I will use the simpler if less practical formulation in what follows.

This way of putting things is not only more precise than "Once we know A, we're in a position to know B"; it also overcomes a problem posed by the paradox of knowability. Let P be a truth that I don't currently know, and let Q be "P and I don't know that P". Then Q is true but unknowable. (To know Q, I would have to know P; but once I know P, then Q is false). So Q is in danger of coming out inscrutable on the first formulation — it is a truth such that having full qualitative information about the world doesn't suffice to know it. Nevertheless it may remain the case that Q is implied by a full qualitative description of the world. The "paradox" gives no special reason to deny that given such a complete qualitative description D, I can know a priori that if D, then Q.

We can also put the scrutability thesis in terms of *epistemic completeness*, where an epistemically complete statement is one that, roughly speaking, epistemically settles everything that could be settled. More precisely, let us say as before that a statement D is epistemically possible (in the broad sense) when D is not ruled out a priori: i.e. when it is not a priori that ~D (or that ~det(D), to cover cases of indeterminacy), i.e., when D is ideally negatively conceivable). Then

> A statement D is *epistemically complete* iff (i) D is epistemically possible, and (ii) for all F, if D&F is epistemically possible, then D implies F.

Then the scrutability thesis, reformulated, says that a complete qualitative description of the world is epistemically complete. The second formulation implies the first, since if D is a complete qualitative description of the world and T is the case, then D&T is true, so D&T is epistemically possible, so (by the second formulation) D implies T. In the other direction: if D&F is epistemically possible, then D does not imply ~F, so (by the contrapositive of the first formulation) ~~F, so F, so (by the first formulation) D implies F. (Worries about indeterminacy are handled by the observation that if D&F is epistemically possible, D does not imply ~det(F), so the indeterminacy of F is excluded.)

The residual unclarity, of course, is in the notion of a "complete qualitative description of the world"? What counts as a complete qualitative description? One idea is that a complete enough qualitative description is one that specifies all truths; but this will not do for our purposes, since it renders the scrutability thesis trivial. Intuitively, a qualitative description of the world is a basic description from which many other truths might be derived.

A second and promising idea says that a complete qualitative description is a complete description in terms of fundamental natural properties (plus indexical information). That is, it involves a description in terms of fundamental microphysical properties (perhaps such as mass, charge, position, and spin), and perhaps also in terms of those fundamental properties (if any) that are not microphysical (on some nonmateralist views, phenomenal or protophenomenal properties). So understood, the scrutability thesis would come to the claim that the fundamental natural truth about the world, in conjunction with indexical truths, implies (a priori) all truths.

We might formalize this by understanding this sort of description of the world as an *ontologically complete* description of the world: roughly speaking, one that metaphysically necessitates all truths about the world. In order a resulting scrutability thesis to be tenable, the relevant sort of metaphysical necessitation must be 1-necessitation

(recalling that a statement is 1-necessary if its primary intension is true in all centered metaphysically possible worlds). More precisely, we can say:

> A statement D is *ontologically complete* if (i) D is 1-possible, and (ii) if D&F is 1-possible, then D⊃F is 1-necessary.

The resulting scrutability thesis is:

> **STRONG SCRUTABILITY**: If D is an ontologically complete truth, then D is epistemically complete.

The strong scrutability thesis is interesting and not obviously false, and is closely related to thesis (2) connecting ideal negative conceivability and possibility (in fact it is a consequence of that thesis). But it does not cut things quite finely enough for our purposes here. The thesis would be denied by a materialist who holds that it is positively conceivable that there be zombies (or other worlds physically identical to ours and phenomenally distinct), but that zombies are not metaphysically possible. (I have elsewhere called this view type-B materialism, as opposed to type-A materialism on which zombies are not even conceivable.) According to such a materialist, phenomenal truths about the world are not implied by the complete fundamental truth about the world, which is microphysical, so strong scrutability is false. This sort of denial of strong scrutability raises issues somewhat distinct from those I am concerned with here. For now, I am concerned with potential gaps between negative and positive conceivability, but this denial is best assimilated to a potential gap between positive conceivability and possibility. So it is useful to factor out a weaker scrutability thesis which this denial does not contravene.

The weaker scrutability thesis requires a sense of "complete qualitative description" such that on the type-B materialist view, a microphysical description is not a complete qualitative description. Intuitively, the microphysical description seems incomplete (in a sense) as a qualitative description precisely because it does not specify the phenomenal truths, and the phenomenal truths seem to be (in a sense) qualitative truths. And the sense in which these are qualitative truths seems to correspond to the fact that such truths will be required for a

fully clear and distinct conception of what the world is qualitatively like. That is, they are required for a description of the world to the limits of positive conceivability.

A qualitatively complete description of the world, then, should be understood as a description to the limits of positive conceivability. That is, it is a description which specifies a unique positively conceivable situation. We can define this more precisely as follows:

> A statement D is *qualitatively complete* if (i) D is positively conceivable, and (ii) if D&F is positively conceivable, then D implies F.

On the type-B materialist view, a complete microphysical description of the world will not be qualitatively complete. For various phenomenal truths F, D&F will be positively conceivable, as will D&~F, but D will imply neither F nor ~F. On this view, a complete qualitative description of the world will require at least something akin to a full microphysical and phenomenal description.

With the notion of qualitative completeness in hand, we can now formulate our final version of the scrutability thesis.

> **SCRUTABILITY**: If D is a qualitatively complete truth, then for all S, S ⊃ (D implies S).

Or equivalently:

> If D is a qualitatively complete truth, then D is epistemically complete.

We can say that S is an *inscrutable truth* if S is true and some qualitatively complete truth D does not imply S. The scrutability thesis above says that there are no inscrutable truths.

It is easy to see that if S is inscrutable, then both D&S and D&~S (for a relevant D) are in the twilight zone. D&S and D&~S are negatively conceivable since their negations (D⊃~S, D⊃S) are not a priori. But they are not positively conceivable: if they were, then D could not be qualitatively complete. So if there are inscrutable truths, there are

inhabitants of the twilight zone, and negative conceivability implies neither positive conceivability nor possibility.

Are there any inscrutable truths? To assess this, it helps to have an idea of what a qualitatively complete truth involves. Such a truth presumably must include at least full microphysical information, including microphysical laws. It may or may not require phenomenal information (a type-A materialist view will deny this), but it cannot hurt to include it (specified in the sort of "pure phenomenal" vocabulary discussed by Chalmers 2002). On any of the type-A materialist, type-B materialist, and property dualist views, a qualitatively complete truth will imply the complete microphysical and phenomenal truth, so anything not implied by the former will not be implied by the latter. Indexical information is required, since more than one conjunction of complete objective truths with indexical claims will be positively conceivable. Finally, in order to rule out situations containing extra nonphysical, nonphenomenal goings-on, qualitative completeness requires a "totality" claim, holding that the world is a minimal world that satisfies the physical, phenomenal, and indexical claims specify.

Elsewhere (Chalmers and Jackson 2001), this conjunction of microphysical, phenomenal, indexical, and totality claims is referred to as PQTI. It is not entirely implausible that PQTI is itself a complete qualitative description of the world: for the most part, even where there are candidates for truths not implied by PQTI, these do not seem to be associated with intuitions of distinct positively conceivable situations analogous to the intuitions in the zombie case. In any case, even if PQTI is not itself qualitatively complete, we can at least say that any inscrutable truth will be a truth not implied by PQTI.

