I

Modal Epistemology and the Rationalist Renaissance

GEORGE BEALER

The term 'modal epistemology' may be understood in three ways. First, as the theory of *modal* knowledge—knowledge of what is necessary and possible. Second, as the theory of *possible* knowledge—what sorts of knowledge are *possible*. Third, as the intersection of the first two: the theory of *possible modal knowledge*—that is, of what modal knowledge is possible.

The primary question of modal epistemology in this third sense is this. What is the relationship between the a priori and the modal? Most traditional rationalists held that, for all p, p is necessary iff p is knowable a priori. But Saul Kripke (1980) taught us that this traditional equivalence fails in both directions. His meter-stick case is a counter-example to the right-to-left direction. And he argued, along with Putnam (1975), that natural kind identities (e.g., water = H_2O) are counter-examples to the left-to-right direction. In addition, every 'actualization' of any true contingent proposition (e.g., the proposition that *in the actual world* Aristotle taught Alexander) is another sort of counter-example to the left-to-right direction. In fact, because of actualizations, the

I wish to thank Iain Martel for insightful discussions and advice during the preparation of the chapter, Mark Moffett, Michael Peirce, and Stephen Biggs for discerning critical suggestions on the completed manuscript, and Tamar Gendler and John Hawthorne for judicious editorial guidance and general philosophical acumen.

left-to-right direction wrongly implies that every true proposition is knowable a priori. Finally, independently of such broadly Kripkean considerations, the left-to-right direction also fails if there are necessary propositions that in principle cannot be thought. There are such unthinkable propositions if, for example, there could be individual entities that cannot coexist in any one possible world (or properties F and G that cannot both have instances in any common world), and whose unique descriptions are beyond any one person's descriptive powers. At least relative to minds like ours, it seems that there could be such things.

These considerations lead naturally to the further question: for which *types* of proposition p does the traditional equivalence hold? (Or, for which types do the separate halves of the equivalence hold?) The safest interesting generalization is this: the equivalence holds for all (or at least most) *semantically stable* propositions p—roughly, propositions that are invariant across communities whose epistemic situations are qualitatively identical.[1] What this means is that p is the sort of proposition immune to scientific essentialism and externalism generally, as well as to the other counter-examples to traditional rationalism. Semantically stable propositions include virtually all central propositions of the traditional a priori disciplines—logic, mathematics, and philosophy. Accordingly, these disciplines can, at least in principle, be independent of the empirical sciences, as traditional rationalists believed. Since this autonomy thesis is at the same time consistent with the truth of scientific essentialism and other phenomena that brought down traditional rationalism, this view may rightly be called *moderate rationalism*.

I have defended a qualified form of moderate rationalism in a series of papers (1987, 1992, 1994, 1996, 1999). A large number of other contemporary philosophers have also become convinced of one form of moderate rationalism or another. While agreeing with their general conclusion, I believe that in many cases their path to it is flawed—often in the conceptual and logical preliminaries. Additional problems arise from terminology itself: sometimes traditional debates are distorted, even trivialized, by nonstandard uses of well-established (and acceptably clear) ordinary-language expressions and traditional philosophical vocabulary. As Frank Jackson puts it (1998: 31), such terminological distortions can turn 'interesting philosophical debates into easy exercises in deductions from stipulative definitions'.

[1] More precisely, (for thinkable p) p is semantically stable iff, necessarily, if p plays some cognitive role in the mental life of a community c, then it is necessary that for any other community c in qualitatively the same epistemic situation as c, no proposition can play that role other than p itself. See Bealer (1987, 1994). I should note that Eli Hirsch (1986) advocates a similar, independently arrived at, position in his elegant paper.

My aim here is to correct a number of these problems, laying the groundwork for a more acceptable modal epistemology and its correct application. In section 1, I clarify a number of (often nontrivial) conceptual and terminological preliminaries concerning intuition (and, in particular, modal intuition), modal error, conceivability, metaphysical possibility, and epistemic possibility. Section 2 is concerned with the appropriate logical (and semantical) framework for modal epistemology, the main conclusion being that two-dimensionalism is unfit for this role and that a certain nonreductionist approach to the theory of concepts and propositions is required instead. In section 3 I turn to the positive story—a moderate rationalist modal epistemology which includes an account of what it is to understand one's concepts and, as a corollary, an account of an important family of modal errors (namely, those that arise from misunderstanding one's categorial concepts). In section 4, I examine moderate rationalism's impact on modal arguments in the philosophy of mind—for example, Yablo's disembodiment argument and Chalmers's two-dimensional modal arguments. I close by defending a less vulnerable style of modal argument, which nevertheless wins the same anti-materialist conclusions sought by these other arguments.

1 Intuition, Conceivability, Possibility

1.1 *Intuition and the A Priori*

Intuition is the source of all a priori knowledge—except, of course, for that which is merely stipulative. The use of intuitions as evidence (reasons) is ubiquitous in our standard justificatory practices in the a priori disciplines—Gettier intuitions, twin-earth intuitions, transitivity intuitions, etc. By intuitions here, we mean *seemings*: for you to have an intuition that A is just for it to *seem* to you that A. Of course, this kind of seeming is *intellectual*, not experiential—sensory, introspective, imaginative. Typically, the contents of intellectual and experiential seeming cannot overlap. You can intuit that there could be infinitely many marbles, but such a thing cannot seem experientially (say, imaginatively) to be so. Intuition and imagination are in this way distinct. Descartes was right, I believe, to distinguish sharply between imagination and understanding, especially intuitive understanding.[2]

Intuition is different from belief: you can believe things that you do not intuit (e.g., that Rome is the capital of Italy), and you can intuit things that you do not believe (e.g., the axioms of naïve set theory). The experiential parallel is that

[2] [See Introduction, sect. 2.2—eds.]

you can believe things that do not appear (seem sensorily) to be so, and vice versa. Intuition is in similar ways different from other propositional attitudes (judging, guessing, etc.) and from common sense. After surveying the alternatives, I can see no choice but that intuition is a *sui generis* propositional attitude.

The set-theoretic paradoxes establish an important moral: namely, that intuition can be fallible, and that a priori belief is not unrevisable. Infallibilism and unrevisability have often been red herrings in modal epistemology. An alternative tradition—from Plato to Gödel—recognizes that a priori justification is fallible and holistic, relying respectively on dialectic and theory construction.

The sort of intuitions relevant to the a priori disciplines are *rational* intuitions, not *physical* intuitions; only the former present themselves as necessary. According to traditional usage, 'thought experiments' appeal, not to rational intuition, but to physical intuitions (and the like). Here one constructs a hypothetical case about which one tries to elicit, say, a physical intuition deriving from one's implicit mastery of relevant physical laws (as, for example, in Newton's bucket thought experiment). The contrast with Gettier cases, de Morgan's laws, and so forth is plain.

A tendency of late has been to stretch the traditional term 'a priori knowledge' by artificially restricting what is meant by 'experience'—for example, by omitting wholesale Locke's second category of experience, knowledge by reflection (or introspection). Accordingly, one's knowledge of one's self-intimating conscious states is wrongly classified as a priori. An easy way to avoid such confusions is to give a *positive* characterization of a priori knowledge—as opposed to the customary negative characterization as knowledge not based on experience. Perhaps one could do so along the following lines: x knows p a priori iff x knows p and this is direct intuitive knowledge or stipulative knowledge or is based wholly upon such knowledge and/or intuitional evidence.[3]

It is our standard epistemic practice to use intuitions as evidence (or reasons): by virtue of having an intuition that p, one has a prima-facie reason or prima-facie evidence for p. Much as we take our ostensible sense perceptions to be prima-facie evidence if we lack special reason not to do so, it would be unreasonable not to do the same for intuitions. I have argued (1992), however, that we have no special reason not to take intuitions this way and, moreover, that if we deny that intuitions are prima-facie evidence, we are put in an epistemically self-defeating situation. For these reasons, I conclude, we are justified in continuing with our standard practice. In what follows I will assume that this is correct.

[3] Is a priori knowledge exhausted by conceptual analysis? No, not unless the latter includes various necessities that traditionally were thought to be synthetic, not analytic.

There is no relevant phenomenological difference between modal and nonmodal intuitions. For example, there is no relevant phenomenological difference between your intuition that any arbitrary object that has a shape has a size and your intuition that it is possible for something to have a shape and a size, or your intuition that it is not possible for there to be something having a shape but no size. Nor are there good grounds for thinking that modal intuition is not prima-facie evidence whereas nonmodal intuition is. In particular, modal intuition's tie to the truth has a satisfactory explanation and is no more prone to uncorrectable error (see section 3). For these reasons, it would be unreasonable to deny the evidential force of modal intuition and, in turn, unreasonable to deny that just as your nonmodal intuitions are a (fallible) guide to nonmodal truth, so your modal intuitions are a (fallible) guide to modal truth. As a special case, therefore, it would be unreasonable to deny that your possibility intuitions are a guide to possibility. In what follows I will assume that this too is correct.

1.2 *Conceivability and Imaginability*

There is a venerable tradition of taking conceivability and inconceivability to be the evidential basis for, and guide to, a priori knowledge of possibility and impossibility. I think this is a mistake. Intuition is, as we have seen, comparatively easy to characterize—at least provisionally. Not so conceivability and inconceivability (at least if the scholarly literature is any indication). Or maybe I am wrong about this; maybe these, too, are easy to characterize. Perhaps when I say 'It is conceivable that p', all I am saying (at least conversationally) is that I have an intuition that p is possible; and when I say 'It is inconceivable that p', all I am saying is that I have an intuition that it is impossible that p. If so, a great amount of unnecessary confusion would be avoided if we were simply to stop using 'conceivable' and 'inconceivable' and to confine ourselves to talking directly about possibility and impossibility intuitions. The same goes for 'imaginable' and 'unimaginable'.

Suppose, however, that this easy idiomatic gloss on 'conceivable' and 'inconceivable' is not correct, and that these terms are instead taken at face value as literal expressions of certain modal facts: it is conceivable that p iff it is possible for someone to conceive that p; it is inconceivable that p iff it is not possible for someone to conceive that p. Then we have a pair of problems. First, unlike intuitions of possibility and impossibility, conceivability and inconceivability would not be suited to play their reputed evidential role in modal epistemology. That it is possible, or impossible, to conceive that p is itself a mere modal fact. But in order for someone to acquire evidence (reasons), something must *actually happen*: a datable psychological episode must occur (the occurrence of

a sensation, an introspective or imaginative experience, a seeming memory, an intuition). Modal facts do not occur. Nothing *happens* when something is conceivable or inconceivable. So something's merely being conceivable or inconceivable cannot provide anyone with evidence (reasons) for anything.

Not only that, our beliefs about what is conceivable and inconceivable can be highly inferential and are often theoretical. True, one way you can come to believe that it is possible for someone to conceive that p is for *you actually* to conceive that p. But why should your conceiving that p provide you with evidence that p is possible? I can see no reason why it should, unless conceiving that p involves intuiting that p is possible.[4] But this takes us back to relying evidentially on modal intuition.[5] In any case, many of our beliefs about conceivability arise, not by way of actual conceivings, but indirectly via our (implicit) modal epistemology—our general beliefs about what classes of propositions can and cannot be conceived. And in the case of inconceivability, something like this *must* be the case. The mere fact that I tried, but happened to fail, to conceive that p is not a good guide to what is in principle possible in this regard for any being whatsoever; maybe I am just not sharp enough. Of course, there is another way we could come to have beliefs about what is and what is not possible to conceive: we can just have modal intuitions to that effect. But then, once again, we are back to relying on modal intuitions as our source of evidence, except that these intuitions are one step removed, for they do not directly concern the possibility or impossibility of p, but only an associated psychological possibility or impossibility of conceiving that p. The moral is simple: in the matter of evidence for possibility and impossibility, talk of conceivability and inconceivability is an idle complication that only breeds confusion. Again, these points hold for 'imaginable' and 'unimaginable'.

As observed at the outset, most traditional rationalists held that, for all p, p is necessary iff p is knowable a priori. Many traditionalists also accepted an associated equivalence between conceivability and possibility: it is possible that

[4] Suppose instead that x conceives that p iff x conceives *of* a possible situation in which p. But what is it to conceive of a situation in which p? In one sense of 'conceive', conceiving of a situation is merely thinking of it. But this sort of conceiving would provide no evidence that p is possible: after all, one can think of impossible situations (e.g., that the square root of 2 is rational); moreover, if one succeeds in thinking of a situation that happens to be possible, that could be a matter of pure chance (e.g., a result of reading the works of the monkeys at the typewriters). Conceiving of a possible situation can be evidential only if, in the very conceiving of the situation, the situation *seems* possible—that is, only if the conceiving involves the intuition that it is possible. As Stephen Yablo tells us: 'In slogan form: *conceiving involves the appearance of possibility*' (1993: 5). But then we are right back to relying on modal intuitions.

[5] As we saw above, intuiting that p is possible *does* provide evidence that p is possible, for this is just a special case of the fact that intuiting that q provides evidence that q.

p iff it is conceivable that p. On the literal, face-value interpretation, this equivalence is: p is possible iff it is possible for someone to conceive that p. On the supposition that x conceives that p iff x intuits that p is possible, the traditional equivalence is then: p is possible iff it is possible for someone to intuit that p is possible. This equivalence fails in both directions. (1) We know that the right-to-left direction fails, because intuitions are fallible (e.g., in view of the paradoxes). (2) The left-to-right direction fails in the event that there are unthinkable propositions p of the sort discussed at the outset.[6]

Returning to the general matter of terminology, my overall judgment is that it is safest simply to avoid the minefield of 'conceivability' and 'inconceivability' and to stick to the unproblematic idiom of intuitions of possibility and impossibility.[7]

1.3 *Metaphysical Possibility and Epistemic Possibility*

The modal expressions 'could', 'can', 'might', 'possible' are used in diverse ways that fall into two broad classes: (i) epistemic and (ii) nonepistemic. (An analogous division holds for 'must' and 'necessary'.) The first thing to note about these two classes of uses is that they do not mark out two associated *genuses*—epistemic possibility and nonepistemic possibility—of some common still more general modal category *Possibility*. The analogous point holds for the specific properties expressed by the various epistemic and nonepistemic uses of 'could': for any pair of properties, one of which is associated with the former class and the second with the latter, these properties are not *kinds* or *species* of some common, still more general modal category *Possibility*. Take, for example, the properties expressed, respectively, by the 'could'-of-less-than-complete-certainty and the 'could'-of-metaphysical-possibility. There simply is no common general modal category *Possibility* of which these two properties are kinds or species. (This of course is not to say that such properties are entirely unrelated, but if they are related, this would result from a much deeper connection—perhaps revealed in the genesis of our concepts.)

[6] It has been suggested (e.g., Chalmers 1996: 68) that we understand 'conceivable' to be equivalent to 'conceivable by a being with maximal cognitive powers', but this is not the standard use of the term in the history of philosophy. It is likewise nonstandard to use 'p is conceivable' to mean 'the possibility of p would be affirmed at the close of a priori deliberation by a being with maximal cognitive powers' (or something of the sort). But, for what it is worth, on these nonstandard readings, the resulting equivalences are threatened by much the same problems as those that undermined traditional rationalism.

[7] In support of the centrality of modal intuition, Stephen Yablo tells us that 'modal intuition *must* be accounted reliable if we are to credit ourselves with modal knowledge' (1990: 179).

In logic and metaphysics the primary focus is, in Kripke's words, necessity *tout court*. The dual of this use of 'necessary' is the weakest of the nonepistemic uses of 'possible'—weakest in the sense that its extension always includes the extensions of all other nonepistemic uses of 'possible'.

Because 'possible' and 'necessary' have diverse uses in ordinary language (both epistemic and nonepistemic), confusion easily creeps in. In logic and metaphysics there is an easy antidote. Most of us are inclined to hear the terms 'contingent' and 'noncontingent' univocally. So we can almost always head off misunderstanding in logic and metaphysics simply by 'translating' modal remarks into this univocal idiom. p is necessary iff p is true but not contingent. p is impossible iff p is false but not contingent. p is possible iff p is necessary or contingent. (That is, p is possible iff either p is true but noncontingent or p is contingent.) For example, when I say, 'Despite the recent proof, there is still a possibility that Fermat's Last Theorem is false', I am certainly not asserting that Fermat's Last Theorem is either necessarily false or contingently false! Likewise, when I say, 'It could have turned out that Hesperus was not Phosphorus', I am assuredly not asserting that it is either necessarily false or contingently false that Hesperus is Phosphorus. Using this 'translation' test will dispel virtually all such confusions.

Various species of possibility (in the identified sense) may be isolated as follows: p is nomologically possible iff p and the laws of nature are compossible (i.e., it is possible for p and the laws of nature to be true together); p is physically possible iff p and the laws of physics are compossible.[8] What about logical possibility? Well, mimicking this pattern of definition, we would have: p is logically possible iff p and the laws of logic are compossible. But *every* possibility p is compossible with the laws of logic, so logic rules out nothing that is not *already* ruled out. Therefore, according to this definition, logical possibility and possibility coincide. That is, logical possibility is not a *species* of possibility; rather, it is just possibility itself. And this is precisely how 'logical possibility' has historically been used by a great many philosophers—for example, by Kripke himself at the advent of *Naming and Necessity*. What about 'metaphysical possibility'? This is a technical term which Kripke stipulatively introduced, solely for heuristic purposes, as a synonym of his term 'logically possible'—that is, of 'possible'. Thus, according to this standard philosophical usage, p is possible iff p is logically possible iff p is metaphysically possible iff p is necessary or contingent iff either p is true but noncontingent or p is contingent.

Despite this fully documented history, some people insist on distinguishing logical possibility and metaphysical possibility and so are led to the following: p is logically possible iff p is merely *consistent* with the laws of logic (i.e., not

[8] For challenges to this suggestion, cf. Fine, Ch. 6 below.

ruled out by logic alone). This usage, however, invites confusion.[9] There are many logically consistent sentences that express obvious impossibilities (e.g., 'Bachelors are necessarily women', 'Triangles are necessarily circles', 'Water contains no hydrogen'). If you buy into calling mere logical consistency a kind of possibility, why not keep going? For example: p is 'sententially possible' iff p is consistent with the laws of sentential logic. Then, since 'Everything is both F and not F' is not ruled out by sentential logic (quantifier logic is what rules it out), would it be possible in some sense (i.e., sententially possible) that everything is both F and not F?![10] Certainly not to my ear! At this juncture it seems to me that the best policy is simply to eschew the problematical term 'logically possible' and to confine ourselves to the well-demarcated terms 'logically consistent' and 'metaphysically possible'.

Let us return to standard epistemic uses of 'could'.[11] To illustrate some of them, consider any thinkable necessary truth p. The first use is the 'could'-of-ignorance: absent what we deem to be adequate evidence (or adequate justification) one way or the other about p, we can truly say, 'It could be that p, and it could be that not p. We just do not know yet.' (For example, this can be truly said of Goldbach's conjecture.) But once we have adequate evidence (justification) one way or the other, what was meant in speaking *that way* can no longer be truly said. Second, there is the 'could'-of-less-than-complete-certainty: if we have less than complete certainty about p (even if we have adequate evidence, or justification, for p), we can still truly say, 'We still could be mistaken; we know we can be wrong about almost anything.' (For example, even though we now have a proof of Fermat's Last Theorem, this can still be truly said of it.) Third, there is the 'could'-of-qualitative-evidential-neutrality: for a posteriori necessities, we can often truly say, 'It could have turned out that p, and it could have turned out that not p.' And this is so, even though, meant this way, it cannot be said of any traditional a priori necessities. For example, meant this way, 'Whether Hesperus was Phosphorus could have turned out

[9] This is certainly not to say that the (very well-studied) notion of logical consistency is unimportant.

[10] Some people also hypothesize a use of 'possible' for what they call 'conceptual possibility': p is possible (in the hypothesized sense) iff it is impossible for anyone to know a priori that p is false. Two points: first, even if there were such a use of 'possible' in ordinary English, it would (just as in earlier cases) not expresses a *kind* of possibility; second, there is in fact no such use of 'possible' in ordinary English. Here is one among many counts on which this is so: if p is an unthinkable proposition, then the proposition that p is false is likewise unthinkable. Thus, it is impossible for anyone to know—and hence, think—that p is false. So, on the hypothesized use, it would be possible (i.e., 'conceptually possible') that p. And this would be so even if p were *logically inconsistent*!

[11] These uses of 'could' need not correspond to distinct *literal meanings*; it is enough that they are standard uses of the term in the sort of ordinary contexts relevant to modal epistemology.

either way' would be true, even though, when meant the same way, 'Fermat's Last Theorem could have turned out either way' would be false.

A few semi-formal remarks about these epistemic uses of 'could' might be helpful. Suppose someone intends the 'could'-of-ignorance when uttering the sentence 'It could be that p' in some relevant conversational context.[12] Then, the asserted proposition would be the proposition that results when an associated propositional operation $\Diamond_{\text{ignorance}}$ is applied to the proposition P. (In symbols: $\Diamond_{\text{ignorance}}$ p.) The truth-conditions are: the proposition that $\Diamond_{\text{ignorance}}$ p is true iff it is unknown, one way or the other, whether p. Likewise, the 'could'-of-less-than-complete-certainty may be represented with the operator '$\Diamond_{\text{uncertainty}}$'. The truth-conditions are: the proposition that $\Diamond_{\text{uncertainty}}$ p is true iff it is not completely certain that p. Finally, the 'could'-of-qualitative-evidential-neutrality may be represented with '$\Diamond_{\text{qual-evid-neut}}$'. The truth-conditions are: the proposition that $\Diamond_{\text{qual-evid-neut}}$ p is true iff it is possible for there to be a population c with attitudes toward p and it is possible for there to be a population c′ whose epistemic situation is qualitatively identical to that of c such that the proposition which in c′ is the epistemic counterpart of p in c is true.[13]

Note that in each of these three biconditionals the whole proposition mentioned on the left-hand side need not be *identical* to that expressed by the associated right-hand side, and intuitively, they are indeed different. This feature allows the above account to avoid various difficulties that undermine other accounts of epistemic uses of 'could'. A case in point is Kripke's account of the 'could'-of-qualitative-evidential-neutrality. As we shall see in section 1.4, the above account avoids problems confronting Kripke's account, while at the same time preserving a thesis latent in Kripke's discussion—namely, that it could have turned out epistemically that p iff p has a qualitative epistemic counterpart (in the above sense) that is metaphysically possible. In other words, this sort of epistemic possibility entails the existence of the associated sort of metaphysical possibility.

Kripke's modal argument against materialism is based on the premise that, for a certain proposition p (i.e., that there could turn out to be pain without firing C-fibers, and firing C-fibers without pain), p's epistemic possibility entails its metaphysical possibility. His argument fails, however. Since p is semantically *unstable*, its epistemic possibility entails merely that some qualitative epistemic counterpart be metaphysically possible; it does not entail that p itself is metaphysically possible. But the latter sort of metaphysical possibility is what

[12] Here and certain other places I use single quotes where, strictly, corner quotes are required.

[13] In symbols: $\Diamond(\exists c)\ \Diamond(\exists c')(\exists p')$[QualitativelyIdentical(c′, c) & Counterparts(<p′, c′>, <p, c>) & True(p′)].

Kripke's argument requires. In section 4.3 we shall see that Chalmers's zombie argument—and other two-dimensional modal arguments—resemble Kripke's argument in this respect, and that they fail for the same reason. If, however, p is semantically *stable*, p's epistemic possibility *does* entail p's metaphysical possibility. This follows from the definition of semantic stability: a semantically stable proposition is, by definition, one that is *identical* to all of its epistemic counterparts (see n. 1). It turns out that, for certain semantically stable propositions p, p's epistemic possibility—and hence p's metaphysical possibility—entails the falsity of associated instances of the Identity Theory. For example, if p is the proposition that some being feels pain but does not have 74,985,263 or more functionally related nonconscious parts (or whatever is the minimum number needed for having firing C-fibers), p's epistemic possibility entails its metaphysical possibility. In turn, this entails that having firing C-fibers is not a necessary condition for being in pain and, therefore, that these two properties are not identical. This epistemic possibility suffices as the first step in the anti-materialist argument described in section 4.4.[14]

1.4 *Modal Error I: Epistemic Possibility and Rephrasals*

An important task for modal epistemology is to identify the sources of—and defenses against—erroneous or misleading modal intuitions. In this section, I will consider one such source: namely, confusion between metaphysical possibility and epistemic possibility. The threat of such confusion plays a significant role in Kripke's discussion of scientific essentialism (hereafter SE). In section 3.3, I will return to the topic of modal error, and the account of it offered by Stephen Yablo.

Kripke holds that, taken literally, there is a conflict between his thesis that, say, it is necessary that Hesperus = Phosphorus and our ordinary intuition that it could have turned out that Hesperus was not Phosphorus (1980: 103–5, 140–4).

[14] The foregoing dialectic is spelled out in Bealer (1994). Note that, besides the indicated epistemic intuition, most of us have a direct metaphysical intuition that there could be a being that feels pain but lacks 74,985,263 or more functionally related nonconscious parts. (Our little 'translation' test can be used to establish that this is indeed the 'could' of metaphysical possibility.) Like the epistemic intuition, this intuition also contradicts the thesis that firing C-fibers is a necessary condition for being in pain. And, because this intuition is semantically stable, it too avoids the sort of scientific essentialist worries that confront the traditional modal arguments (i.e., traditional multiple-realizability, zombie, and disembodiment arguments, including Yablo's disembodiment argument; cf. sect. 4.2). We thus have *two* arguments against the Identity Theory that are immune to scientific essentialist worries.

Kripke takes there to be a conflict because he believes that '*it could have turned out that p* entails that p could have been the case' (1980: 141–2). And he believes that, if conflicts like this cannot be resolved, his argument for SE would be foiled.

His resolution is to hold that the apparent conflict among our intuitions is an illusion. All, or most, of our intuitions are correct, but many are *misreported*. For example, our intuitions supporting the necessary identity of Hesperus and Phosphorus are correctly reported, but when we report our apparently contrary intuitions, we confuse ordinary possibility with the possibility of a certain kind of epistemic situation:

> And so it's true that given the evidence that someone has antecedent to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation [to ours], and call two heavenly bodies 'Hesperus' and 'Phosphorus', without their being identical. So in that sense we can say that it might have turned out either way. (1980: 103–4)

Generalizing from Kripke's remarks, one arrives at the following rephrasal schema: The inaccurate statement 'It could have turned out that A' is to be rephrased with the accurate statement 'It is possible that a population of speakers in an epistemic situation qualitatively identical to ours would make a true statement by uttering "A" with normal literal intent'.

But this sort of metalinguistic rephrasal is untenable because of familiar problems concerning fine-grained intensional content. For example, it runs afoul of the Langford–Church translation test.[15] It also runs afoul of the sort of arguments Tyler Burge (1979) and Stephen Schiffer (1987) give against metalinguistic rephrasals of attitude reports.

There is, however, an extremely simple resolution of the problem: namely, to deny that there is conflict in the first place. When people say that it could have turned out that Hesperus was not Phosphorus, they are simply not contradicting the SE thesis that it is necessary that Hesperus = Phosphorus. Why? Because they are just employing a straightforward epistemic use of 'could': namely, the 'could'-of-qualitative-evidential-neutrality. As we saw in the previous section, this use of 'could' does not collide with the metaphysical use. End of story. Kripke took there to be a conflict because '*it could have turned out that p* entails that p could have been the case'. True enough—when the two uses of 'could' are the same. But our little 'translation' test from section 1.3 easily shows

[15] Church (1950: 98) describes this test thus: '[W]e may bring out more sharply the inadequacy of [an analysis] by translating into another language . . . and observing that the two translated statements would obviously convey different meanings to [a speaker of the other language] (whom we may suppose to have no knowledge of English).'

that in the context of Kripke's discussion his first and second uses of 'could' are not the same. (When Kripke granted, 'It could have turned out that Hesperus was not Phosphorus', he was assuredly not committing himself to the claim that it is either necessary that Hesperus ≠ Phosphorus or contingent that Hesperus ≠ Phosphorus. But when he told us, 'Hesperus could not be different from Phosphorus', he was committing himself to the claim that it is neither necessary that Hesperus ≠ Phosphorus nor contingent that Hesperus ≠ Phosphorus.) As soon as we see this, the appearance of conflict vanishes. Moreover, by identifying the first occurrence of 'could' with the 'could'-of-qualitative-evidential-neutrality and by analyzing it in the way proposed in section 1.3, we are able to preserve Kripke's underlying insight, but without the problematic features of his official metalinguistic approach.

Besides Kripke's approach (just discussed), there have been many others. For example, Kripke offers a second style of rephrasal based on the idea that the intuition that the Hesperus–Phosphorus identity 'could have turned out otherwise' is misreported, and that it is correctly reported with definite descriptions ('the so and so heavenly body', 'the such and such heavenly body') rather than names ('Hesperus', 'Phosphorus'). This, too, is seriously flawed (see Bealer, 1994). Another general approach also holds that the intuitions are indeed in conflict, but that the conflict results from an understandable conceptual illusion: specifically, a subtle slipping back and forth between very similar but conflicting intuitions—for example, an intuition about the thing, whatever it is, and an intuition involving the thing's contingent macroscopic properties. A third general approach (see section 4.1 for criticism of this idea) is to hold that, although the conflict is genuine, one of the conflicting intuitions is of a generally unreliable sort and so on that ground should be dismissed. In any case, the simple logico-linguistic account given in the previous section vitiates these three approaches, for each is based on the false assumption that there *is* a conflict in the intuitions at issue.

1.5 *Epistemic Possibility in Two-Dimensionalism*

A final approach to the problem just discussed, which is espoused by David Chalmers, is a variant of the account I have given. On Chalmers's approach (as I understand it) the intuitions at issue are correctly reported from the start, but the intuition-reports are *ambiguous*. There simply is no conflict between the reported intuitions as long as the reports are understood correctly. If (i) 'Hesperus could not have been different from Phosphorus' is understood metaphysically and (ii) 'It could have turned out that Hesperus ≠ Phosphorus' is understood epistemically, there is no conflict. In other words: (iii) although it

is not metaphysically possible, it is nevertheless epistemically possible that Hesperus ≠ Phosphorus. Up to this point, this is just what I have said.

The difference emerges in the way (iii) is dealt with. I represent (iii) in the following straightforward way (where p is that Hesperus ≠ Phosphorus): $\neg\Diamond p \;\&\; \Diamond_e p$. Here I am using '$\Diamond$' for metaphysical possibility and '$\Diamond_e$' for the relevant epistemic use of 'possible', whatever it is. By contrast, to deal with (iii), Chalmers posits a very different sort of ambiguity: namely, an ambiguity in the *sentence* 'Hesperus ≠ Phosphorus' itself. As we will see (section 2.2), Chalmers holds that there are two distinct sorts of meaning: 'primary' and 'secondary'. He holds in particular that a sentence is semantically correlated to what he calls a 'primary proposition' and a 'secondary proposition'. (For Chalmers (1996: 63–4) , a proposition, whether primary or secondary, is a function from possible worlds to truth-values; accordingly, he holds that a proposition is possible iff it is true in some possible world iff it has the value True for some possible world.) Let p_1 be the primary proposition of 'Hesperus ≠ Phosphorus', and p_2 its secondary proposition. Then, according to Chalmers (if I understand him), 'It is epistemically possible that Hesperus ≠ Phosphorus' is equivalent to 'p_1 is possible [i.e., true in some possible world]'. And 'It is not metaphysically possible that Hesperus ≠ Phosphorus' is equivalent to 'p_2 is not possible [i.e., true in no possible world]'. Accordingly, (iii) would be represented thus: $\neg\Diamond p_2 \;\&\; \Diamond p_1$. So, despite all the talk of epistemic possibility, the epistemic use of 'could' disappears; the only use of 'could' is the one for metaphysical possibility. (This fact comes to the foreground in Chalmers's zombie argument; see section 4.3.)