Possible candidates might fall into a number of classes.

(i) Ordinary macroscopic truths. One might first question whether ordinary macroscopic truths about the natural world, such as "grass is green" and "there is water in my pool" can be derived by a priori reasoning from PQTI. I have argued elsewhere (Chalmers 1996; Chalmers and Jackson 2001) that they can be, and I will not repeat that case here. But the basic idea is that straightforward a priori reasoning from PQTI puts one in a position to know all about the physical

composition, the phenomenal appearance, the spatial structure, and the dynamic behavior of macroscopic systems, along with facts about their relation to oneself and their distribution in space and time; and this information in turns puts one in a position to know all ordinary macroscopic truths S about such systems, as long as one possesses the concepts involved in S. The information will include all the information on which ordinary perceptual or theoretical knowledge that S might be based, along with sufficient information to conclusively rule out skeptical counterpossibilities to S. If so, it is very plausible that PQTI implies S.

One worry arises with names and natural kind terms: someone might object that truths involving these terms cannot be a priori entailed by PQTI, as the relevant a priori connections are not built into the semantic content of these terms. In response, recall that we are working with a speaker-relative conception of apriority and primary intension. It may be in cases such as 'Neptune' or even 'water', the primary intension and a priori connections of a term varies between speakers, so that if "semantic content" must be common to all speakers, primary intensions and a priori connections are not determined by semantic content. But all that is required here is that the conditional "PQTI⊃S" be a priori for any given speaker. This thesis is quite compatible with the variation in primary intension, and it can be argued for straightforwardly along the lines of the previous paragraph.

(It might be thought that Kripke's epistemological arguments tell against even speaker-relative a priori entailments, but it is easy to see they have no power against the sort of specific entailment at issue here. At most, they suggest that a term's primary intension cannot be captured by a description. Special issues come up for expressions that are used with semantic deference, as when a speaker defers to other speakers in fixing a term's reference. I think that even these expressions can be accommodated on this framework, but for present purposes it is easiest to stipulate that we are concerned only with nondeferential uses.)

(ii) Certain mathematical truths. Someone might suggest that there are true mathematical statements that are not a priori, i.e. that are not

knowable even on ideal rational reflection. For example, one might suppose that certain Gödelian statements in arithmetic (the Gödel sentence of the finite human brain?), or certain statements of higher set theory (the continuum hypothesis or its negation?) may be determinately true without being ideally knowable. If such truths exist, they will plausibly not be implied by a qualitatively complete description of the world, so they will be inscrutable.

However, it is not at all clear that such statements exist. In any given case, one can argue that either the statements in question are knowable under some idealization of rational reasoning, or that the statements are not determinately true or false. In the arithmetical case, one can argue that for any statement S there is some better reasoner than us that could know S a priori. Our inability to know a given Gödel sentence plausibly results from a contingent cognitive limitation: perhaps our limitations in the ordinal counting required for repeated Gödelization (which can be shown to settle all truths of arithmetic), or even our contingent inability to evaluate a predicate of all integers simultaneously (Russell's "mere medical impossibility"). In the case of unprovable statements of set theory, it is not at all clear that truth or falsity is determinate. Most set theorists seem to hold that the relevant cases are indeterminate (although see Lavine forthcoming for an argument for determinacy); and even if they are determinate in some cases, it is not out of the question that possible beings could know the truth of further axioms that settle the determinate truths.

There is more to say about this issue. I think that the mathematical case is the most significant challenge to scrutability, and even if it fails, it clearly raises important questions about just what sorts of idealizations are allowed in our rational notions. For now, however, it suffices to note that there is no strong positive reason to hold that cases of mathematical determinacy without apriority exist.

(iii) Vague statements.

It is plausible that some statements involving vague predicates (e.g., "person X is bald") cannot be known to be true or false, even given complete qualitative information. Complete qualitative information will tell us how much hair a person has, but may leave the question of

tallness unsettled. On the standard view of vagueness, this will not be a case of inscrutability, since the statements themselves will be neither true nor false. On the epistemic theory of vagueness, however, such statements are determinate even if we cannot know their truth-values: there is a precise border between the bald and the nonbald, but we cannot know where it is. Some versions of the epistemic theory may hold that this is due to rational limitations on our part (so that more intelligent creatures might be able to locate the border between baldness and nonbaldness); but we can consider a version on which even this epistemic connection fails. On such a view, some statements involving vague predicates will be inscrutable truths.

For example, let us assume for simplicity that baldness supervenes on number of hairs. On this version of the epistemic theory, some truths of the form "X is bald" will not be implied by truths of the form "X has n hairs". Here, "X has n hairs and is not bald" will be ideally negatively conceivable, but impossible. Here, the two statements "X has n hairs and is not bald" nor "X has n hairs and is bald" are negatively but not positively conceivable. When we consider any imagined situation in the vicinity — one in which X has n hairs and has further qualitative properties that are inessential here — it verifies neither "X is bald" not "X is not bald". So these statements fall into the twilight zone.

(Contrast the zombie case in the philosophy of mind, where it seems that "physical structure P and conscious" and "physical structure P and not conscious" are each positively conceivable, verified by two different modally imaginable situations. In the baldness case, and other cases of vagueness, there are no two such distinct modally imaginable situations: at best, there are two coherently entertainable descriptions. So in these cases, unlike the zombie case, there is no call to include information about baldness explicitly in a qualitatively complete description of the world; the existing description was already qualitatively complete, at least as far as the matters here are concerned. So the truths about baldness (on the epistemic theory) fall into the gap between negative and positive conceivability (yielding inscrutability), whereas truths about consciousness do not.)

Of course all this is contingent on the truth of the epistemic theory of vagueness, and the epistemic theory is widely regarded as very

implausible. In fact one might trace the implausibility of the epistemic theory at least in part to the way it denies inscrutability. In these cases, it seems that a subject has all the qualitative information that could possibly be relevant, and it seems almost obvious that given that information, the subject is in a position to know all there is to know about baldness here. So it might be argued that the intuitive implausibility of the epistemic theory is grounded in an intuitive endorsement of scrutability, at least in this domain.

One might worry about the status of vague statements even when the epistemic theory is rejected. Here the answer depends on one's view of vagueness. If one accepts the law of the excluded middle (Sv~S), then one must also accept (PQTI implies S)v(PQTI implies ~S). If one rejects the law of the excluded middle (the best option, on my view), then one will reject the corresponding claim about implication. If there are determinacy operators "det" and "indet", and cases in which indet(S), then PQTI must imply indet(S). But I do not think there are any fatal problems for scrutability here, over and above the problems that arise in analyzing vagueness in general.

(iv) Moral claims. Many philosophers hold that the truth or falsity of moral claims, such as "eating animals is bad", is not determined a priori by natural truths. This can be argued by a generalization of Moore's open question argument, suggesting that two people possessing full natural information might disagree on the truth of a claim like this, without either displaying incapacities of reasoning or failing to grasp moral concepts. This view is often combined with the view that moral claims are not strictly truth or false at all; but some philosophers holds that moral claims are true or false despite being epistemically underdetermined by the natural truth in this way. On this view, moral truths are not implied by PQTI. As in the case of vagueness, there do not seem to be distinct positively conceivable situations verifying PQTI&M and PQTI&~M, where M is a moral claim. If so, then moral truths are inscrutable truths.

As in the case of vagueness, this view of morality is controversial, so it certainly does not provide a clear case for inscrutable truths. Proponents of this sort of view often argue for the view by appealing to Kripke's distinction between the a priori and the necessary, but there

are strong disanalogies: Kripke's cases are compatible with an entailment from (negative) conceivability to possibility, and with the scrutability of truth, whereas this view is not. And I think that there are good arguments against the view, based on considerations related to scrutability (Horgan and Tienson 19xx give some related arguments). But in any case, the view helps to illustrate what an inscrutable truth might be.

(A related issue: Yablo (this volume) worries that in the moral case, given the nonapriority of PQTI⊃M and PQTI⊃~M and a principle connecting nonapriority of ~S with primary possibility of S, one could infer the primary possibility of both PQTI&M and PQTI&~M, which seems wrong. In response: if one denies the apriority of the conditionals and also rules out the position above, then the most plausible position remaining is that moral claims are not strictly true or false but are indeterminate; and it is plausible that if this view is true, it is a priori. If so, then it is a priori that ~det(PQTI&M) and that ~det(PQTI&~M). If so, then (given the official characterization of ideal negative conceivability) PQTI&M and PQTI&~M are not ideally negatively conceivable, so they will not be primarily possible.)

(v) Metaphysical claims. Many issues within philosophy are such that it is not obvious that they can be conclusively settled by rationally reasoning from the information in PQTI. This applies especially to questions at issue within metaphysics: Do mereological sums exist? Do all dispositions have categorical bases? Are properties universals or tropes? What is required for identity over time? Do numbers exist? Is an A-theory or a B-theory of time correct? And even: does conceivability suffice for possibility?

There is no space here to consider these issues separately. But in general, an advocate of scrutability can take one of three strategies any given one of these issues. (1) Argue that sufficient rational reflection — perhaps more than has been done to date — can conclusively settle the issue (perhaps the issues about numbers and conceivability fall here). (2) Argue that the issue is positively conceivable either way, so that PQTI needs to be supplemented by further information to yield a qualitatively complete description of the world (perhaps the issues about categorical bases and about time fall

here). (3) Argue that there is no fact of the matter about the issue, or that it can only be settled by terminological refinement (perhaps the issues about mereological sums and identity over time fall here). In each case, there is room for argument, but it certainly seems that there are no clear cases where all three of these strategies fail.

Overall: It seems that there is no clear counterexample to scrutability. At best there are some unclear potential counterexamples, none of which carries enormous antecedent plausibility, although some deserve further investigation.

Stepping back for a moment, why should we accept the scrutability thesis? One way to argue for it is to suggest that any reasonable candidate for an inscrutable truth will be an unknown and unknowable truth, since what we know is limited to information gained through perception (and so present in or implied by PQTI), plus that derived from rational reflection (and so implied by PQTI). One can also argue that all unknown truths stem from either ignorance of the qualitative nature of the world, or from insufficient a priori reasoning; and that the only unknowable truths stem from ignorance of the qualitative nature of the world. If this is so in general, then there are no inscrutable truths.

None of this yields a knockdown argument, but it does give reason to take the scrutability thesis very seriously.

It is useful to generalize the scrutability thesis slightly, so that it applies not only to complete qualitative descriptions of the actual world and to actual truths, but to any complete qualitative descriptions and to any truths. The generalized thesis is that a complete enough qualitative description of *any* world leaves no truth about that world epistemically open. After all, it would be odd if scrutability turned out to be true in this world but not in others; the thesis seems to have a much more general source than that.

> **GENERALIZED SCRUTABILITY:** If D is qualitatively complete, then D is epistemically complete.

Clearly generalized scrutability implies scrutability: the earlier scrutability thesis is the special case where D is a qualitatively

complete *truth*. Scrutability does not logically imply generalized scrutability, but it is natural to think that if scrutability is true, generalized scrutability is probably true. If scrutability holds, it seems unlikely that it holds accidentally, because of the character of the actual world. Rather, its truth would seem to reflect something deep about concepts, truth, and reason. If this is right, then the two theses are likely to stand and fall together.

9 Sideline: Modal Rationalism and Logical Empiricism

As Yablo (this volume) notes, the scrutability thesis has something in common with some versions of logical empiricism. A common logical empiricist thesis was that phenomenal truths analytically entail all truths. The main differences between the theses are (1) in appealing to a qualitatively complete truth, the scrutability thesis allows much more in its entailment base than just phenomenal truths (in particular, it allows the complete microphysical truth); and (2) scrutability as I have characterized appeals not to analyticity but to apriority, which I think is a more basic notion.

Yablo suggests that the problems of logical empiricism may infect the modal rationalism that I advocate. The problem on which Yablo focuses is the underdetermination of theory by evidence. Underdetermination of theory by *local* evidence is no problem for a sufficiently holistic logical empiricist, but underdetermination of theory by *total* evidence is a real problem. It seems easily conceivable that different states-of-affairs might provide the same evidence, or that some truths might leave no trace on our evidence; and there are even pairs of real-life scientific theories (as in quantum mechanics) that save all appearances while making different microphysical claims. These considerations are all tied to the limitations of observation, however: they suggest that phenomenal truths or observational truths underdetermine theory. They do nothing at all to suggest that the complete qualitative truth (including microphysical truths) underdetermines theory. So the most common worries about underdetermination do nothing to threaten the scrutability thesis.

To make a stronger parallel argument, Yablo appeals to a different worry, about the role played by a posteriori considerations of

"reasonableness" and "sensibility" in moving from evidence to theory. There seem to be a number of separate arguments here, although I am not certain that Yablo intended all of them.

First: considerations of reasonableness cannot be reduced to a set of a priori rules (as in "the dream of an a priori inductive logic"). Response: there is no reason to suppose that a priori truths, or a priori entailments, must be reducible to some basic set of explicit formal principles. So this is a red herring. (See Chalmers and Jackson 2001 for more on this.) At best, this point might affect a thesis cast in terms of analyticity, if analyticity is defined in terms of logic plus definitions.

Second: the modal rationalist requires too much of our "grasp of meaning", by requiring knowledge of the relevant conditionals. Response: On the rationalist view, knowledge of the relevant conditionals need not be built into grasp of meaning, and need not be possessed by every subject who possesses the relevant concpts. Where present, the knowledge is usually a product of substantive reasoning, grounded both in possession of the relevant concepts and in rational reflection.

Third: there are cases where rational reflection on qualitative information underdetermines theoretical truth, which is settled only by pragmatic factors. Response: if the pragmatic factors are rationally underdetermined, then these are cases where originally indeterminate statements become determinately true or false because of terminological evolution. (See Chalmers and Jackson 2001.) So there is no point in the process at which there are inscrutable truths.