This picture has very implausible consequences. Intuitively, statement (iii) entails the following: (iv) there is something which, although not metaphysically possible, is nevertheless epistemically possible, namely, *that Hesperus ≠ Phosphorus*.[16] But, on the present theory, (iv) cannot be true. Why? Because on the theory, the only ambiguity in sentences like (iii)—and hence, the only ambiguity in sentences like (iv)—is between the primary and secondary intensions associated with 'Hesperus ≠ Phosphorus', and there is only one modality relevant to such sentences: namely, truth in some possible world. Therefore, on the theory, '$\Diamond_e p$' just collapses into '$\Diamond p$'. Consequently, on the theory, (iv) implies that there is something p that both is and is not possible (true in some possible world): namely, that Hesperus ≠ Phosphorus.[17] A contradiction. And this is so whether p is a 'primary proposition', a 'secondary proposition', or any other sort of proposition. Furthermore, this problem generalizes, yielding

[16] That is, there is something p such that p is not metaphysically possible and p is epistemically possible and p is that Hesperus ≠ Phosphorus. In symbols: $(\exists p)(\neg\Diamond p \;\&\; \Diamond_c p \;\&\; p = [H \neq P])$.

[17] In symbols, $(\exists p)((\neg\Diamond p \;\&\; \Diamond p) \;\&\; p = [H \neq P])$.

the conclusion that, on this theory, there is *not one* proposition that, although not metaphysically possible, is nevertheless epistemically possible.

In the face of this contradiction, Chalmers has no choice but to retreat to a metalinguistic analysis of sentences like (iv), according to which only a *sentence*—never a proposition—can be said to be both epistemically possible and metaphysically impossible.[18] But such metalinguistic analyses run afoul of a host of well-known problems (including, e.g., the very sort of problem that confronted Kripke's metalinguistic rephrasals in section 1.3). For this reason, a satisfactory logico-linguistic treatment of epistemic possibility turns out to be out of the reach of Chalmers's two-dimensionalism.

The fact that Chalmers has no choice but to accept the thesis that sentences (never propositions) are the only sort of thing that can be at once epistemically possible and metaphysically impossible should be no surprise, for it is entailed by the Jackson–Chalmers thesis that sentences, never propositions, are the only sort of thing that can be at once necessary and a posteriori. Like the former thesis, the latter thesis is subject to a wealth of well-known arguments against metalinguistic treatments of 'that'-clauses and their interaction with 'could' and 'possible' and other key expressions ('knows', 'believes', etc.).[19] Such 'mixed' constructions, however, form the very heart of intensional logic—which is, of course, the logical framework of modal epistemology. In section 2, I will discuss a number of other reasons why the Jackson–Chalmers two-dimensional framework is incapable of fulfilling this role.

2 Logical Framework for Modal Epistemology

As I mentioned in the introduction, many defenses of moderate rationalism fail right in the conceptual and logical preliminaries. If we are to have an adequate

[18] There are a few minor variants. Chalmers could hold that there are certain other *sentence-like* entities (asserted sentences, Mentalese representations, etc.) which, by virtue of having both primary and secondary intensions, can be both epistemically possible and metaphysically impossible. But this proposals falls prey to analogues of the problems that undermine the metalinguistic analysis in the text. Chalmers also speaks as though 'concepts' (1996: 56ff.) and 'thoughts' (1996: 65) have primary and secondary intensions, but he offers no theory of what 'concepts' and 'thoughts' are, or of what this relation of *having* a primary and secondary intension is. Maybe Chalmers's 'thoughts' are sentence-like entities (as above). Or maybe they are ordered pairs consisting of a primary intension and a secondary intension. But such proposals fall prey to a host of difficulties. For example, they amount to category errors: thoughts are not *really* English sentences, asserted sentences, Mentalese sentences, or ordered pairs! Moreover, these proposals fall prey to a multitude of insurmountable Benacerraf-style problems; see sect. 2.2. (The present note is also intended to apply to the remarks in the next paragraph in the text.)

[19] Section 2.3 provides a new style of argument. See also Bealer (1993*b*).

modal epistemology, it is essential that we first have an adequate logical framework. In this section, after showing in some detail why the two-dimensionalist approach cannot provide one, I will then briefly outline a more satisfactory framework, the nonreductive algebraic framework.

2.1 *Propositions*

There is very strong logical and linguistic evidence for the following tenets. (1) 'Say', 'mean', 'believe', 'is possible', and so forth often function as predicates, and 'that'-clauses function as singular terms in companion sentences (e.g., 'I believe that A', 'It is possible that A'). I propose to use 'proposition' as a technical term for the sorts of entity—*whatever they turn out to be*—denoted by 'that'-clauses occurring in the indicated family of sentences. (2) Propositions (in this neutral sense) are not sentences, uttered sentences, utterances of sentences, or any other such linguistic entities. For example, speakers of diverse languages can believe various common propositions; animals and infants have beliefs but no language; metalinguistic treatments of 'that'-clause sentences do not pass the Langford–Church translation test, and so forth. (3) Propositions can be very fine-grained. For example, it is a truth of logic that triangles are triangles, but a truth of geometry, not logic, that trilaterals are triangles. (Of course, our proposed use of 'proposition' is consistent with the idea that there are also families of propositions that are not so fine-grained, and, indeed, that there could be a spectrum of 'granularities'.)

Frege's puzzle (how can sentences of the form 'A = A' and 'A = B' be true but be different in meaning) also calls for very fine-grained propositions: 'A = A' means that A = A, and 'A = B' means that A = B; therefore, since the two sentences are different in meaning, the associated 'that'-clauses must denote different propositions (in our neutral sense). How can these propositions be different? Frege's solution no longer seems feasible, for Marcus, Kripke, Kaplan, Putnam, and others have convincingly argued that names do not have descriptive content. The problem is compounded by the fact that, if 'A' and 'B' are names such as 'Hesperus' and 'Phosphorus', the proposition that A = B is both necessarily true and knowable only a posteriori, as Kripke and Putnam have shown.

A great irony in contemporary philosophy of language is that most people originally moved by the Kripke–Putnam arguments advocate theories according to which there exist *no* such necessary a posteriori propositions. I have in mind hidden-indexical theories and most other direct-reference theories and the Jackson–Chalmers two-dimensional theory. This creates a special problem in our context, for there can be no satisfactory modal epistemology without

such propositions. Reserving critical discussion of hidden-indexicalism for another setting, I will here examine two-dimensionalism and follow that discussion with a proposal of what I believe is a more satisfactory framework.

2.2 *Two-Dimensionalism, Possibilia, and Understanding Language*

Frank Jackson (1993, 1998) and David Chalmers (1995, 1996) have each proposed a new 'two-dimensional' logical framework, adapted from prior work of David Kaplan, Robert Stalnaker, Martin Davies and Lloyd Humberstone, David Lewis, and Pavel Tichy (see Jackson 1998: 47 n.). The resulting theory is at once a formal semantics, a foundation for intensional logic, a theory of understanding of words and language, and a modal epistemology. I believe that this framework and the theory that flows from it are inadequate.

2.2.1 *The Possible-Worlds Background*

A preliminary concern is that two-dimensionalism is developed within a metaphysics that assumes both possible worlds and possible individuals (nonactual as well as actual) and that constructs all the intensional entities required by philosophical semantics from such possibilia. This applies to propositions, concepts, senses, meanings, mental contents—and so, presumably, also to properties and relations.

There are, I believe, overwhelming reasons for holding that this analytical order is backwards. The notions of concepts and/or properties and relations are fundamental. They, together with modal notions and certain other logical notions, are the basic notions in terms of which the notions of proposition, state of affairs, maximal state of affairs, and so forth are to be characterized and possibilist language is to be analyzed.

Here is one of many arguments for this assessment.[20] Intuitively, it is necessary *that some proposition is necessary*. On the possible-worlds reduction, the property (concept) of being a necessary proposition is a function (set of ordered pairs) from possible worlds to the set of necessary propositions. But this set includes the proposition that some proposition is necessary (because, as just indicated, this proposition is itself necessary). Thus, this proposition belongs to a set belonging to an ordered pair belonging to the property of being necessary.

[20] Two other problems: (1) identifying properties (red) and propositions (that I am thinking) with functions is, other things being equal, simply a category mistake—these things are not functions, not *really*; (2) this reduction is subject to a wealth of Benacerraf-style problems. See Bealer (1993*a*: 20, 22) and Jubien (2001). These two problems, and that in the text, are avoided by the approach in sect. 2.5.

But, to avoid the problem of logical omniscience, our reductionists are forced to treat this proposition as a *structured proposition*, one of whose constituents is the property (concept) of being necessary. Hence, this property belongs to an ordered set that belongs to a set that belongs to an ordered pair that belongs to the property of being necessary. That is, being necessary $\in \ldots \in$ being necessary. Hence, the property of being necessary cannot be a set-theoretical construct built up entirely from possible *particulars* (possible people, possible stones, and the like). Thus, the possible-worlds construction fails for the property of being necessary—in general, for most every iteratable property. Hence, these properties are irreducible *sui generis* entities. But, then, uniformity supports the thesis that all other properties (concepts) are *sui generis* as well.

Chalmers responds to concerns of this sort by holding that the possible-worlds framework and its reductions of propositions (concepts, senses, meanings, thoughts) to possible-worlds constructs is only a tool, which may be taken for granted 'in much the way one takes mathematics for granted' (1996: 66). And Jackson tells us that refusing the possible-worlds approach would be 'not that different from refusing to count one's change at the supermarket because of the ontological mysteries raised by numbers' (1998: 11). But the implied analogy is unsound: the questions at hand concern the ultimate *foundations* of philosophy; at this point one can no longer simply defer the issue. Is two-dimensionalism a satisfactory framework for understanding the fundamental concepts of modal epistemology: necessity, possibility, meaning, property, concept, proposition, mental content, a posteriori necessity, epistemic possibility, and so forth? Considerations like the one given above strongly suggest that it is not.[21]

2.2.2 *The Key Idea of Two-Dimensionalism*

Jackson puts it thus:

> We can think of the [things] to which a term *applies* in two different ways, depending on whether we are considering what the term applies to under various hypotheses about which world is the actual world, or whether we are considering what the term applies to under various counterfactual hypotheses. (1998: 48, emphasis added)

In the former case, the term has an 'A-intension' (A for actual), a 'primary intension' in Chalmers's terminology; in the latter case, a 'C-intension' (C for counterfactual), a 'secondary intension' in Chalmers's terminology. In conventional

[21] A related point concerns the actualism/possibilism debate. Other things being equal, economy makes actualism preferable to possibilism. It is often claimed, however, that possibilist language has a useful expressive power not available to actualist language. The fact of the matter is that an actualist semantics can be given for possibilist language, and this settles the matter in favor of actualism. See Bealer (1998: 25 f.).

possible-worlds terminology, the intension of a term is the C-intension (secondary intension). The A-intension (primary intension) is very different and is characterized by Jackson as follows:

> What then does the word 'water' *denote* in a world where the kind common to the relevant watery exemplars in that world is kind K under the hypothesis that that world is the actual world? Kind K, of course, be that kind H_2O, XYZ, or whatever. (1998: 49, emphasis added)

Chalmers states the idea thus:

> When the XYZ world is considered as actual, my term 'water' *picks out* XYZ in the world. (1996: 60, emphasis added)

But these claims are simply mistaken. The English word 'water' denotes H_2O, as we all know. Kripke and Putnam taught us that, for *every possible situation*, the English word 'water' denotes H_2O in that situation. This just reflects the standard use of the English expressions 'water' and 'denote'. It is simply false that, for some possible situation, the English word 'water' denotes XYZ in that situation! This is so no matter how we might be 'considering' these situations or what 'hypotheses' we might be entertaining about them. For the same reason, it is false that, for some possible situation, the English word 'water' *applies* to picks out, refers to, designates) XYZ in that situation. All such uses of these key English semantical expressions are blatant violations of Jackson's own terminological dictum (cited at the outset)![22]

In Chalmers the problem is even more graphic, for he holds that *concepts* (as well as words) have primary intensions:

> Take the concept 'water'. [T]he primary intension of 'water' maps the XYZ world to XYZ and the H_2O world to H_2O. . . . [I]t picks out the *watery stuff* in a world. (1996: 57; see also 65)

[22] The following alternative definition (cf. Chalmers 1996: 364 n. 21; 365 n. 25) is no better: the A-intension (primary intension) of the English word 'water' is a function from worlds w to the object in w denoted by the sound-alike word 'water' in the counterpart language of the beings in w whose epistemic situation is qualitatively the same as ours. Analogously for sentences. But, as before, such A-intenstions (primary intensions) do not correspond to any standard notion from the semantics for English. For example, the English word 'water' does not *express*, or have as its *meaning*, a function that identifies what *other* speakers of *other* languages would denote with *their* word 'water'! In the case of sentences, there is another way to make the point. Propositions of the following form fix the meaning relation (and, in turn, reference) for English: the English sentence 'S' means *that S*. For example, the English sentence 'Hesperus=Phosphorus' means *that Hesperus=Phosphorus*. But this proposition cannot be the primary intension because, unlike the primary intension, it is necessary; nor can it be the secondary intension because, unlike the secondary intension, it cannot be know a priori (sect. 2.3 elaborates this argument).

From Putnam we know that, for every possible situation, XYZ is not water in that situation. It is simply a misuse of the English expressions 'concept of being water' and 'pick out' (or 'applies to') to say that, for some possible situation, the concept of being water picks out (applies to) XYZ in that situation. Again, this is so regardless of how we might happen to be considering these situations or what hypotheses we might be entertaining about them. To hold otherwise is to hold that, for some possible situations, the concept of *being water* picks out (applies to) things that are *not water* in those situations! And, by parity, this would lead one to hold that, for certain reptilian worlds, the concept of being a human being applies to beings that are not even human beings in those worlds: namely, to human-looking reptiles! *Reductio ad absurdum.*[23]

These and related considerations show that A-intensions (primary intensions) and their kin do not match any of the standard semantical notions from the theory of reference and meaning for natural language. Likewise, the incongruity between C-intensions (secondary intensions) and the fine-grained distinctions exemplified in Frege's puzzle shows that they too do not match any of the standard semantical notions. The problem here is not merely 'terminological'; rather, it occupies the core of philosophy of language. What is linguistic meaning? What do words and sentences really mean (given that their meanings are not primary or secondary intensions)? What is it to understand a language? How are we to solve Frege's Puzzle? How can a sentence express something that is both necessary and a posteriori? How can language be both public and anti-individualistic in nature? On the Jackson–Chalmers theory, these questions prove unanswerable.