Fourth: the inference from PQTI to macroscopic truths may depend on "peeking", as when one perceptually imagines the appearance of a situation. Response: This point might threaten a scrutability thesis based on PTI, which excludes information about appearances. But given the phenomenal information Q about appearances (in a pure phenomenal vocabulary), peeking comes for free. Knowledge of PQTI yields knowledge of the phenomenology of the appearances, and this puts one in as good a position to reason from those appearances to macroscopic truths as if one had experienced the appearances directly. (See also the further discussion below.)

Fifth: the general exercise of "sensibility" is a posteriori, since it involves introspective knowledge of concept application in one's own mind. Response: No introspective knowledge is needed to know the entailment from PQTI to S. One merely needs to deploy the concepts involved in S; one does not need to observe their deployment.

There is more to say here, but it seems that at least on a first look, the scrutability thesis is unthreatened by its parallels with logical empiricism. Still, the parallels certainly exist, both with logical empiricism and other broadly phenomenalist and anti-realist views. These anti-realist views hold (in a sense) that when it comes to truth, nothing is hidden. The scrutability thesis does not suggest this, and is perfectly compatible with a realist view. But it suggests a weaker thesis: *given* complete qualitative knowledge (and ideal reflection), nothing is hidden. There is perhaps a tiny residue of anti-realism here: if a claim cannot be settled by a priori reasoning on the basic qualitative facts, then there is no fact of the matter about that claim. But one does not have to be a logical positivist to find this thesis attractive. (After all, there are some things that it is perfectly reasonable to be anti-realist about.) Indeed, in almost all cases where a claim cannot be settled in this way, as we have seen, it is independently plausible that the claim is indeterminate. So the scrutability thesis may well embody a principle that is tacit in our reasoning.

10 Open Inconceivabilities

The second potential source of a gap between negative and positive conceivability arises from states of affairs that are *inconceivable*, but that are nevertheless not ruled out a priori.

There are quite likely many prima facie inconceivabilities: a rich source is provided by statements about phenomenal properties quite distinct from our own. For example, the claim that there are creatures with 12-dimensional phenomenal color spaces cannot be ruled out a priori, but it may be beyond our capacity to conceive of a situation verifying this claim. Such a conception might require phenomenal concepts (and ultimately phenomenal experiences to ground those concepts) that we simply lack. If so, such a claim is prima facie

negatively conceivable, but not prima facie positively conceivable. This is not obviously a case of ideal inconceivability, however. We have already seen that the inconceivability here stems from a lack in our repertoire of phenomenal concepts, and this limitation is contingent. If we idealize away from this conceptual lack, then the situation in question will plausibly turn out to be conceivable after all. Presumably there are possible creatures with the relevant concepts, and such creatures would have no difficulty in conceiving of the situations in question.

Still, perhaps there are some features of the world, or of some world, that simply cannot be positively conceived at all. One example might be provided by intrinsic properties that are not phenomenal properties, and are not conceptually related to them. One might argue that the only way to form a conception of an intrinsic property is by direct acquaintance, as in the phenomenal case, or perhaps by a priori reasoning from concepts of intrinsic properties one has direct acquaintance with; think of the missing shade of phenomenal blue. (Of course one might form an extrinsic conception of an intrinsic property, such as "the property that is causally responsible for such-and-such", but this is not good enough here, as such a conception leaves open multiple epistemic possibilities as to the nature of the property.) And one might argue that the only intrinsic properties any subject can be directly acquainted with are phenomenal properties. If so, then any intrinsic properties that are not phenomenal properties will be in the relevant sense inconceivable.

Of course all the assumptions going into the case above are highly contestable, but the possibility of inconceivable features of the world does not seem easy to rule out. This can be exploited to yield perhaps the most plausible example of an open inconceivability: namely, "there are inconceivable features of the world". This statement is by its nature verified by no positively conceivable situation, but it is also not easy to rule out a priori. Unless some way can be found to rule out this statement a priori, it will be (ideally) negatively conceivable but not positively conceivable, and hence will be in the twilight zone.

More precisely:

> S is an *open inconceivability* if S is negatively conceivable, but for all qualitatively complete D, D implies ~S.

(Note: to handle indeterminacies, the last clause should hold that for all D, D implies ~det(S). If S is negatively conceivable, det(S) is not ruled out a priori; if nevertheless for all D, D implies that S is indeterminate, then S should be an open inconceivability.)

We have seen that "There are no inconceivable features of the world" is one potential open inconceivability. Another is "There is no qualitatively complete description of the world". At a more specific level, the case of nonphenomenal intrinsic properties, on the assumptions above, will provide examples of open inconceivabilities insofar as there are ways to express relevant inconceivable truths (e.g., "There are nonphenomenal intrinsic properties", if nothing else). Still, none of these yield clear cases, so we can at least formulate a relevant thesis opposed to them:

> **NOINCONCEIVABILITY:** No S is an open inconceivability.

It is not clear how best to argue for this thesis. One might argue for any property, there is some creature than can form a conception of it — perhaps any intrinsic property can be known by acquaintance, and any non-intrinsic property by description. And one might argue that this principle is itself a priori. If that is so, then it plausibly follows that there are no open inconceivabilities. But the central claim here is far from obvious.

Like inscrutabilities, open inconceivabilities (if they exist) provide a gap between negative and positive conceivability. They differ, however, in that they may not provide a gap between negative conceivability and possibility. That gap depends on whether the open inconceivabilities in question correspond to real possibilities (e.g. properties we can't form a conception of), or whether they correspond to impossibilities that we cannot rule out a priori (perhaps all properties are conceivable, but we can't rule out the alternative a priori). If all open inconceivabilities are of the former sort, then negative conceivability might still be a guide to possibility — it is just that the possible will outstrip the positively conceivable. If some are of the latter sort, then negative conceivability

will be an imperfect guide to possibility.

11 The Structure of the Twilight Zone

We have seen that potential members of the twilight zone stem from at least two classes: inscrutabilities and open inconceivabilities. In fact is is not hard to see that all members of the twilight zone stem from these two classes.

> **NEGPOS:** Ideal negative conceivability entails ideal primary positive conceivability.

Then it is not hard to demonstrate the following thesis: Negpos is true iff generalized scrutability and noinconceivability are true."

Proof: Left-to-right: Given negpos, any qualitatively complete statement will be epistemically complete, so generalized scrutability will be true. Given negpos, any negatively conceivable statement will be positively conceivable, so will be entailed by some qualitatively complete D, so noinconceivability will be true. Note that the second part requires the principle that any positively conceivable statement is implied by some qualitatively complete statement. This seems reasonable, as it encapsulates the idea that a statement verified only by "uncompletable" situations will not be ideally positively conceivable.

Right-to-left: Let S be negatively conceivable. Noinconceivability implies that there is a D such that ~(D implies ~det(S))). Generalized scrutability implies that D settles S's truth-value, so D implies S. (Note that the determinacy operator in the definition of open inconceivability is needed here, in order to exclude the possibility that D implies neither S nor ~S, due to indeterminacy.)

So in order to close a potential gap between (ideal primary) negative and positive conceivability, it is necessary and sufficient to rule out generalized inscrutabilities and open inconceivabilities.