2.2.3 *Understanding Language*

As indicated, Jackson and Chalmers also intend their theory to provide an account of what it is to understand a word, a sentence, and, more generally, a language. Put briefly, their view is that to understand an English sentence is just to know its A-intension (i.e., primary intension). As Jackson tells us: '[U]nderstanding the sentence only requires knowing the A-proposition' (1998: 77); '[U]nderstanding "Water covers most of the Earth" does not require knowing the conditions under which it is true, that is, the proposition it expresses [i.e., its C-intension]. Rather it requires knowing how the proposition expressed depends on the context of utterance—in this case, how it depends on which stuff in the world of utterance is the watery stuff of our acquaintance in it [i.e., its

[23] In Chalmers there is also a use/mention problem: standard quotation names (e.g., 'water') are, throughout, used ambiguously as names of both words and concepts; and, contrary to standard usage, concepts are said to have 'meanings' and 'references'.

A-intension]' (1998: 73–4).[24] But this account is surely mistaken. Consider the two-planet version of the twin-earth example: in that world, earth exists with all its inhabitants, languages, etc., and, in addition, so does a twin earth that is a macroscopic duplicate of earth except that XYZ fills the lakes and rivers, etc. Given that on the Jackson–Chalmers theory A-intensions (primary intensions) are defined on 'centered worlds', the words and sentences of Twin English would have the same A-intensions as they do in English. Moreover, Twin English would be syntactically and phonetically the same as English. It would then follow that in this envisaged world you already would understand, and know how to speak, Twin English.[25] I find this wholly counter-intuitive. In radio 'conversations' with your twin earth counterpart, each of you would seriously misunderstand the other's sentences. Suppose each of you simultaneously asserts 'Water contains hydrogen.' Then, each one would misunderstand the sentence he receives. After all, you know that in your language 'Water contains hydrogen' means *that water contains hydrogen*, but you wrongly believe that in your counterpart's language 'Water contains hydrogen' means *that water contains hydrogen*. (Analogously for your counterpart.) In fact, *no* sentence in your counterpart's language can even express this proposition! (And the other way round.) Still worse, not only does the present theory wrongly imply that your counterpart *does* understand your sentence, it also implies that even your most intimate friends *cannot* understand it! (Likewise for you counterpart's most intimate friends and his sentence.)

2.3 *Two-Dimensionalism, the Necessary A Posteriori, and Frege's Puzzle*

The primary tenet of the Jackson–Chalmers theory of the necessary a posteriori is that only sentences, never propositions, have this key property. But this is not so.

[24] Accordingly, Jackson tells us that his theory of understanding language 'can be put in Stalnaker's terminology by saying that understanding requires knowing the propositional concept associated with a sentence, though not necessarily the proposition expressed, and in Kaplan's by saying that understanding requires knowing character but not necessarily content' (1998: 72 n. 26). But, as Chalmers emphasizes, 'Kaplan uses his account to deal with indexical and demonstrative terms like "I" and "that", but does not extend it to deal with natural-kind terms such as "water", as he takes "water" to pick out H_2O in all contexts (the sound-alike word on Twin Earth is simply a different word)', (1996: 365–6 n. 25). (Kaplan seems clearly right.)

On the Jackson-Chalmers theory of linguistic understanding, not even your most intimate friend understands what you do by 'Water contains hydrogen'; only your twin-earth counterpart does!

[25] This is avoided if your words' primary intensions were *partial funcitons* defined for a centered world only if in that world you truly occupy the relevant location. Still, all the problems in sects. 2.3 and 2.4 (and the privacy problem just below) survive.

As I indicated earlier, to avoid the problem of logical omniscience (i.e., knowing every truth of logic if you know any), Jackson and Chalmers must incorporate structured propositions (ordered sets, labeled trees, or whatever). Jackson ought to accept this solution, for he himself endorses an analogous solution to a very similar problem (1998: 34). The idea is that, if a 'that'-clause denotes one of these structured propositions, the latter's 'constituents' would be either logical particles or intensions, associated in the obvious way with corresponding primitive expressions occurring in the embedded sentence. For example, 'that everything is even or odd' might denote the structured proposition: <universal generalization, <being even, disjunction, being odd>>. Universal generalization is the logical particle associated with 'everything'; being even, the primary intension (and also the secondary intension) of 'is even'; disjunction, the logical particle associated with 'or'; being odd, the primary (and also secondary) intension of 'is odd'.

Consider the following intuitive schemas: (i) If a sincere English speaker x utters an English sentence 'A' with literal, assertoric intent, then x thereby sincerely asserts that A; (ii) If x sincerely states that A, then x believes that A. Now, famously, Saul Kripke sincerely uttered 'It is necessarily true that Hesperus = Phosphorus' with literal, assertoric intent. So, by (i), Kripke thereby sincerely asserted that it is necessarily true that Hesperus = Phosphorus. Hence, by (ii), Kripke believed that it is necessarily true that Hesperus = Phosphorus. As history showed, this proposition which Kripke believed proved to be highly nontrivial. Like so many of us, Kripke once believed the negation of this proposition, and only with argument did he come to believe the unnegated proposition. What proposition is this? Within Jackson–Chalmers semantics there are two candidates: (a) the secondary intension of the embedded sentence 'It is necessarily true that Hesperus = Phosphorus' and (b) the primary intension of this sentence.[26] But neither option works.

(a) The secondary intension is the following structured proposition (where h is the constant function from worlds to Hesperus—i.e., to Phosphorus): <necessity, <h, identity, h>>. But this structured proposition is a triviality, so cannot be the one Kripke believed. (I think this would be Jackson's and Chalmers's assessment, for they take primary, not secondary, intensions to be the objects of the attitudes.)

(b) The primary intension of 'It is necessarily true that Hesperus = Phosphorus' is the following structured proposition (where h and p and are the

[26] I am supposing that hidden-indexicalism is not available (see Bealer (forthcoming)). Nor is treating 'water' as synonymous to the description 'the *actual* watery stuff of our acquaintance', because of various well-known difficulties.

primary intensions of 'Hesperus' and 'Phosphorus', respectively): <necessity, <h, identity, p >>. Unlike the secondary intension, this structured proposition is not a triviality. But this proposition is true iff the structured proposition <h, identity, p > is necessary. The latter structured proposition, however, is the primary intension of 'Hesperus = Phosphorus', and this primary intension is contingent, according to Jackson and Chalmers. It follows, therefore, that the former structured proposition must be false. Thus, given option (b), the proposition Kripke believed when he uttered 'It is necessarily true that Hesperus = Phosphorus' was false.

Now it is a truism that the sentence 'Hesperus = Phosphorus' is necessary iff the sentence 'It is necessarily true that Hesperus = Phosphorus' is true. Since all participants in the present debate (two-dimensionalists and their opponents alike) accept the left-hand side, they must accept the right-hand side as well: namely, that the sentence 'It is necessarily true that Hesperus = Phosphorus' is true. This, of course, is the reason Kripke chose to utter this sentence: he believed that it is true. Now when he uttered this sentence, what he said—namely, that it is necessarily true that Hesperus = Phosphorus—was true. And, since he was sincere, he believed this proposition. Therefore, the proposition he believed when he uttered 'It is necessarily true that Hesperus = Phosphorus' was true. But this contradicts the conclusion we reached in the preceding paragraph. So option (b) cannot be right, either.

Since neither option (a) nor (b) works, Jackson–Chalmers semantics is evidently unable to represent our sample sentence—and a great many other such sentences mixing modalities and attitudes—and, for this reason, it is evidently unable to deal successfully with one of the central phenomena of modal epistemology: the necessary a posteriori.[27]

[27] Someone might propose a more complicated two-dimensional semantics designed to avoid the above problem. It seems, however, that problems of the same general type will continue to recur. But even if this is not so, how far from commonsense semantics should one be willing to go? Incidentally, there is a quicker, but less rigorous, way to formulate the above problem. Kripke, and almost everyone else, accepts the following sentence: 'It is a posteriori that it is necessary that Hesperus = Phosphorus'. But if the second 'that'-clause denotes the secondary intension of 'Hesperus = Phosphorus', then the whole sentence is false because the proposition that this proposition is necessary is a priori (vs. a posteriori). Conversely, if the indicated 'that'-clause denotes the primary intension, then the whole sentence is again false because the primary intension is a contingent (vs. necessary) proposition. Put another way, when you utter 'Hesperus = Phosphorus' with sincere assertoric intent, what are you asserting and consciously and explicitly thinking? The necessary a priori secondary intension? The contingent a posteriori primary intension? Both?! The ordered-pair of the two?! None of these answers is acceptable. By the way, many of my criticisms in this section can be applied *mutatis mutandis* to the theory that intentional states have two kinds of content—'narrow' and 'wide'.

The underlying problem is, of course, that two-dimensionalism has no solution to Frege's puzzle. As we saw in section 2.1, a solution requires an explanation of how, for example, the propositions that Hesperus = Hesperus and that Hesperus = Phosphorus can be different. So (by generalization), a solution also requires an explanation of how the proposition that it is necessarily true that Hesperus = Hesperus and the proposition that it is necessarily true that Hesperus = Phosphorus can be different. Since, as we have just seen, two-dimensionalism cannot explain this difference, it has no general solution to Frege's puzzle. That is to say, two-dimensionalism cannot solve the first problem in philosophy of language. What is missing, of course, is a satisfactory theory of fine-grained propositions. If one is feasible (see sect. 2.5), however, the complicated apparatus of two-dimensional semantics is rendered superfluous.

2.4 *Two-Dimensionalism, Public Language, and Anti-Individualism*

In this section we will see that two-dimensional semantics runs into two other fundamental challenges in philosophy of language: namely, how to deal with the public and anti-individualistic character of language. A convenient way to bring this out is to examine Jackson and Chalmers's commitment to implicit descriptivism.[28] By implicit descriptivism, I mean the thesis that, if a Jackson–Chalmers primary intension (e.g., being-the-watery-stuff-of-our-acquaintance) were analyzed completely, the result would be a descriptive analysis. (It is understood that the envisaged analysis may be infinitary.) This implicit descriptivism is consistent with the thesis that, phenomenologically, primary intensions present themselves as nondescriptive simples. Since Jackson's commitment to implicit descriptivism is more straightforward (e.g., 1998: 40 n.), I will formulate the discussion in Jackson's idiom. (For Chalmers's commitment, see n. 29 below.)

[28] This implicit descriptivism plays the role in the Jackson-Chalmers account of the alleged 'a priori link' between the microphysical and macrophysical and between the physical and the mental, so if their implicit descriptivism is problematic, so is this account. (For another problem, see Bealer 1997, 2000.) Further, since Chalmers's implicit descriptivism also plays an essential role in his zombie argument, that argument will likewise be deficient, as we shall see in sect. 4.3. A note on terminology. In my remarks on the two-planet version of the twin-earth story in section 2.2.3, I noted that two-dimensionalism's account of linguistic understanding is rendered untenable by the fact that (as defined) the primary intensions of corresponding words in English and Twin English are identical. In n. 23 I indicated how this consequence could be avoided if the notion of primary intension were redefined in a certain way. In the present section, 'primary intension' may be understood in either the first or second way; the criticisms in this section apply either way.

2.4.1 *Platitudinous Conceptions*

According to Jackson, the primary intension of 'water' (i.e., being-the-watery-stuff-of-our-acquaintance) is encoded in our 'conception' of water. Hence, if this conception were converted into a Ramsified definition (i.e., an implicit-turned-direct second-order definition), the result would be a correct analysis of the original primary intension. Since this qualifies as a descriptive analysis, Jackson's view qualifies as an implicit descriptivism in the above sense.

What, for Jackson, is a 'conception'? It is a certain kind of theory: namely, one that is revealed by one's intuitions about possible cases. For example, in connection with his 'conception' of free action, Jackson tell us, '[M]y intuitions about possible cases reveal my theory of free action' (1998: 32). On one natural way of taking this, a conception is an aggregation of widely accepted platitudes (general truths)—for example, that water is the stuff of our acquaintance that is 'a clear potable liquid and all that; for short, being watery' (1998: 38). Though this is not Jackson's official way of taking 'conception', for the time being let us adopt it, for it is important to see that the resulting view (which many two-dimensionalists seem to accept) suffers from two serious problems. (a) Either it wrongly implies that natural language is not *public* in the way we know it is, or it leads to a mistaken account of what it is to *understand* a language. (b) It wrongly implies that platitudinous conceptions are *a priori* and, since in many cases they are not, that ordinary English speakers do not understand many everyday English words.

(a) The Public-Language Dilemma.[29] Suppose that there are two English-speaking communities c_1 and c_n who are 'joined' to one another by a chain of English-speaking communities $c_2, \ldots, c_{n-1}$, each of which has regular contact with its flanking communities, and with whom it has significant overlap in water platitudes (i.e., the theory-like general principles). But suppose that at the extremes c_1 and c_n (e.g. the English speakers on "Waterworld" and "Dune") have little or no overlap in their water platitudes. In this case, we would still want to say that they share the term 'water'. Indeed, if the two

[29] In effect, the argument of sect. 2.3.3 already demonstrates the public language dilemma. On the one hand, we saw that when primary intensions are defined in the original way, the resulting theory of linguistic understanding is mistaken, for it wrongly implies that every English speaker would already understand Twin English, and conversely. On the other hand, we saw that this unwanted consequence could be avoided by redefining primary intensions in accordance with the corrective given in n. 23. But given this definition, it trivially follows that primary intensions of every English speaker's words are private (not public), for these primary intensions are defined only in those centered worlds in which that speaker is the person at the center. The argument I am about to give (as well as the public-language argument in sect. 2.4.2) is independent of this earlier argument for the public-language dilemma.

groups got together, we can easily imagine how the conversations about water would go. In such conversations, the participants would mean and understand the same thing by their 'water' sentences (namely, what those sentences mean *in English*), and they would be communicating as well. But this would not be so on the present construal of 'conceptions', for their conceptions of water differ so radically.

There is, however, a natural fall-back position, namely, to base the implicit-turned-direct definition of water on the water platitudes *common to* the various communities of English speakers. The problem with this is that (at least in some worlds) hardly any water platitudes would be shared by all these communities, and, therefore, the resulting platitudes would not support an implicit-turned-direct definition that picks out water uniquely. Well, I take that back. The shared platitudes could support such a definition, but only if they included highly *deferential* platitudes (including platitudes concerning causal information chains). But this would only trigger the other horn of the dilemma. If the primary things you understand by 'water' are, for example, that it refers to the thing that English speakers refer to with 'water' (and/or other such deferential facts), then even though you are able to use the word passably in various conversations, you would not really *understand* it. Moreover, even though this fall-back position has many variations, as far as I can see they all either let in too much or exclude too much, making this an unsolvable problem for the Jackson–Chalmers picture (even when understood as in section 2.4.2).

(b) A Priori Conceptions and Linguistic Understanding. On Jackson's account of linguistic understanding (1998: 73 f.), an ordinary English speaker understands the English word 'water' only if he or she is in a position to know a priori the propositions making up our ordinary conception of water. So, if (as we are supposing temporarily) our ordinary conception consisted of widely accepted platitudes, we should be in a position to have a priori knowledge of those platitudes (or, at least, that a majority of them hold). But this knowledge is in fact a posteriori, not a priori. Why? Because *intuition* constitutes the evidential basis of a priori knowledge;[30] and, even if we initially have intuitions supporting the platitudes, we are readily disabused of them as soon as we come upon our intuitions that it could have turned out (i.e., it is epistemically possible) that, as a result of systematic illusions, water is not really clear or potable or . . . (for the majority of the platitudinous properties of water). Given this, it

[30] Except for stipulative knowledge, which is not relevant here, since we may assume that our English speaker (and, for that matter, *every* English speaker; see sect. 3.2 below) did not stipulatively introduce the word 'water' to English.

follows on the present theory that, since the requisite knowledge is not a priori, ordinary English speakers do not understand the English word 'water'.

2.4.2 *The Jackson–Chalmers Picture*

It is now time to see what happens when we take 'conception' in the way Jackson really intends—that is, as *the whatever it is* that our intuitions about relevant possible cases reveal. Understood this way, Jackson's picture and Chalmer's picture pretty much coincide.[31] When 'conceptions' are taken this way, however, there are two unacceptable consequences: (a) once again, language cannot be public in the way we know it is; (b) language cannot be anti-individualistic in the way we know it is.

(a) Public Language. Rather than establishing this point in detail, I will give an example indicative of the problem. It should be clear how to extend the argument to other examples. Consider Kripke's original meter-man. Suppose that he perished immediately after he stipulatively introduced his word 'meter'. Suppose that meter-man's stipulation was witnessed by someone, and that a causal naming chain flowed out from this person and eventually came to include utterances to which you have been party (but with only a deferential understanding). In this connection, suppose that the primary intension of your word 'meter' was fixed at the time solely by virtue of your being party to these utterances. When this primary intension is analyzed, the result is a description of a causal information chain descending from your exposure to these utterances through other people and terminating in a naming ceremony in which a name denoting one meter was stipulatively introduced. By contrast, when the primary intension of meter-man's word 'meter' is analyzed, the result is a description of a naming ceremony in which he himself is stipulatively introducing his word for one meter by reference to the length of a stick he is looking at. Since there are plainly centered worlds in which the denotations of

[31] How so? Well, Chalmers tell us that 'what makes an actual-world X *qualify* as the referent of "X"' is provided by an 'analysis of the primary intension [which is] an a priori enterprise' (1996: 59). Primary intensions just encode *all* the correct, determinate answers to 'questions about what our concept [*sic*] *would* refer to if the actual world turned out in various ways' (1996: 60). 'The true intension can be determined only from detailed consideration of specific scenarios: What would we say if the world turned out this way? What would we say if it turned out that way?' (1996: 57 f.) Thus, on Chalmers's picture (and Jackson's too), if no finitary descriptive analysis is available, then there always exists a default descriptive analysis consisting of a definite description 'the x such that . . .' formed from an infinitary (cf. Chalmers 1996: 84 n. 45) conjunction of conditionals in which (i) the antecedent characterizes some relevant possible case (scenario) from the perspective of the individual at the center of a corresponding centered world, and (ii) the consequent correctly identifies the indicated entity x with one of the entities described by that antecedent.

these two descriptions come apart, it follows that the primary intensions of meter-man's word 'meter' and your word 'meter' are different.[32] Similar considerations (i.e., a different description for each personalized mode of access) show that this conclusion generalizes to all speakers who have a word 'meter' in their vocabularies: the primary intension of each such speaker's word differs from that of meter-man's word (and, indeed, from everyone else's). Thus, meter-man's word is private not public.

(b) Anti-Individualism and Linguistic Misunderstanding. On the Jackson–Chalmers picture, the primary intension of, say, *your* word 'know' is uniquely determined by your 'conception of knowledge', where the latter is uniquely revealed by your intuitions about relevant possible cases (e.g., your intuition that it is, or is not, possible for a certain Gettier scenario to occur and the person in the case to have the relevant bit of knowledge; and so forth). On this picture, therefore, you could not fail to understand your word 'know'. To illustrate, suppose there are various people (perhaps you?) who resolutely insist that the Gettier examples are examples of knowledge. Then, as Jackson tells us, 'In these cases, it is . . . misguided to accuse them of error' (1998: 32); '[They are] right about what counts as knowledge in *their* sense' (1998: 36). Naturally, this generalizes from your word 'know' to your other words.

This picture, however, is mistaken, for it overlooks the phenomenon of anti-individualism, which is convincingly illustrated by Tyler Burge's (1979) arthritis example. When the patient in Burge's example sincerely utters 'I have arthritis in my thigh', he asserts and believes that he has arthritis in his thigh, which of course is impossible. (If the patient did not believe this, he would not be relieved if the doctor were to tell him that arthritis in the thigh is impossible because arthritis is a joint disease. But he plainly would be relieved!) Now suppose that right after his original assertion the patient had gone on to consider the question whether having arthritis in the thigh is possible, he would have had the intuition that it is—or, at least, he would have lacked the (correct) intuition that it is impossible. The Burgean explanation would be that the patient does not understand his concept, for if he did, he would instead have had the intuition that arthritis in the thigh is impossible. If this is right, as it surely is, it is clear the patient does not understand what he himself means when he utters

[32] Indeed, these two primary intensions do not even have the same *extension* in the centered *actual* world where meter-man is flagged as the person at the center and the time of the dubbing is flagged as the time at the center. Nor do these primary intensions have the same extension in the centered *actual* world where you are flagged as the person at the center and the time your path first crossed the aforementioned causal information chain is flagged as the time at the center. And these facts also generalize.

his word 'arthritis'; in other words, he does *not* understand his own word 'arthritis'. Such is the lesson of anti-individualism. But since, on the Jackson–Chalmers picture, it is impossible in the example for the patient not to understand his word 'arthritis', this picture cannot accommodate the fundamental anti-individualist nature of this and a large family of other natural-language expressions.[33]

2.5 *A More Satisfactory Framework*

Consider some truisms. The proposition that A & B is the conjunction of the proposition that A and the proposition that B. The proposition that not-A is the negation of the proposition that A. The proposition that Fx is the singular predication of the property F of x. The proposition that there exists an F results from existentially generalizing on the property F. And so on. These truisms tell us what these propositions are essentially: they are by nature conjunctions, negations, singular predications, existential generalizations, etc. These are rudimental facts that require no further explanation and for which no further explanation is possible. This was pretty much the dominant view on propositions in the history of logical theory.

By adapting techniques developed in the algebraic tradition in extensional logic, one is able to develop this nonreductionistic approach. Examples like those just given isolate fundamental logical operations—conjunction, negation, singular predication, existential generalization, etc. Intensional entities are then taken as *sui generis* entities; the aim is to analyze their behavior with respect to the fundamental logical operations.

How are we to integrate definite descriptions? On the Fregean approach, the singular term 'the F' refers to the unique item satisfying the predicate 'F' if there is one; if there is not, 'the F' has no reference, in which case the sentence 'The F Gs' lacks truth-value. To incorporate this intuitive theory, we consider a logical operation *the* (which is akin to the Frege–Church operation ι) associated

[33] The Jackson–Chalmers approach does not deal with a number of related challenges facing an account of linguistic understanding. (1) A characterization of overly deferential, incomplete understanding (e.g., as in Putnam's beech/elm case). (2) A solution to the circularity problem created by the fact that a speaker might fail to understand some of the *auxiliary* words or concepts involved in the characterization of relevant possible cases. (3) An account of other ways of failing to understand a word or concept besides the sort of incomplete understanding associated with deference, for example, the sort of misunderstanding created by the ubiquity of *false* information in one's web of belief. (4) An account of the phenomenon of mere local misunderstanding versus full-fledged misunderstanding. (The account in section 3 is designed to deal with these and other concerns.)

with the word 'the'. One may think of the values of *the* as 'individual concepts': *the*(F) would then be the individual concept of being the F. Consider the property of being G and the individual concept of being the F. What is the relation between them and the proposition that the F Gs? Not singular predication: when the operation of singular predication is applied to the property of being G and the concept of being the F, the value is the proposition that the concept of being the F Gs. This is a *very* different proposition! The relation of singular predication is thus not the relation holding between the property of being G, the concept of being the F, and the proposition that the F Gs. Rather, the relation holding between them is a quite distinct kind of predication, which may be called *descriptive predication*. Thus, the proposition that the F Gs is the result of descriptively predicating G of the concept that results from applying our operation *the* to F.

In this informal account, the entities playing the subject roles in descriptive predication are individual concepts (or properties). But we could instead think of them more generally as modes of access or modes of presentation (*Arten des Gegebenseins*). Besides these purely Platonic entities, certain *constructed* entities also present things to us. For example, pictures do. Certain *socially* constructed entities also function in this capacity; the most prominent are linguistic entities. Indeed, linguistic entities provide the only access most of us have to various historical figures. These linguistic entities have the important feature of being *public*, shared by whole communities. Names are one kind of linguistic entity that provide us with this kind of access to objects.[34]

This suggests that certain Frege-style puzzles may be dealt with by relying on such non-Platonic modes of presentation. (I will use double-quoted expressions to denote such names: "Hesperus", "Phosphorus", etc.) The object "Hesperus" presents = Hesperus = Phosphorus = the object "Phosphorus" presents. This is so despite the fact that these two non-Platonic modes are distinct (i.e., "Hesperus" ≠ "Phosphorus").

The key idea is that relevant logical operations should be defined for all modes of presentation, non-Platonic as well as Platonic. So, in particular, the

[34] In what follows, names will be understood, not as mere phonological or orthographic types, but as fine-grained entities whose existence is an empirical fact and for which it is essential that they name what they do. For example, Cicero the Illinois town and Cicero the famous orator share a name in the phonological or orthographic sense, but not in the fine-grained sense. In the latter sense, but not the former, the existence of the two names is an empirical matter: the name of the town is fairly new; the name of the orator is very old. Given that the name of the town exists, it is essential to it that it name the town; likewise, given that the name of the orator exists, it is essential to it that it name the orator. This conception meshes with Kripke's rigid-designator theory of names. See Kripke (1980: 8 n. 9), Kaplan (1989: 603 ff.), Bealer (1993*b*: 35 f) Fine (1994).

operation of descriptive predication may take as arguments, say, the property of being a physical object and the non-Platonic mode "Hesperus". The result of this descriptive predication would be a proposition. Likewise, for being a physical object and "Phosphorus". The point is that these non-Platonic modes—as opposed to descriptive properties obtained from them by means of *the*—are the arguments in these descriptive predications.

The resulting two propositions have the following salient features. First, as noted, they are distinct (since "Hesperus" and "Phosphorus" are distinct). Second, they are necessary, reflecting Kripke's doctrine that every physical object is *necessarily* a physical object. Third, they are distinct from all propositions expressible by the use of definite descriptions (with or without actuality operators); in other words, they have *no* descriptive content (except, of course, for that associated with the predicate 'physical object'). Finally, these non-Platonic modes of presentation are neither *in* nor *parts* of these propositions. These propositions are seamless; only in our logical analyses do these modes appear.

Taken together, these features are exactly the salient features of the proposition that Hesperus is a physical object and the proposition that Phosphorus is a physical object. This suggests the hypothesis that they *are* these propositions. If this is correct, the analogous thing will hold for identity propositions of the sort responsible for Frege's puzzle itself. For example, the logical analysis of the proposition that Hesperus = Phosphorus will be a descriptive predication involving both "Hesperus" and "Phosphorus", whereas the logical analysis of the proposition that Hesperus = Hesperus will only involve "Hesperus". This ensures that the former proposition is nondescriptive, necessary, and a posteriori, while the latter is nondescriptive, necessary, but a priori. We thus have a provisional solution to this instance of Frege's puzzle.

I do not say that all instances of Frege's puzzle (or even the above instance) have exactly this style of solution. What I do claim (Bealer, forthcoming) is that, by using this and certain other techniques provided by the nonreductive algebraic framework, all instances have solutions. If so, theories that deny the existence of necessary a posteriori propositions would be baseless.

3 Rationalist Modal Epistemology

With an adequate logical framework in place, we are now ready to outline a moderate rationalist account of modal epistemology. Central to this account of our modal knowledge is an analysis of what it is to understand one's concepts, which will be sketched in section 3.1. This account, in turn, leads in section 3.2 to an account of our a priori knowledge of *categories*, an account that can then

be used to explain our knowledge of a posteriori necessities of the kind discussed by SE. Finally, these accounts of the source and justification of our modal knowledge will be applied in section 3.3 to the question of modal error.

3.1 *Moderate Rationalism and Understanding Our Concepts*

I argued (Bealer 1992) that intuitions are in fact evidence. And in Bealer 1987 and 1996 I argued that intuition is in fact a *basic* source of evidence, and that the only satisfactory explanation of this fact is provided by *modal reliabilism*: the doctrine that something counts as a basic source of evidence iff there is an appropriate kind of modal tie between its deliverances and the truth. Modal reliabilism implies a form of moderate rationalism and an associated autonomy of the a priori disciplines. But why should the indicated tie to the truth exist? The answer is provided by the analysis of what it is to understand one's concepts. Giving this analysis creates an exegetical dilemma—either to give a very compressed sketch or to provide a lengthy exegesis and accompanying justification. In this setting, I have opted for the former, being all too aware of its shortcomings.[35]

There are two senses in which a subject can be said to possess a concept. First, a weak nominal sense:

> A subject possesses a given concept at least nominally iff the subject has natural propositional attitudes toward propositions that have that concept as a constituent content.

Possessing a concept in this sense is compatible with what Tyler Burge (1979) calls misunderstanding and incomplete understanding of a concept ('misunderstanding' for cases where there are errors in the subject's understanding of the concept, and 'incomplete understanding' for cases where there are gaps), and it is compatible with having concepts partly in virtue of the mere attribution practices of third-party interpreters. Second, there is a robust sense of concept possession, which requires *understanding* the concept:

> A subject understands a concept iff (i) the subject at least nominally possesses the concept and (ii) the subject does not do this with misunderstanding or incomplete understanding or merely by virtue of satisfying our attribution practices or in any other such manner.

[35] I give a more complete discussion in Bealer 1987 and 1999, and the full presentation in my forthcoming book *Philosophical Limits of Science*.

I will use the technical term 'determinately understand a concept' for this kind of concept possession.

Examples provide the basis for our analysis. Here is an illustration (see Bealer 1987). Suppose that in her journal a sincere, wholly normal, attentive woman introduces *through use* (not stipulation) a new term 'multigon'. She applies the term to various closed plane figures having several sides (pentagons, octagons, chiliagons, etc.). Suppose her term expresses some definite concept—the concept of being a multigon—and that she determinately understands this concept. By chance, she has neither applied her term 'multigon' to triangles and rectangles nor withheld it from them; the question has just not come up. Eventually, however, she considers it. Her cognitive conditions (intelligence, etc.) are good, and she determinately understands these concepts. Suppose that the property of being a multigon is either the property of being a closed, straight-sided plane figure, or being a closed, straight-sided plane figure with five or more sides. (Each alternative is listed under 'polygon' in my desk *Webster's*.) Then, intuitively, when the woman entertains the question, she would have an intuition that it *is* possible for a triangle or a rectangle to be a multigon if and only if being a multigon = being a closed, straight-sided plane figure. Alternatively, she would have an intuition that it is *not* possible for a triangle or a rectangle to be a multigon if and only if being a multigon = being a closed, straight-sided plane figure with five or more sides. That is, the woman would have *truth-tracking* intuitions. If she did not, the right thing to say would be that either the woman does not really understand one or more of the concepts involved, or her cognitive conditions are not really good.

Our judgments about other relevant examples also fit this pattern. Naturally, these judgments are defeasible, given that all sorts of incidental factors might in misleading ways affect a person's dispositions to have such truth-tracking intuitions. Consistent with this, however, are the following ideas. The person (or an appropriate epistemic *Doppelgänger* of the person) would have *ever more reliable* intuitions if his cognitive conditions (intelligence, etc.) were to improve and his auxiliary conceptual repertory were to enlarge (given, of course, that in the process of these developments there is no shift in the way the person understands his original concepts or propositions involving them). Thus, as the person's cognitive conditions improve and his auxiliary conceptual repertory enlarges, the degree to which the person's intuitions are truth-tracking will ever more accurately reflect how well the person understands the relevant concepts. This suggests that determinate understanding can be explicated in terms of the associated metaphysical possibility of this sort of truth-tracking intuition: determinate understanding is that mode of understanding that

constitutes the categorical base of this possibility.[36] I will work up to the final analysis in stages.