In order to close a potential gap between negative conceivability and possibility, it is necessary to rule out generalized inscrutabilities (if S is a generalized inscrutability, then both D⊃S and D⊃~S will be negatively conceivable for a relevant D, but both cannot be possible). It

is not necessary to rule out all open inconceivabilities, but one must rule out all *impossible* open inconceivabilities. If some open inconceivabilities are also impossibilities, then negative conceivability does not entail possibility. But if all open inconceivabilities are possibilities (which is not entirely implausible), then an entailment between negative conceivability and possibility is not threatened.

12 From Positive Conceivability to Possibility

Does (ideal primary) positive conceivability imply (primary) possibility? A counterexample to this principle must involve what I have elsewhere called a *strong necessity*: a statement that is falsified by some positively conceivable situation (considered as actual), but which nevertheless true in all possible worlds (considered as actual). For such necessities to exist, the space of positively conceivable situations must outstrip the space of possible worlds.

There are certainly no clear examples of strong necessities, and the only candidates are highly tendentious. I have discussed this matter at some length elsewhere (Chalmers 1999), so I will say only a little about some possible candidates here.

(i) *The existence of God*. On many theist views, a god exists necessarily, so that every possible world contains a god. But a theist may hold that it is not a priori that a god exists, and that a godless world is positively conceivable, even on rational reflection. On this view, it is natural to hold that "no god exists" is primarily positively conceivable, but not primarily possible. So if this view is correct, then "a god exists" is a strong necessity. Of course, the theist thesis here is highly controversial, so this is not a strong counterexample, but it illustrates what a counterexample must involve.

One further worry: if god does not exist necessarily, then "A necessary god exists" is impossible. But it may seem that a necessary god is at least conceivable (see Yablo 1999). In response, I deny that a necessarily existing god is ideally conceivable. A god's existence may be conceivable, but to conceive of a god's necessary existence is much harder, especially given its conceivable nonexistence. In effect, one must conceive (metamodally!) that conceivability does not imply

possibility. But it is not clear that this is more than prima facie negatively conceivable. On my view, it is a priori, if non-obvious, that conceivability entails possibility (see below for the sketch of an a priori argument). If so, then the denial of the entailment is not ideally conceivable, and so neither is the necessary existence of a god.

(2) *Laws of nature*. Some philosophers hold that the laws of nature are metaphysically necessary. On some views of this sort (e.g. those discussed by Fine and Sidelle in this volume), this necessity arises for broadly Kripkean reasons: the reference of terms such as "mass" is fixed a posteriori to a certain very specific property, so that worlds with different laws do not contain mass. I think this view is implausible, but in any case it is compatible with an entailment from primary conceivability to primary possibility. If G' is a counternomic statement (say, an adjusted statement of gravitational laws with a different constant), then G' is both primarily conceivable and primarily possible. G' is verified by a metaphysically possible world W considered as actual, although not by W considered as counterfactual. (Considered as counterfactual, W contains "schmass", not mass.) So there are no strong necessities here.

There is a stronger view on which the laws of every world are are exhausted by actual-world laws, applying to actual-world properties. On this sort of view, even "schmass" worlds are metaphysically impossible: G' will be primarily conceivable but not even primarily possible. On this view, laws of nature are strong necessities. There is no reason to accept this view, however. (It is notable that Fine and Sidelle quickly dismiss such a view as too extreme to be plausible.) Proponents of necessary laws usually appeal to Kripke's necessary a posteriori for support, but the Kripkean cases support at best the weak view in the previous paragraph. Nothing here gives reason to suppose that worlds with different laws are impossible; at best, it suggests that they are misdescribed as breaking our laws. So there is no good reason here to deny the conceivability-possibility thesis.

(3) *Response-enabled concepts*. Yablo (this volume) considers a class of *response-enabled* concepts whose extension, and whose meaning, is fixed by our responses. He suggests that "oval" is in this class: the reference of "oval" is fixed by picking out whatever looks oval-shaped

to us, irrespective of any pure geometric description of their shape. I think the example is imperfect, since "oval" is arguably a pure geometric concept, picking out certain geometric shapes regardless of the responses they cause in us. But there may be other terms that function as Yablo suggests, so I will play along with his suggestion that "oval" works this way.

Following Yablo, let "cassini" be a term for a certain class of mathematically defined geometric figures, of a sort that actually cause "oval"-responses. Then (given Yablo's view of "oval"), it is not a priori that cassinis are ovals. So "cassinis are not ovals" is ideally negatively conceivable. "Cassinis are not ovals" is also plausibly ideally positively conceivable, since it is verified by a situation in which cassinis are not the sort of object that cause "oval"-responses. But Yablo suggests that "cassinis are ovals" is nevertheless true in all worlds considered as actual: in all such worlds, cassinis fall under the extension of "oval". If this is correct, then ideal (negative or positive) primary conceivability does not entail primary possibility.

In response: As I have characterized considering-as-actual, "cassinis are not ovals" is true of some worlds considered as actual. Let W be a world in which cassinis do not cause oval-responses. Let us grant that it is a priori that if W is actual, cassinis do not cause oval-responses. (Yablo raises no objection to this.) We can also note that if "oval" functions as described, then a material conditional such as "if Hs do not cause oval responses, then Hs are not ovals" is pretty clearly a priori. (Yablo himself allows that there may be an apriori connection here.) It follows that the material conditional "if W is actual, cassinis are not ovals" is a priori. So as I have defined things, "cassinis are not ovals" is true in W considered as actual, and is primarily possible.

Why does Yablo resist this straightforward conclusion? It seems to me that he is operating with a different conception of how statements are evaluated in considering a world as actual, one tied to the "if it had turned out" locution and to certain claims about "conceptual necessity". I am not sure that I fully grasp this conception, but for present purposes I need not deny that it is coherent or that it captures some feature of our concepts. But it is clearly distinct from the conception I am operating with, on which considering as actual

involves a priori reasoning about epistemic possibilities. Yablo gives no reason to deny that this conception is coherent, or that it yields the results I have suggested. So the conceivability-possibility link that I have advocated is unthreatened by Yablo's discussion.

Something similar applies to other claims that Yablo discusses, including "unless we are greatly misled about the circumstances of visual perception, what looks green is green". If this (or something like it) is a priori, then like all a priori statements, it will automatically hold in all worlds considered as actual, at least on my conception of considering as actual (though perhaps not on Yablo's). And the statement "Fs are not red", where "F" involves a complete intrinsic characterization of something that is actually red, will be a posteriori, primarily positively conceivable, and primarily possible on my conception: it will be straightforwardly true in a world (considered as actual) where Fs do not look red.

At one point Yablo raises another worry about response-dependent concepts: physical and phenomenal truths may not imply truths about yellowness (say), since an characterization of the relevant phenomenology may not enable one to identify the "yellow" responses a priori. I think this is not a problem. The relevant responses are phenomenal kinds, characterized by what it is like to have them. In particular, knowledge of what it is like to experience an object (in normal circumstances) enables knowledge of whether the object is yellow, with no further empirical justification required. Further, physical and phenomenal knowledge enable knowledge of what it is like to experience the relevant objects (in normal circumstances), with no further empirical justification required. It follows that the physical and phenomenal truths imply the truths about yellowness.