3.1.1 *A Priori Stability*

Suppose x understands a given proposition p in some mode m (determinately, indeterminately, etc.). (For brevity, I will say that x understands p m-ly.) Then, I will say that x settles with *a priori stability* that p is true iff, for cognitive conditions of some level l and for some conceptual repertory c, (1) x has cognitive conditions of level l and conceptual repertory c, and x attempts to elicit intuitions relevant to the question of whether p is true, and x seeks a theoretical systematization based on those intuitions, and that systematization affirms that p is true, and all the while x understands p m-ly, and (2) necessarily, for cognitive conditions of any level l' at least as great as l and for any conceptual repertory c', which includes c, if x has cognitive conditions of level l' and conceptual repertory c', and x attempts to elicit intuitions bearing on p and seeks a theoretical systematization based on those intuitions, and all the while x understands p m-ly, then that systematization also affirms that p is true.[37] In other words, once x achieves cognitive conditions l and conceptual repertory c, theoretical systematizations of x's intuitions always yield the same verdict on p as long as p continues to be understood m-ly throughout. That is, p thereafter always gets settled the same way.

Using this notion of a priori stability we arrive at the following candidate analysis:

> determinate understanding = the mode m of understanding such that, necessarily, for all x and property-identities p understood m-ly by x, p is true iff it is possible for x to settle with a priori stability that p is true.[38]

[36] Some people have objected that relying in this way on intuition's tie to the truth is unacceptable, for it simply amounts to invoking a 'dormative virtue'. The analogy fails, however, for in the present context the explanandum is a modal fact—i.e., intuition's qualified *necessary* tie to the truth. And necessities call for a very different sort of explanation from that called for by contingencies. In the explanation of necessities, it is wholly appropriate to articulate essences, and it is of the essence of determinate understanding of concepts that intuitions involving those concepts be correct—modulo suitably good cognitive conditions, notably intelligence. This is compatible with its being of the essence of intelligence to have the complementary property. In fact, this complementarity is paradigmatic of functionally definable sets of basic (i.e., non-Cambridge) properties.

[37] When I speak of higher-level cognitive conditions, I do not presuppose that there is always commensurability. In order for the proposal to succeed, I need only consider levels of cognitive conditions l' and l such that, with respect to *every* relevant dimension, l' is at least as great l.

[38] By property-identities p, I mean the following. Suppose a primitive predicate 'F' expresses a given concept. Then the associated property-identities p are propositions expressible with sentences of the form 'The property of being F = the property of being A', or the denials of such sentences (where A is some possible formula). The reason for the restriction to property-identities is to avoid certain potential counter-examples to the definition (and, in turn, to moderate rationalism).

The sufficiency claim in this biconditional is a *correctness* property. The necessity claim is a *completeness* property. The correctness property tells us about the potential *quality* of x's intuitions: it is metaphysically possible for x to get into a cognitive situation such that, from that point on, theoretical systematizations of x's intuitions yield only the truth regarding p, given that x understands p m-ly throughout. The completeness property tells us about the potential *quantity* of x's intuitions: it is metaphysically possible for x (or a qualitative epistemic counterpart of x) to have enough intuitions to reach a priori stability regarding the question of p's truth, given that x understands p m-ly throughout.

3.1.2 *Scientific Essentialism*

As it stands, the completeness clause clashes with scientific essentialism, which tells us that there are property-identities that are essentially a posteriori. For example, this clause wrongly requires that x be able to settle with a priori stability that the property of being water = the property of being H_2O. To solve this problem, we need to weaken the completeness clause so that it requires only *categorial mastery*. The easiest way to do this is to replace the original completeness requirement with the weaker requirement that it merely be possible for x to settle with a priori stability that there could exist some true proposition that is a twin-earth counterpart of p.[39] Thus:

> determinate understanding = the mode m of understanding such that, necessarily, for all x and property-identities p understood m-ly by x,
>
> (a) p is true *if* it is possible for x to settle with a priori stability that p is true.
> (b) p is true *only if* it is possible for x to settle with a priori stability that p has a counterpart that is true.

Note that, if a test proposition p is *semantically stable*, it is entirely immune to scientific essentialism. Consequently, for semantically stable property-identities, the weakened completeness clause entails the strong completeness clause of the earlier analysis, which in turn implies a form of moderate rationalism and autonomy of the a priori disciplines.

3.1.3 *Anti-individualism*

This weakening of the completeness clause, however, creates a problem concerning the *noncategorial* content of our concepts. Suppose x has mere categorial mastery of a certain pair of concepts—say, the concept of being a beech and the

[39] For example, if p is the proposition that being water = being H_2O, x would need to be able to settle with a priori stability that there could be a twin-earth relative to which there is a true proposition that is the counterpart of the proposition that being water = being H_2O.

concept of being an elm. None the less, x might not be able to determine whether trees are beeches or elms, no matter how long and carefully he studies them. In this case, x certainly would not understand these concepts determinately (although the above analysis implies, wrongly, that he would). His 'web of belief' would be too sparse. What x needs is, roughly, enough information to 'begin doing the science' of beeches and elms.

We can resolve this difficulty by making use of the idea of *truth-absorption*. If x has categorial mastery of certain of his concepts but none the less does not understand them determinately, then, by absorbing ever more true beliefs, x will eventually switch out of his deficient mode of understanding and come to understand those concepts determinately. By contrast, people who already understand their concepts determinately can always absorb more true beliefs without switching out of their determinate understanding. This suggests the following:

> determinate understanding = the mode m of understanding such that, necessarily, for all x and all p understood m-ly by x,
>
> (a) p is true *if* it is possible for x to settle with a priori stability that p is true.
>
> (b.i) p is true *only if* it is possible for x to settle with a priori stability that p has a counterpart that is true. (for property-identities p)
>
> (b.ii) p is true *only if* it is possible for x to believe m-ly that p is true. (for p believable by x)[40]

The reason why this analysis is successful is that, absent intuition, one's web of belief is the default basis on which determinateness rides. Where there is the possibility of a priori intuitions, however, they are determinative.

Thus, to understand a proposition determinately is to understand it in a certain mode. What distinguishes this mode from other natural modes of understanding are three essential properties:

(a) correctness
(b.i) categorial completeness
(b.ii) noncategorial completeness

(a) A mode m has the correctness property iff, necessarily, for all individuals x and all propositions p that x understands in mode m, p is true *if* it is possible for x (or someone starting out in qualitatively the same sort of epistemic situation as x) to settle with a priori stability that p is true, all the while understanding p in mode m.

[40] Perhaps 'believe' should be strengthened to 'rationally believe', and p restricted to propositions that x can rationally believe.

(b.i) A mode m has the categorial completeness property iff, necessarily, for all individuals x and all true (positive or negative) property-identities p which x understands in mode m, it is possible for x (or someone starting out in qualitatively the same sort of epistemic situation) to settle with a priori stability that there exists some true twin-earth-style counterpart of p, all the while understanding p in mode m.

(b.ii) A mode m has the noncategorial completeness property iff, necessarily, for all individuals x and all true propositions p that x understands in mode m and that x could believe, it is possible for x to believe p while still understanding it in mode m.

3.2 *Categories*

How does one justify the step from the empirical information that all and only samples of water are samples of H_2O (i.e., from the coextensiveness of water and H_2O) to the *modal* conclusion that, necessarily, water = H_2O? One bridges this 'modal gap' by combining the empirical information with *intuitions about hypothetical cases*—for example, twin-earth cases (see Bealer 1987).[41] What accounts for these intuitions? The analysis (just given) of what it is to understand one's concepts provides an answer. But in the case of these crucial intuitions still more light can be shed by an idealized rational reconstruction (ibid.).

The first step is to divide our concepts as follows: (1) category concepts (predication, number, identity, property, relation, proposition, quality, quantity, stuff, compositional stuff, functional stuff, etc.); (2) content concepts (phenomenal concepts and concepts of psychological attitudes); (3) naturalistic concepts. What distinguishes the concepts in the first two categories from those in the third is that they are *semantically stable*. The idea is that our mastery of these semantically stable category and content concepts is what drives the concrete-case intuitions (twin earth, etc.) that bridge the modal gap.

Consider the following categorial principle: if a sample of a given purely compositional stuff has such-and-such composition, then, necessarily, all other samples of that purely compositional stuff also have that composition. This principle is semantically stable, so the analysis of what it is to understand our concepts tells us that if we determinately understand the concepts involved in the principle, then our intuitions concerning the principle will be reliable. Accordingly, the principle can be known a priori.

[41] Not by the implicit descriptivist route of Jackson and Chalmers. Incidentally, a dialectically critical datum in the mind–body debate is that we simply lack crucial psychophysical analogues of these twin-earth intuitions (see Bealer 1994).

Now by simple universal instantiation on this principle, we get the following principle: if water is a purely compositional stuff, then if a sample of water has such-and-such composition, then, necessarily, all other samples of water also have that composition. But this principle in turn implies the following: if water is a purely compositional stuff, *then* if all and only samples of water here on earth have such-and-such composition, then if there were a twin-earth that is macroscopically like earth but on which the samples corresponding to the samples of water on earth had composition so-and-so (≠ composition such-and-such), those samples would not be samples of water. Instantiating on 'such-and-such composition' and 'composition so-and-so', we obtain: if water is a purely compositional stuff, *then* if all and only samples of water here on earth have H_2O as their composition, then if there were a twin earth that is macroscopically like earth but on which the samples corresponding to the samples of water on earth had XYZ (≠ H_2O) as their composition, those samples would not be samples of water. The consequent of this principle (following '*then*') states the chief twin-earth intuition used in the defense of SE.

But what about the antecedent of this principle: namely, that water is a purely compositional stuff? This principle is itself necessary, but how is it known? A simple answer would follow if we were to make the simplifying assumption that 'water' was originally introduced by means of a Kripkean baptism, in which the baptizer picked out water by, among other things, identifying its category: namely, purely compositional stuff. On this oversimplified account, the baptizer would then be in a position to know a priori that, if it exists, water is a purely compositional stuff. This a priori knowledge would be an instance of the kind of a priori knowledge one gets when giving any stipulative definition. Now, through use, the term 'water' enters the vocabulary of other speakers, so that whoever *determinately understands* the term will then be in a position to elicit a priori intuitions supporting the thesis that water is a purely compositional stuff. By this process, the baptizer's original a priori knowledge is transmitted to other speakers. (A similar story holds for 'H_2O'.)

Of course, to come to know the *specific* SE necessity that all and only samples of water are samples of the purely compositional stuff H_2O (and therefore that, necessarily, water = H_2O), something more is needed. We must supplement the above a priori knowledge with the a posteriori scientific knowledge that all and only samples of water here on earth are samples of the purely compositional stuff H_2O. This blending of a priori knowledge and a posteriori knowledge is what accounts for knowledge that is both necessary and a posteriori.

Next we remove the oversimplifications in this picture. One obvious source of oversimplification is the fact that the baptism story is wholly implausible in the case of a common noun so central to practical life as 'water'. But let us

maintain this fiction a moment longer, for the main source of oversimplification lies elsewhere: namely, in the fact that 'water' was successfully put into use long before anyone had any idea that various standard samples of water were composed of a single purely compositional stuff. So it is implausible that any baptizer would (on pain of vacuity) have restricted the reference of 'water' to a purely compositional stuff: for all the baptizer knew, there was no such stuff. But still, in fixing the reference of 'water', our (fictitious) baptizer would have needed to invoke *some* relevant categorial concept, if only to distinguish water from, say, the *functional stuff* drink and the water-like *macroscopic stuff* (whose actual extension is the same as the actual extension of the purely compositional stuff water), and so forth. The way to do this would have been by means of an implicit categorial concept that, when analyzed, is equivalent, not to the fundamental concept of being a compositional stuff, but to an ordered conjunction of default categorial conditionals. For example, the concept of being the stuff S such that (1) if all and only samples of S in our acquaintance are samples of some purely compositional stuff, then S is that purely compositional stuff; (2) if there is no such purely compositional stuff and if, instead, all and only instances of S in our acquaintance are instances of some not too complicated impure compositional stuff (akin to the impure compositional stuff jade), then S is that impure compositional stuff; (3) if there is no such pure or impure compositional stuff and if, instead, all and only instances of S in our acquaintance are instances of some not unwieldy macroscopic stuff, then S is that macroscopic stuff; and so forth. (This illustration is merely heuristic.)

Suppose, then, that our baptizer introduced the term 'water' by means of some such implicit categorial concept. By virtue of this, the baptizer would know a priori that water falls under that concept. Then the baptizer would still be in a position to have all the pro-SE twin-earth intuitions that we are trying to explain. For our baptizer would be in a position to know a priori that water satisfies the conjunction of default conditionals. And, much as before, the pro-SE intuitions are immediate logical consequences of this analysis plus the relevant corresponding general categorial principles. In particular, the hypothesis that all and only samples of water in our acquaintance are samples of a purely compositional stuff is an instance of the antecedent of the first conditional in the conjunction of default conditionals and, at the same time, the antecedent of the twin-earth conditional.

Now, just as in the simpler setting, the baptizer's more refined a priori categorial knowledge about water (i.e., the knowledge that, when analyzed, is equivalent to the ordered conjunction of default conditionals) can be transmitted through use to other speakers—specifically, to those who have come not just to use the word 'water', but to understand it determinately. Then these

speakers would likewise be in a position to have all the indicated pro-SE twin-earth intuitions. The rest follows *mutatis mutandis*.

Finally, let us remove the fiction that the term 'water' was introduced by a Kripkean baptizer. This is an unrealistic picture of how common nouns (in this case, naturalistic mass terms) typically function in natural language; rather, the community of speakers over time repeatedly reaffirms—and sometimes refines—the conventions governing the use of such terms. This process of reaffirmation and revision determines the implicit a priori categorial content of such terms, and having the potential (in relevantly good cognitive conditions) for a priori knowledge of that categorial content is a necessary condition of a speaker's determinate understanding of such terms (as was made clear in our analysis of determinate understanding).

The above picture may be generalized to all other terms to which SE is applicable. Moreover, the account may be used to explain further features of our intuitions—or, more accurately, our lack of intuitions—concerning natural kinds. For example, it explains why we lack intuitions that, *per impossibile*, would underwrite a priori knowledge of necessities that SE deems to be a posteriori.[42] For example, why do we lack an intuition that it is metaphysically impossible for there to be a sample of water containing no hydrogen? Our rational reconstruction provides an answer. *The general categorial principles that would underwrite such natural kind intuitions intuitively do not hold.* Specifically, the proposition in this water-without-hydrogen example is an instance of the following general categorial principle: for all purely compositional stuffs W and U, it is metaphysically impossible for there to be a sample of W that contains no U. But this general categorial principle is intuitively false. According to our rational reconstruction, to have an intuition that it is metaphysically impossible for there to be a sample of water containing no hydrogen, that intuition would need to be underwritten by this (or a kindred) general categorial principle concerning purely compositional stuffs. Since it is not, we lack such an intuition. So goes the explanation.

I will close with a question about modal error, which I will then try to answer in section 3.3. Consider the following false modal proposition: it is metaphysically possible for there to be a puddle of water containing no hydrogen. This proposition is an instance of the following general categorial principle: for all purely compositional stuffs W and U, it is metaphysically possible for there to be a sample of W containing no U. But, like its instance, this categorial principle is intuitively false. (On the contrary, the following general

[42] I certainly lack such intuitions. If someone reports having them, this would be an instance of the phenomenon considered in my example (4) in Bealer 1998: 28.

categorial principle is intuitively true: it is metaphysically possible for there to be compositional stuffs W and U such that, necessarily, every sample of W contains U.) Because the cited general categorial principle is intuitively false, the above explanation scheme predicts that we should *not* have an intuition that it is possible for there to be a puddle of water containing no hydrogen. The problem is that, prior to learning of twin-earth and the other pro-SE examples, a great many of us *did* have this possibility intuition and many others like it. How can our intuition have been erroneous in this way?

3.3 *Modal Error II: Categorial Misunderstanding*

Stephen Yablo (1993) presents an account of modal error—specifically, how anti-SE modal intuitions can be in error. (Yablo's discussion is stated in the idiom of 'conceivability' and 'inconceivability'; I will be reformulating it in what follows in the idiom of modal intuition, as defended in section 1.2.) Yablo is not concerned with errors resulting from conceptual illusions, limitations on intelligence, inattentiveness, and so forth. Nor is he concerned with the allegedly contradictory intuitions that exercised Kripke (e.g., the intuition that the identity of Hesperus and Phosphorus is necessary if true and the intuition that it could have turned out either way whether Hesperus = Phosphorus). We saw in section 1.4 that Kripke's worry dissolves as soon as we remember the distinction between the 'could'-of-metaphysical-possibility and the 'could'-of-qualitative-epistemic-neutrality and apply our little 'translation' test. Yablo's underlying concern is rather with full-fledged errors in intuitions about metaphysical possibility that arise in the context of scientific essentialism. Yablo holds that these errors have two potential sources, in each case mistaken *beliefs*: (a) mistaken a posteriori beliefs (e.g., someone who mistakenly believes that Hesperus ≠ Phosphorus might have the intuition that Hesperus could outlast Phosphorus) or (b) mistaken beliefs regarding the relationship between such a posteriori beliefs and associated modal truths (someone might deny that, if Hesperus = Phosphorus, then necessarily Hesperus cannot outlast Phosphorus). I am here less interested in class (a), for practiced dialecticians have the ability to proceed using exclusively 'pure' a priori intuitions: namely, those that survive even under the hypothesis that such a posteriori beliefs (both pro and con) are unjustified or mistaken.[43]

[43] In fact, by exercising this ability in the context of pure a priori philosophizing, one's natural kind intuitions will actually diminish, thereby all but eliminating disagreements of the sort associated with class (a).

How do people come to have erroneous modal intuitions belonging to the second class? Yablo's answer is that they are somehow produced by underlying class (b) beliefs, which, by hypothesis, are false. But it is plausible that the preponderance of such class (b) beliefs, albeit false, would at least be *justified*, and, as we know, such justification should ultimately be a matter of intuitions—presumably, intuitions about relevant concrete cases (twin-earth, etc.). But since a person's class (b) beliefs, which are justified by these concrete-case intuitions, are, by hypothesis, false, presumably a number of these justifying intuitions must themselves be false. What explains why these justifying intuitions go wrong? If the explanation is that they too are produced by false class (b) beliefs, we risk going in a circle. It seems, therefore, that we need something besides, or at least in addition to, Yablo's belief-based explanation of class (b) intuition errors.

Suppose two empirically well-informed, dialectically skilled philosophers have conflicting concrete-case SE intuitions (twin-earth, etc.). For example, suppose Putnam has the intuition that in his twin-earth example the samples of XYZ would not be water, whereas Carnap has the contrary intuition. One of them is in error. How are we to explain this error without going in a circle?

Our analysis of what it is to understand a concept determinately (and, specifically, the role of categorial mastery in that analysis) provides the missing pieces. One candidate explanation using these ideas is that either Putnam or Carnap simply does not determinately understand—indeed, misunderstands—the concept of being water. In some cases, this is no doubt the right explanation, but surely not in the case of Hilary Putnam or Rudolph Carnap. Certainly, *they* do not misunderstand the everyday concept of being water! For them, a subtler explanation is therefore needed. What has happened is that Carnap is the subject of a *local* categorial misunderstanding of the concept; that is, his categorial mastery of the concept is locally disrupted. An example will help to explain what I mean.

A student of mine, musing about prime numbers, realized that he did not know whether negative integers could be prime or, indeed, whether in the definition of prime number the domain is restricted to natural numbers or whether it includes all integers. Fortunately, he had a firm intuition that primes are divisible only by themselves and one, and he had the intuition that every negative integer, $-n$, is the product of itself and the number one and is also the product of n and -1, from which he inferred that negative integers cannot be prime. Then he had the intuition that 3 is prime but that $3 = 1 \times 3$ and $3 = -1 \times -3$, from which he concluded that only natural numbers were permitted in the definition of prime. In view of this performance, the student plainly *understood* the concept of being prime all along; specifically, he had full categorial mastery

of it. What went wrong early on was that he suffered a *local* lapse in his categorial mastery. His a priori (dialectical) process, however, was able to correct this lapse, therein manifesting his categorial mastery of the concept.

Carnap is in a somewhat similar situation. Not only does he have the intuition that, on twin earth, samples of XYZ would be water, he also would (if asked) have the general categorial intuition that water is a macroscopic stuff (individuated by its macroscopic properties). But this categorial misunderstanding is (we may suppose) only local: it is correctable by the a priori (dialectical) process—specifically, by careful examination of further cases, say, *other* sorts of twin-earth cases (e.g., the diamond and cubic zirconium twin-earth case[44]), and by systematization of the results. That is, left to his own a priori devices, Carnap would in the fullness of time become a scientific essentialist.

The point is that, at least in a large family of cases, the quality of one's categorial mastery holds the key, not only to the correctness of one's intuitions, but also to their *incorrectness*; furthermore, whether or not that mastery has lapsed only locally is the key to whether or not it is correctable a priori. And this, of course, is the solution to the problem at the close of the previous section. For, presumably, if Carnap has the intuition that in the twin-earth example the samples of XYZ would be water, he would likewise have the intuition that it is possible for there to be a puddle of water containing no hydrogen. The source of the intuition error in each case is the same local categorial misunderstanding.

The larger moral of this discussion is thus that, besides Yablo's class (a) and class (b) belief-based errors, there are two other classes: (c) those resulting from local categorial misunderstanding and (d) those resulting from out-and-out categorial misunderstanding.

4 Modality and the Mind–Body Problem

In this closing section, I shall examine the bearing that the above formulation of moderate rationalism has on modal arguments in philosophy of mind.[45]

4.1 *Hill's Critique and Categorial Misunderstanding*

Christopher Hill (1997) gives a new style of criticism of the familiar modal arguments (e.g., Kripke's) against the Identity Theory. His idea is that the

[44] The diamond-appearing samples on twin earth are samples of cubic zirconium (the comparatively cheap material from which fake diamonds are commonly made on earth). Would Carnap really have had the intuition that those samples are diamonds?!

[45] See also the closing paragraph of sect. 1.3.

possibility intuitions used in these anti-materialist arguments belong to a large, general class of possibility intuitions that are united by a shared form of psychological explanation and that are typified by error-prone intuitions of the kind we have just been considering (e.g., that it is possible for there to be water without hydrogen, heat without molecular motion, etc.). Because of this shared pathology Hill concludes from this that the possibility intuitions upon which these anti-materialist arguments are based lose their evidential force and that, consequently, those arguments do not go through.

But in section 3.3 we saw that, even though this type of possibility intuition (water without hydrogen, etc.) can be subject to error resulting from local categorial misunderstanding (as in the Carnap example), the error is typically correctable a priori (as long as the concepts involved really are understood). Indeed, such intuitions typically just disappear when one submits them to the sort of pure a priori dialectical process mentioned early in section 3.3 (and n. 43). So, for those who have already gone through this a priori process, the threat of this kind of anti-SE modal error pretty much disappears, at least for mainstream SE concepts (water, heat, gold, etc.) and concepts closely akin to them. Because intuitions in this family are thus not subject to any in principle pathology, they are immune to Hill's attack, and for those of us with this kind of SE dialectical background, it is perfectly legitimate to rely on them evidentially. Furthermore, in connection with our overall theme of moderate rationalism, the specific kind of pathology involved in these pre-SE style errors pertains entirely to semantically unstable intuitions and is not found in semantically stable modal intuitions. (This, of course, is not to say that the latter intuitions are immune to error.)

Despite this happy outcome, however, there is still an SE worry concerning certain anti-materialist modal intuitions with which Hill is concerned.

4.2 *Yablo's Disembodiment Argument*

Stephen Yablo (1990) gives an argument for the metaphysical possibility of his own disembodiment. The argument proceeds in three steps. First, he elicits the intuition that his disembodiment is indeed possible. Second, he uses his belief-based account of modal error to argue that, dialectically, anyone wishing to undermine the evidential force of this intuition must show that it has been corrupted either by a false class (a) belief or by a false class (b) belief. Third, he argues that his materialist opponents cannot do this without violating the presumption that an intuition is correct unless an independent reason for doubting it can be given.

I will make two points. First, many people report not having any intuitions like Yablo's, and they report having the opposite intuitions. Moreover, many

people who do have intuitions like Yablo's find themselves unable to shake their suspicions about them. For these reasons, one should hardly expect this disembodiment argument to produce many converts; to win converts, anti-materialists will need to provide additional argument.

Second, since 'I' is paradigmatically semantically *unstable*, so is the intuition report 'I could exist as a purely mental being'. So the usual SE question is then apt: what category of thing is 'hit' by use of the semantically unstable term 'I'? A *person*, to be sure. But consider a parallel case: although we know that 'water' hits a stuff, SE teaches us that one can (as in the Carnap case) be locally in the dark about what *subcategory* of stuff is hit (compositional, functional, macroscopic, etc.; see section 3.2). So, too, we could be locally in the dark with 'I': what prevents the subcategory from being that of *essentially embodied person* (at least for some utterances of 'I')? SE seems to open up this threat, and something needs to be done to close it.

We know that, at least initially, intuitions involving semantically unstable concepts are often prone to errors of the sort produced by local categorial misunderstandings and that, as a result, the presumption of correctness should at that point be suspended for such intuitions. Nevertheless, we also saw that this presumption should be restored for intuitions involving (concepts closely akin to) mainstream SE concepts as soon as one has gone through the relevant a priori process (twin-earth examples, etc.). Now, given that we have already gone through this process for mainstream SE concepts, is there, as a result, a presumption of correctness for our 'I' intuitions?

Consider an analogy. Unlike water's categorial profile, fire's remains puzzling (at least to me), despite our SE background. Is fire a process (akin to digestion and photosynthesis), a stuff (as the ancients seemed to think), a state (much as liquid is a state of matter), or what? The problem is not, I gather, a lack of relevant empirical information. In view of this, it is at least plausible that many (most?) of us suffer from some kind of local categorial misunderstanding or incomplete understanding of the concept of fire. This seems to provide reason for withholding (at least for now) the presumption of correctness from various modal intuitions about fire. Can there be fire without there being some physical object that is burning? Many people have the negative intuition. Others positive. (Presumably, some ancients would have been among the latter, for if fire were really an element, fire without any corresponding burning physical object ought to be possible.)

Now it is not implausible that we are in a somewhat similar situation with respect to the subcategory of 'I'. If this is right, our dialectical situation differs from that which Yablo describes. Specifically, at least for now, the presumption of correctness is rightly questioned for the pivotal semantically unstable

intuition that I could exist as a purely mental being, and for this reason, Yablo's disembodiment argument at least supplemental support.[46]

4.3 *Two-Dimensional Modal Arguments against Materialism*

4.3.1 *Chalmers's Zombie Argument*

Whereas the possibility of disembodiment would undermine the thesis that physical properties are substantive necessary conditions for mental properties, the possibility of zombies would undermine the thesis that they are sufficient conditions. But only the right sort of zombies would have this effect: namely, zombies that are *perfect physical duplicates* of us—that is, zombies that have *exactly the same* physical properties as we have. Of course, the weak zombie intuition that there could be zombies that, physically, are *functional* duplicates refutes various forms of materialistic functionalism (e.g., that of Lewis, Harman, Shoemaker). But this intuition does not refute traditional materialism generally; in particular, it does not refute materialism of the traditional *matter-chauvinist* variety (espoused by certain Identity Theorists and also by brute-supervenience advocates). These 'right wing' materialists (as opposed to their more liberal-minded functionalist cousins) believe that, amongst the various arrays of physical properties which, physically, are functionally equivalent to one another, only certain arrays (perhaps even some unique array) are sufficient conditions for our actual mental properties. What makes the difference is the kind of matter involved: it must be outright identical to the kind of matter we have in the actual world.[47]

[46] There is another tension in Yablo's philosophy of mind. On the one hand, he defends the metaphysical possibility of his own disembodiment. On the other hand, he adopts the supervenience of the mental on the physical as a (for him intuitive) premise in his account of mental causation (1992). But if it is possible for embodied beings—say, you and I—to be disembodied, shouldn't it also be possible for us to switch bodies with no accompanying purely physical change? My intuition of the former possibility seems to have no more nor less evidential weight than my intuition of the latter. If such body switching is possible, however, supervenience would fail: for example, after the switch your original body would have various mental properties which it lacks in the actual world: e.g., being the body of a certain thinking being, namely, *me*. A related source of tension is that there are people who have anti-supervenience intuitions that are just as strong as Yablo's disembodiment intuitions. It seems that Yablo's method would direct these people to give their anti-supervenience intuitions the same evidential weight that Yablo gives his disembodiment intuitions, thereby threatening stalemate between these people and Yablo over supervenience.

[47] Neither the weak zombie intuition (the possibility of physical functional duplicates having no mental properties) nor the strong zombie intuition (the possibility of perfect physical duplicates of us having no mental properties) is anywhere close to being universal and, of course, the reliability of these intuitions is hotly challenged. (Compare these intuitions with the multiple-realizability

Even though a full refutation of materialism requires the indicated strong zombie intuition, Chalmers's anti-materialist zombie argument is built on a zombie intuition that seems distinctly weaker (even though it is not as weak as the anti-functionalist zombie intuition). Chalmers's intuition may be reported thus: the primary intension (i.e., 'primary proposition') of the sentence 'Such-and-such physical properties are instantiated even though no qualitative conscious properties are' is possibly true. Here 'such-and-such physical properties' is meant to be a specification of the physical properties instantiated in the actual world.[48] This specification would of course contain various microphysical expressions—'electron', 'charge', and the like. For simplicity, let us assume that this specification is entirely microphysical; doing so will not affect the argument substantially.

Two-dimensional semantics is designed so that, typically, the primary intension and the secondary intension of natural kind predicates differ markedly from one another. If this is so for any of the expressions just indicated, then the mere fact that the primary intension of 'Such-and-such physical properties . . .' is possible does not tell us whether the *secondary* intension is possible. In turn, it does not tell us whether it is possible for such-and-such physical properties to be instantiated even though no conscious properties are. But, as noted, materialism will be refuted only if these very physical properties (electron, etc.)—not some contingently associated properties (primary intensions)—can be instantiated absent mental properties. Ironically, the whole point of two-dimensionalism was to separate the meaning of natural kind expressions—'water', 'heat', and, one would think, 'electron'—into two distinct kinds, primary and secondary. If two-dimensionalism does its job *uniformly*, Chalmers's intuition fails to refute materialism.

Chalmers is aware of this alleged 'wrong-intuition' problem and tries to show that 'it relies on an incorrect view of the semantics of physical terms' (1996: 135). According to Chalmers, 'Not only is reference to electrons fixed by the role that electrons play in a theory; the very concept of an electron is

intuition that there could be intelligence absent exactly our sort of complex electrochemical property.) For these reasons, most anti-materialists know that they will have few converts if their arguments depend on such intuitions—just as they know that they will have few converts if they rely on either the weak disembodiment intuition (that there could be a disembodied being) or the strong disembodiment intuition (that *we* could be disembodied).