(4) *Psychophysical laws*. A final example is given by some type-B materialist views, on which there is an epistemic gap between the physical and the phenomenal, but no ontological gap. On such a view, zombies (and the like) are positively conceivable but not possible. Type-B materialists often appeal to Kripkean cases for support, but it is not hard to see that these do not help, since those cases are compatible with the primary conceivability-possibility link, and the mere primary possibility of zombies causes problems for materialism.

In response, some type-B materialists deny that zombies are even primarily possible. On such a view, psychophysical laws (of the form "if P, then Q" for physical P and phenomenal Q) are strong necessities.

Again, this view is highly controversial, so it does not provide any clear counterexample. This view is usually put forward on the grounds that it is the only tenable way to preserve materialism, given the epistemic gap; but of course that falls well short of a positive argument for the view, especially when the truth of materialism is at issue. Indeed, one can suggest that the conceivability-possibility link that holds elsewhere itself provides a strong argument against this view. Given the discussion above, it seems that the strong necessities required here will be unique. (Even if one thinks that God and laws of nature provide partners in crime, it is notable that the sort of strong necessity at issue there cannot save materialism: in those cases, strong necessities connect ontologically distinct existences!) Some type-B materialists (e.g. Loar 1997, 1999, and Hill 1998) have bitten this bullet and tried to give an explanation of why strong necessities should uniquely arise in the phenomenal domain. I have argued elsewhere (Chalmers 1999) that these explanations fail.

In summary: in each case, the claim that there are strong necessities rests on very controversial assumptions. One might more plausibly argue in reverse: in each of these cases, the elsewhere unbroken link between conceivability and possibility provides an argument against the assumptions in question. In any case, there are no clear counterexamples to the conceivability-possibility thesis here.

Still, all this at best makes a negative case for the conceivability-possibility thesis, by defeating potential counterexamples and explanations. It remains to make a positive case for the thesis, giving reasons why we should expect it to be true. I make a start on this case in Chalmers (1999). I hope to expand on this further elsewhere, but here I will just recapitulate the case briefly.

The argument involves locating the roots of our modal concepts in the rational domain. When one looks at the purposes to which modality is put (e.g. in the first chapter of Lewis 1986), it is striking that many of these purposes are tied closely to the rational and the psychological:

analyzing the contents of thoughts and the semantics of language, giving an account of counterfactual thought, analyzing rational inference. It can be argued that for a concept of possibility and necessity to be truly useful in analyzing these domains, it must be a *rational* modal concept, tied constitutively to consistency, rational inference, or conceivability.

It is not difficult to argue that even if not all conceivable worlds are metaphysically possible worlds, we *need* a rational modal concept tied to rational consistency or conceivability to best analyze the phenomena in question. We might call the corresonding notion of possibility *logical* possibility. For example, even if all worlds with different laws of nature are metaphysically impossible, it will still be tremendously useful to have a wider space of logically possible worlds (or world-like entities) with different laws, to help analyze and explain the hypotheses and inferences of a scientist investigating the laws of nature. Such a scientist will be considering all sorts of rationally coherent possibilities involving different laws; she will make conditional claims and engage in counterfactual thinking about these possibilities; and she may have terms and concepts that are co-extensive at all worlds with our laws, but that intuitively differ in meaning because they come apart at worlds with different laws. To analyze these phenomena, the wider space of worlds is needed to play the role that possible worlds usually play.

Further, there is no bar to a space of such worlds. If one does not want simply to postulate them, one can easily construct them in an 'ersatz' way. For example, one can identify them with equivalence classes of qualitatively complete (or epistemically complete) descriptions (for such a construction, see Chalmers forthcoming c). One can then introduce means of semantically evaluating expressions at these worlds, on both epistemic and subjunctive dimensions. The worlds and the semantic evaluation are perfectly well-behaved, yielding a modal space that is useful for all sorts of purposes. (If one has qualms about using the term 'world' for these entities, nothing turns on the word: one can equally call them 'scenarios', or some such, instead.)"

One can then argue that this space of worlds suffices to account for all modal phenomena that we have reason to believe in. Such a space will

analyze such rational and psychological matters as counterfactual thought, rational inference, and the contents of thought and language as well as any other modal space can. And with the help of two-dimensional semantic evaluation, it can accommodate such "metaphysical" modal phenomena as the concept/property distinction, a posteriori necessities, and so on. These phenomena emerge directly from two-dimensional semantic evaluation over a single space of worlds. The two-dimensional semantics in question will be grounded in a priori conceptual analysis plus nonmodal facts about the actual world. (The first dimension is grounded straightforwardly in a priori conditionals. The second dimension is grounded in a priori conditionals, such as 'if water is H_2O, it is necessary that water is H_2O', plus empirical nonmodal facts, such as 'water is H_2O'.) So one modal space plus conceptual analysis plus nonmodal facts gives us everything, as long as this modal space is tied constitutively to the rational domain.

If this modal space is all, we have *modal monism*, with a single modal primitive. The believer in strong necessities, by contrast, must embrace a *modal dualism*, with distinct primitive modalities of logical and metaphysical possibility, neither of which is reducible to the other. There is no good reason to accept such a modal dualism, when modal monism can explain all the untendentious phenomena. There are no further modal data for a distinct metaphysical modality to explain: what needs to be explained is already explained. This is not just a simplicity argument: One can argue further that there is no distinct *concept* of metaphysical possibility for the second modality to answer to. The momentary impression of such a concept stems from a confused understanding of such ontic/epistemic distinctions such as that between apriority and necessity, and that between concept and property, all of which are easily subsumed under a modal monism with the help of some two-dimensional semantics.

Ultimately, there is just one circle of modal concepts, including both the rational modal concepts (validity, rational entailment, a priority, conceivability) and the metaphysical modal concepts (possibility, necessity, property). The result we are left with is *modal rationalism* in more senses than one: a priori access to modality, and constitutive ties between the modal and rational domains.

CONCLUSIONS

We can sum up the lay of the land by labeling some varieties of modal rationalism:

> **WEAK MODAL RATIONALISM:** (Ideal primary) positive conceivability entails (primary) possibility.
>
> **STRONG MODAL RATIONALISM:** Negative conceivability entails possibility.
>
> **PURE MODAL RATIONALISM:** Positive conceivability $<\supset$ negative conceivability $<\supset$ possibility.

Then pure modal rationalism is equivalent to the conjunction of weak modal rationalism with negpos. The left-to-right direction here is obvious, and the right-to-left direction follows from the observation that possibility entails negative conceivability (no primary possibility is ruled out a priori). Combining this with the previous result about the nature of the twilight zone, it follows that pure modal rationalism is equivalent to the conjunction of weak modal rationalism, generalized scrutability, and noinconceivability.

It follows that to establish pure modal rationalism, we must rule out strong necessities, generalized inscrutabilities, and open inconceivabilities. Here I am most confident about the first, reasonably confident about the second, and unsure about the third. I have outlined a case against strong necessities here, and given tentative reasons to be doubtful about inscrutabilities, while the status of open inconceivabilities is unclear. In any case, it seems to me that each of these three is a distinct and substantial philosophical project, and that the investigation of each raises deep philosophical questions and promises significant philosophical rewards.

If weak modal rationalism is the best we can establish, then we will have done enough to support conceivability arguments as traditionally used, although the overall picture of modality and of modal epistemology will remain somewhat messy. We will have distinct notions of positive and negative conceivability, and thus a mild dualism within the rational modal sphere (though it will be a dualism that is

forced on us by the phenomena). If there are generalized inscrutabilities, then although conceivability will guarantee access to a possible world, it may not yield access to all truths in that world. And if there are open inconceivabilities, there will be worlds that conceivability offers no access to.

Pure modal rationalism yields a simpler picture of modal space, and a correspondingly elegant epistemology. Looking at its three components in turn: the first says that positive conceivability gives us access to only possible worlds, the third says that it gives us access to all the possible worlds, and the second says that we can know all the truths about these possible worlds. In effect, we have a telescope that gives us access to all and only the stars, and that tells us the exact composition of every star. If this thesis is true, the epistemology of modality, at least when idealized, will be simple and beautiful.

APPENDIX: THE MIND-BODY PROBLEM

With these conceivability-possibility theses in hand, it is interesting to apply them to various conceivability arguments against materialism in the philosophy of mind.

Historically, the most important such argument has been Descartes' conceivability argument. This argues from the conceivability of my existing without a body to the possibility of my existence without a body, and from there to the claim that I am not physical. The soundness of this argment is often doubted, and the standard reasons for doubt can be expressed straightforwardly in the current framework. The sense in which it is clearly conceivable that I am disembodied is primary positive conceivability, from which the 1-possibility of disembodiment follows. The sense in which physical things are essentially physical involves 2-necessity (as do all claims of de re necessity). But the 1-possibility of disembodiment is quite compatible with the 2-impossibility of disembodiment, so the claim that I am physical is not threatened by Descartes' argument.

More recently, the knowledge argument and the zombie argument against materialism have been widely discussed. Here, let P be the conjunction of physical truths about the world, and let Q be a

phenomenal truth. The zombie argument claims that zombies, and therefore P&~Q, are primarily positively conceivable. (Here, Q might be "someone is conscious".) The knowledge argument claims that Q cannot be derived a priori from P, so that P&~Q is primarily negatively conceivable. (Here, Q might be "someone is having a such-and-such experience".) From here, both step to the denial of materialism. If we use the current framework to analyze these arguments, a first pass might yield something like the following:

> (1) P&~Q is ideally primarily positively (negatively) conceivable.
> (2) If P&~Q is ideally primarily positively (negatively) conceivable. then P&~Q is primarily possible.
> (3) If P&~Q is primarily possible, materialism is false.
> ----
> (4) Materialism is false.

Here, premise (2) is a special case of the two main conceivability-possibility theses already outlined. It is notable that the zombie argument requires a weaker epistemic-modal premise here: the thesis connecting positive conceivability to possibility is weaker than the thesis connecting negative conceivability to possibility. (It requires excluding only strong necessities, not inscrutabilities.) This is offset to some extent by the fact that the knowledge argument requires a weaker epistemic premise: the claim that P&~Q is negatively conceivable is weaker than the claim that it is positively conceivable.

What of the epistemic premise (1)? This is widely although not universally accepted in the knowledge argument case, and to a somewhat lesser extent in the zombie case. A materialist might deny it in two ways: either by denying even the prima facie conceivability of P&~Q, or by accepting prima facie conceivability but denying ideal conceivability. Some type-A materialists will deny even prima facie conceivability, but this denial is not easy to defend, since it runs counter to a very strong intuition. Others accept prima facie conceivability but deny ideal conceivability, holding that there may be a deep epistemic connection between P and Q, and a deep a priori contradiction in the notion of a zombie.

The second position, exploiting the gap between prima facie and ideal conceivability, may seem particularly promising, especially in the view (which I have elsewhere called 'type-C materialism') that holds that there is a deep epistemic connection that we have not found yet, or perhaps cannot find. But there are two problems. First, this materialist will concede that zombies are not just prima facie but secunda facie positively conceivable, and we have seen that secunda facie conceivability is an extremely good guide to possibility. Second, defeating ideal conceivability will require an a priori entailment from physical to phenomenal, which will require an analysis of phenomenal concepts that can support that entailment. Given the structural-dispositional nature of the physical concepts in P, this requires a structural or functional analysis of phenomenal concepts. But there is good reason to believe that any such analysis of phenomenal concepts is a misanalysis. So while type-C strategy is an interesting strategy that deserves investigation, I think we have reason to believe that it will not succeed.

That leaves premise (3). Here, one runs up against the same problem as in the Cartesian argument. Materialism requires that the physical truths *secondarily* necessitate all truths, and so requires that P&~Q is secondarily possible. But there is no clear inference from the primary possibility of P&~Q to its secondary possibility. So the argument seems to be unsound as it stands. (An eliminativist who denies Q might deny (3) for different reasons, but I will set that position aside here.)

In this case, unlike the Cartesian case, the argument can be rescued. First, one can observe that *if* P and Q both had identical primary and secondary intensions (up to centering), then premise (3) would be straightforwardly true. Further, it is very plausible that the most important phenomenal concepts do indeed have the same primary and secondary intensions (see Chalmers 2002b), so that Q at least can be accommodated here. And even if this is false, Q's primary intension can be seen as the secondary intension of some other truth Q', which stands to Q roughly as "watery stuff" stands to "water". As long as P has the same primary and secondary intension, then the primary possibility of P&~Q will entail the secondary possibility of P&~Q', which will itself entail the falsity of materialism.

A loophole emerges: it is not clear that P has the same primary and secondary intension. It can reasonably be argued that physical concepts have their reference fixed by some dispositional role, but refer to an underlying categorical property. If so, their primary intensions pick out whatever plays a certain role in the world (irrespective of categorical nature), while their secondary intensions pick out instances of a certain categorical property (irrespective of its role). If so, the purported "zombie world" in which the primary intension of P&~Q holds may be a world in which the secondary intension of P is false, so we cannot infer the secondary possibility of P&~Q (or P&~Q').

However, this loophole opens up only a small space for the materialist. Consider the conceived world W, in which the primary intension of P&~Q holds. Because the primary intension of P holds, this world must be structurally-dispositionally isomorphic to the actual world, with the same patterns of microphysical causal roles being played. If P's secondary intension fails, it can only be because these microphysical causal roles have different categorical bases in W (or just possibly, no categorical bases at all). This difference is the only microphysical difference between our world and W. If physicalism is true, it is this difference that is responsible for the presence of consciousness in our world and its absence in W.

What results is a view on which the existence of consciousness is not necessitated by the structural or dispositional aspects of the microphysics of our world, but is necessitated by the categorical aspects of microphysics (the underlying categorical basis of microphysical dispositions), perhaps in combination with structural/dispositional aspects. This is an important view: it is the view put forward by Russell (1926) and discussed in recent years by Maxwell (1978), Lockwood (1989), and others. In effect, the view holds that consciousness stems from the underlying categorical aspect of microphysics. On this view, the nature of the categorical aspect is left open by physical theory, but it turns out to involve special properties that are collectively responsible for constituting consciousness. We can call these special properties *protophenomenal*: they might not themselves be phenomenal properties, but they stand in a constitutive relation to phenomenal properties. We can call the view as a whole

panprotopsychism.

It is not clear whether this sort of panprotopsychism qualifies as a version of physicalism. That question turns on whether the underlying protophenomenal properties are best counted as physical properties, or not. We need not settle that question here: We need only note that if it is a sort of physicalism, it is a quite unusual sort, and one that many physicalists do not accept. In many ways, it has more in common with nonmaterialist views, in virtue of its postulation of fundamental protophenomenal properties whose nature is not revealed to us by physical theory.

In any case, we are now in a position to reformulate the relevant argument:

> (1) P&~Q is ideally primarily positively (negatively) conceivable.
> (2) If P&~Q is ideally primarily positively (negatively) conceivable, then P&~Q is primarily possible.
> (3) If P&~Q is primarily possible but not secondarily possible, then panprotopsychism is true.
> (4) If P&~Q is secondarily possible, materialism is false.
> ----
> (5) Materialism is false or panprotopsychism is true.

The argument (in both versions) is valid, and I have given reasons to accept all of the premises. Note that one can substitute the secondary possibility of P&Q' for the secondary possibility of P&Q in the third and fourth premises, if necessary. Note also that I have said nothing about the role of indexicals and centering. One might think that these raise another loophole in the argument (around premise (3)), by opening another gap between primary and secondary possibility. It is not hard to give a fuller version that takes this role into account (see Chalmers 1998), but I omit the details here for reasons of space.

Finally, a note on Stalnaker's paper in this volume, concerning the zombie argument. Stalnaker (through his character "Anne") questions the argument, by questioning the inference from conceivability to possibility. In effect, he invokes a notion of conceivability distinct from

those discussed here, which we might call *1-2-conceivability*. It is 1-2-conceivable that S if it is primarily conceivable that S is secondarily possible, or more precisely, if "possibly S" is 1-conceivable, where the modal operator here represents 2-possibility. Stalnaker accepts that zombies are 1-2-conceivable (property dualism is not a priori false, and if property dualism is true, then zombies are 2-possible). But he notes that the 1-2-conceivability of S does not entail the possibility of S: the epistemic possibility of property dualism is compatible with the truth of materialism, and with the 2-impossibility (and 1-impossibility) of zombies.

Stalnaker is right that the conceivability of zombies, in this sense, does not directly entail the falsity of materialism. But this sort of conceivability plays no role in the arguments I have given. What is relevant is simply the *1-conceivability* of P&~Q. In the knowledge argument, the argument aims to directly establish the non-apriority of P⊃~Q, and so the prima negative conceivability of P&~Q. In the zombie argument itself, the claim is that it is conceivable that *in the actual world*, P holds but no-one is conscious. (Of course I know that I am conscious, but this is a posteriori knowledge; that issue can also be bypassed by considering only the epistemic possibility that P holds while *others* in the actual world are zombies.) That is, the claim is that P&~Q is primarily positively conceivable. Stalnaker says nothing to cast doubt on this claim (or the analogous claim about negative conceivability), and he says nothing to cast doubt on the inference from primary conceivability to primary possibility. So his discussion leaves this argument untouched.

Acknowledgements

I have presented this material at Arizona State University, New York University, Princeton University, and the Mighty Midwestern Metaphysical Mayhem conference at Notre Dame in summer 1999. Thanks to many on those occasions and elsewhere for discussion of these issues. Special thanks to Chris Evans, Tamar Gendler, John Hawthorne, Bernie Kobes, and David Sosa for comments on the paper, and to Steve Yablo for very helpful discussion and correspondence.

References

Bealer, G. 1996. A Priori Knowledge and the Limits of Philosophy. *Philosophical Studies* 81:121-42.

Benacerraf, P. 1965. What numbers could not be. Philosophical Review 74:47-73.

Benardete, J.A. 1964. *Infinity: An Essay in Metaphysics*. Oxford University Press.

Block, N. & Stalnaker, R. 1999. Conceptual analysis, dualism, and the explanatory gap. Philosophical Review 108:1-46.

Chalmers, D.J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Chalmers, D.J. 1998. Mind and modality. Lectures given at Princeton University. http://www.u.arizona.edu/~chalmers/papers/mm.html.

Chalmers, D.J. 1999. Materialism and the metaphysics of modality. Philosophy and Phenomenological Research 59:473-96.

Chalmers, D.J. 2002a. On sense and intension. http://www.u.arizona.edu/~chalmers/papers/intension.html

Chalmers, D.J. 2002b. The content and epistemology of phenoenal belief. In (A. Jokic & Q. Smith, eds) *Consciousness: New Philosophical Essays*. Oxford University Press. http://www.u.arizona.edu/~chalmers/papers/belief.html

Chalmers, D.J. (forthcoming a). The nature of epistemic space. http://www.u.arizona.edu/~chalmers/papers/espace.html

Chalmers, D.J. & Jackson, F. 2001. Conceptual analysis and reductive explanation. *Philosophical Review* 110:315-61. http://www.u.arizona.edu/~chalmers/papers/analysis.html.

Davies, M. & Humberstone, I.L. 1981. Two notions of necessity. Philosophical Studies 58:1-30.

Evans, G. 1977. Reference and contingency. The Monist 62:161-89.

Fine, K. (this volume).

Hawthorne, J. 2000. Before-effect and Zeno causality. Nous 34:622-33.

Hill, C.S. 1997. Imaginability, conceivability, possibility, and the mind-body problem. Philosophical Studies 87:61-85.

Horgan, T. 1986. Psychologism, Semantics, and Ontology. Nous 20: 21-31.

Jackson F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press.

Lavine, S. forthcoming. *Skolem was Wrong*. Book manuscript.

Lewis, D. 1986. *On the Plurality of Worlds*. Blackwell.

Loar, B. 1997. Phenomenal states (second version). In (N. Block, O. Flanagan, and G. Güzeldere, eds) *The Nature of Consciousness: Philosophical Debates*. MIT Press.

Loar, B. 1999. On David Chalmers' *The Conscious Mind*. Philosophy and Phenomenological Research 59:465-72.

Lockwood, M. 1989. *Mind, Brain, and the Quantum*. Oxford: Blackwell.

Maxwell, G. 1978. Rigid designators and mind-brain identity. In (C.W. Savage, ed.) *Perception and Cognition: Issues in the Foundations of Psychology* (Minnesota Studies in the Philosophy of Science, Vol. 9). Minneapolis: University of Minnesota Press.

Menzies, P. 1998. Possibility and conceivability: A response-dependent account of their connections. In (R. Casati & C. Tappolet, eds) *Response-Dependence* (European Review of Philosophy, volume 3). CSLI Press.

Quine, W. 1961. *Word and Object*. MIT Press.

Russell, B. 1926. *The Analysis of Matter*. Kegan Paul.

Sidelle, A. (this volume). On the metaphysical contingency of laws of

nature.

Stalnaker, R. (this volume). What is it like to be a zombie?

van Cleve, J. 1983. Conceivability and the Cartesian argument for dualism. Pacific Philosophical Quarterly 64:35-45.

Yablo, S. 1993. Is conceivability a guide to possibility? Philosophy and Phenomenological Research 53:1-42.

Yablo, S. 1999. Concepts and consciousness. Philosophy and Phenomenological Research.

Yablo, S. (this volume). Coulda, woulda, shoulda.