[48] Recall that, for Chalmers, the primary intension ('primary proposition') of a sentence is possible iff it is true in some possible world iff it has the value True for some possible world. Note, also, that in Chalmers's framework, the above intuition is equivalent to the intuition that 'Such-and-such physical properties . . .' is epistemically possible. See sect. 1.5 and also the close of sect. 1.3.

defined by that role, which determines the application of the concept across all worlds' (1996: 136). If Chalmers were right about this, it would follow that in the case of 'electron' (and kindred terms), the secondary intension would be effectively the same as the primary intension. In this case, the secondary intension of the strong zombie sentence 'Such-and-such physical properties . . . ' would likewise be effectively the same as its primary intension. So, if the latter is possible, the former should be, too. In this event, Chalmers's intuition would refute materialism after all.[49]

But it is not his critic's view of the semantics of physical terms, but rather Chalmers's own view, that is unsatisfactory.[50] I mention three of its problems. First, the hypothesized deviation from normal two-dimensional semantics looks extremely *ad hoc*, and it is unexplained why there should be an about-face just at the point that it is needed to save the zombie argument. Second, since this semantics is committed to a Jackson-style implicit-theory descriptivism, it is refuted by the two dilemmas (see section 2.4.1) that plague the 'Platitudinous Conception' version of implicit descriptivism (for example, reminiscent of the people in communities c_1 and c_{11}, people who understand the term 'electron' can have very significant disagreements about the correct theory of electrons). The third problem may be brought out by means of an example.

Consider three worlds, w_1, w_2, w_3. In w_1 all particles have a certain property F that plays no identifiable causal role in the physical theory discoverable by

[49] I say 'effectively the same' because of the technicality that primary intensions are defined on centered worlds whereas secondary intensions are defined on ordinary (uncentered) worlds and, therefore, that primary and secondary intensions can never be identical. In what follows, I will for simplicity suppress the difference associated with this and related technicalities.

[50] True, Chalmers argues against the matter-chauvinists' view (that, in the case of 'electron', the primary and secondary intensions are different and, in turn, that the property of being an electron differs from the property of playing the electron role—i.e., the role electrons play in physical theory), and he does this by invoking the following claim: 'The notion of an electron that has all the extrinsic properties of actual protons does not appear to be coherent' (1996: 136). But his argument is a fallacy based on a confusion between necessary and sufficient conditions. Chalmers's matter-chauvinist opponents are free to agree with him that the indicated description of an electron is incoherent. Their point is that, even if all actual extrinsic electron properties were necessary for something's being an electron (thereby preventing electrons from having the envisaged extrinsic proton properties), these extrinsic electron properties would not be sufficient—something more is needed to be an electron. Chalmers's argument plainly does not touch this point. Given this, his argument does nothing to prevent our matter-chauvinists from going on to make the further claim that only the sum total of all these physical properties (including the 'something more') is sufficient for our conscious properties. Consequently, the matter-chauvinists' view of the body–mind relationship is also untouched by Chalmers's argument.

w_1's inhabitants. This holds, in particular, for the particles the inhabitants on w_1 call 'electrons'. Let us call these particles Felectrons. World w_2 is a kind Putnamian 'twin-earth world' of w_1: the particles in w_2 fall into the same number of kinds as those in w_1, and they interact with one another in the very same pattern as do the corresponding particles in w_1. There is a difference, however: namely, that instead of having property F, these particles have a certain property G. Like F in w_1, G plays no identifiable causal role in the physical theory discoverable by w_2's inhabitants. The particles these people call 'electrons' I will call Gelectrons. w_3 is a more or less symmetrical world. World w_3 may be thought of along the lines of Putnam's 'two-planet' twin-earth world. The particles on one half of w_3 (call it 'Rightland') are numerically the same as those in w_1: they are all F-particles (including Felectrons). The particles on the other half of w_3 (call it 'Leftland') are numerically the same as those in w_2: they are all G-particles (including Gelectrons). In interactions with one another, the F-particles behave exactly as they do in w_1; likewise, in interactions with one another, the G-particles behave exactly as they do in w_2. In interactions between F-particles and G-particles, however, something wholly novel occurs—say, mutual annihilation. In other words, the laws governing FF interactions in w_1 still govern them in w_3, and the laws governing GG interactions in w_2 still govern them in w_3. In addition to these laws, there are further laws governing FG interactions. (I believe it to be possible for all three sets of laws to hold in worlds w_1 and w_2 as well. On this scenario, the difference would be that GG-laws and FG-laws are uninstantiated in w_1, and FF-laws and FG-laws are uninstantiated in w_2. Although this is how I prefer to think of the example, this is not crucial.)

In the language of the 'Rightlanders' in w_3, would their term 'electron' apply to Gelectrons as well as Felectrons? Heavens no, Gelectrons actually destroy their paradigmatic electrons! Conversely, in the language of the 'Leftlanders', their term 'electron' would apply to Gelectrons but not Felectrons. Now, Rightlanders are numerically and epistemically the same people as they were in w_1, and the 'electron-ish things of their acquaintance' in w_1 and w_3 are numerically the same. Certainly, if the Rightlanders' term does not apply to the Gelectrons in w_3, the term 'electron' in the w_1 language would not apply to Gelectrons in w_3, either. And if it does not apply there, surely it would not apply to them in w_2. We thus have a counter-example to Chalmers's semantical picture. For in w_2 Gelectrons interact with one another and other G-particles in exactly the way FF-laws characterize the interaction of Felectrons with one another and other F-particles, but the w_1 term 'electron' does not apply to Gelectrons in w_3, contrary to what Chalmers's semantics predicts. This (perhaps

supplemented with other examples) shows that fundamental particles, forces, and so on cannot be defined by Ramsifying on any semantically stable base.[51]

We may draw two further conclusions. First, there can be physical properties that are 'hidden' in one world (as F is in w_1 and G is in w_2) but that can nevertheless be extremely significant physically—and not at all 'hidden'—in other worlds (as F and G are in w_3). Indeed, the difference between F and G plays a dominant role in the 'revealed' physical theory for w_3: namely, in the FG-laws. In other words, 'hiddenness' is not an in-principle property of F and G.[52] Second, and most importantly, if, as in the example, the physical properties F and G can be revealed to have different physical consequences (F-particles annihilate G-particles but not other F-particles), what is to prevent them—or properties akin to them—from being revealed to have different mental consequences as well? For example, what prevents F-particles from necessitating consciousness even though G-particles do not? Nothing prevents it, says the matter-chauvinist, thus defeating Chalmers's zombie argument.

Chalmers has one last response to the matter-chauvinist: namely, to abandon his original dualism and to replace it with another, which he enunciates thus: 'The dualism of "physical" and "nonphysical" properties is replaced on this [new view] by a dualism of "accessible" and "hidden" physical properties, but the essential point remains' (1996: 136).[53] But this is not so. As we just saw, the matter-chauvinist's challenge does not turn on essentially hidden properties, and, consequently, it impinges with equal force on this fall-back 'dualism' of alleged accessible and hidden physical properties.

[51] By the way, in w_3 there are two distinct sequences of properties (the sequence consisting of F-particle followed by G-particle and the sequence consisting of G-particle followed by F-particle) that simultaneously satisfy the conjunction of all three sets of laws. Could we not break this symmetry by including semantically unstable terms in the theory? Yes, but this would result in a difference between the primary intension and the secondary intension of 'electron', again contradicting Chalmers's position.

[52] How, would the terms 'Felectron' and 'Gelectron' be introduced in w_3? It appears that there is no alternative but to turn from Lewis's picture of theoretical terms to Kripke's: since natural kinds are not in general definable by Ramsification on a semantically stable base, they are successfully named only by taking advantage of our situatedness at some stage of naming or other. At the same time, although the significance—physical or mental—of what we are naming might be locally unclear, it need not be hidden in principle.

[53] Chalmers also makes the suggestion that the sorts of physical property upon which matter-chauvinism is based are 'protophenomenal'; but such renaming is one more violation of the terminological maxim alluded to at the outset. Successful reductions of Xs to Ys are trivialized by saying that Ys were really just proto-Xs all along.

By the way, Chalmers entertains, favorably, a kind of Identity Theory according to which there are 'protophenomenal' physical properties that, either alone or in combination, are identical to

4.3.2 *Two-Dimensional Modal Arguments and the Identity Thesis*

Kripke really offered two modal arguments against the Identity Thesis. One aimed to establish that it is metaphysically possible for something to have firing C-fibers without pain (thereby refuting the sufficiency condition). The other aimed to establish that it is metaphysically possible for something to be in pain without having firing C-fibers (thereby refuting the necessity condition). When the first argument is reconstructed in the style of Chalmers's two-dimensional zombie argument, the resulting argument has the following main premises (analogous to those in Chalmers's argument): (a_1) the modal premise that the primary intension of 'Something has firing C-fibers without being in pain' is possibly true; (b_1) the semantical premise that this sentence's secondary intension is identical to its primary intension. Similarly, when Kripke's second argument is reconstructed in this way, it has two analogous premises: (a_2) the primary intension of 'Something is in pain without having firing C-fibers' is possibly true; (b_2) this sentence's secondary intension is identical to its primary intension. But each of these arguments fails because its semantical premise is false (just as the corresponding semantical premise in Chalmers's original zombie argument, in section 4.3.1, was false). The underlying reason is that there will always be alternate worlds in which, say, you* (the qualitative epistemic counterparts of you) use 'C-fiber' for a natural kind very different from real C-fibers (e.g., in a science fiction case, you* and me* use 'C-fiber' for silicon fibers in the heads of the human-appearing androids who, along with you*, populate the planet). (As we saw in section 1.3, Kripke's arguments also failed because of this sort of semantic instability of 'C-fiber'.)

The above two-dimensional argument against the sufficiency condition is in principle unsalvageable (see section 4.4). It turns, out, however, that the above argument against the necessity condition can be salvaged as long as the secondary intension of 'firing C-fibers' entails (i.e., has as a necessary condition) the secondary intension of some predicate *C* for which the associated pair of premises hold: (a_3) the modal premise that the primary intension of 'Something is in pain without having *C*' is possibly true; (b_3) the semantical premise that this sentence's secondary intension is the same as this primary intension. Now, in fact, there are such predicates *C* (e.g., '74,985,263 or more functionally related nonconscious parts'—or whatever is the minimum number needed for

our phenomenal properties. This Identity Theory, however, is inconsistent with a possibility to which Chalmers's larger argument is committed: viz., the possibility of disembodied (and hence, nonphysical) beings whose phenomenal properties are nevertheless the same as ours.

having firing C-fibers). Furthermore, (a_3) and (b_3) trivially entail that the secondary intension of 'Something is in pain without having *C*' is possibly true. Therefore, since the secondary intension of 'firing C-fibers' entails the secondary intension of *C*, we obtain the conclusion: the secondary intension of 'Something is in pain without having firing C-fibers' is possibly true. But, for any English sentence S, if S's secondary intension is possibly true, it is metaphysically possible that S. It follows that it is metaphysically possible that something is in pain without having firing C-fibers. In other words, the Identity Thesis fails in this instance because having firing C-fibers fails to be a necessary condition for being in pain.

Ironically, this success on the part of two-dimensionalism reveals that two-dimensionalism is just a gratuitous complication, contributing nothing substantive to the refutation of the Identity Thesis. Why? Because the above two-dimensional argument can, without loss, be dropped in favor of an 'equivalent' but essentially simpler 'one-dimensional' argument.[54] The argument has two premises: (a_4) it is epistemically possible that something be in pain without having *C*; (b_4) the proposition that something is in pain without having *C* is semantically stable. Since a semantically stable proposition's epistemic possibility entails its metaphysical possibility, it follows from (a_4) and (b_4) that it is metaphysically possible that something be in pain without having *C*. Therefore, since the property of having firing C-fibers entails the property of having *C*, it follows that it is metaphysically possible that something is in pain without having firing C-fibers. The desired result.

The moral. Two-dimensional arguments against the Identity Thesis (viz., against its necessity condition) go through iff the primary and secondary intensions of 'Something is in pain without having *C*' are identical iff the sentence is semantically stable. (I suppress difference in individual- and community-centered meaning.) Hence, the two-dimensional argument goes through in just those cases where the primary/secondary distinction plays no role in so far as the sentence at issue is semantically stable and so able to underwrite an essentially simpler one-dimensional argument. Two-dimensionalism is just a third wheel: it succeeds only where it is not needed.

The indicated simpler pattern of argument generalizes. If we wish to show that a semantically unstable property U is not a necessary condition of a

[54] Equivalent in that these biconditionals hold: The primary intension of 'Something has pain without having C' is possibly true iff it is epistemically possible that something has pain without having *C*. Suppressing issues of private vs. public meaning (sects. 2.2.3; 2.4), the primary and secondary intensions of 'Something has pain without having *C*' are the same iff the sentence is semantically stable. The secondary intension of 'firing C-fibers' entails the secondary intension of *C* iff the property of having firing C-fibers entails the property of having *C*.

semantically stable property S, it suffices to find some necessary condition U′ of U such that: (a) it is epistemically possible for S to be instantiated without U′, and (b) U′ is semantically stable. This general strategy provides a systematic routine for disarming scientific essentialism's threat to a large family of modal arguments in philosophy. And it does so without commitment to any particular semantical theory (especially one whose adequacy is already in doubt on independent grounds) but merely a commitment to the semantic stability/instability distinction. The first goal of my early paper on mental properties (1994) was to isolate and defend this general strategy—and to apply it, in particular, against the Identity Thesis. The second goal was to demonstrate that this method's 'dual' (i.e., the corresponding method for showing unstable U not to be sufficient for stable S) is unsound, and, in particular, that it cannot refute the Identity Thesis.

4.4 *A Middle Way*

If Yablo's disembodiment argument were to go through, it would imply that mental properties do not have physical properties as substantive *necessary* conditions. But we saw that the argument is not likely to convince his materialist opponents, and their reservation is not without foundation, given the semantic instability of the key modal intuition. Analogously, if Chalmers's zombie argument were to go through, it would imply that mental properties do not have physical properties as *sufficient* conditions. Our notion of semantic instability allows us to see the failure of this argument in a similar way. From this point of view, the most that this style of argument could ever succeed in showing is that mental properties do not have any *semantically stable* physical properties as sufficient conditions. But this leaves untouched matter-chauvinism, which, in this idiom, amounts to the view that, even if semantically stable physical properties cannot be sufficient conditions for mental properties, certain semantically unstable physical properties can.[55]

For me the challenge has been to find a way to overcome this dialectical situation. More specifically, my goal (1994, 1997, 2000) has been to find a way to rely on only very weak, but very compelling, semantically stable intuitions (as in the previous subsection), thereby skirting entirely the issue of scientific essentialism.[56] When these epistemically safe intuitions are combined with our moderate rationalism, they yield all the same anti-materialist conclusions as the original, but inconclusive disembodiment and zombie arguments.

[55] Many matter–chauvinists believe that this is the secret of phenomenal qualities (sensing red, etc.). This phenomenon (and others) also creates problems for the Knowledge Argument.

[56] See the close of sect. 3.1 and n. 14.

REFERENCES

Bealer, George (1987), 'Philosophical Limits of Scientific Essentialism', *Philosophical Perspectives*, 1: 289–365.

—— (1992), 'The Incoherence of Empiricism', *Aristotelian Society*, supp. vol. 66: 99–138.

—— (1993*a*), 'A Solution to Frege's Puzzle', *Philosophical Perspectives*, 7: 17–61.

—— (1993*b*), 'Universals', *Journal of Philosophy*, 91: 185–208.

—— (1994), 'Mental Properties', *Journal of Philosophy*, 91: 185–208.

—— (1996), '*A Priori* Knowledge and the Scope of Philosophy', *Philosophical Studies*, 81: 121–42.

—— (1997), 'Self-Consciousness', *Philosophical Review*, 106: 69–117.

—— (1998), 'Propositions', *Mind*, 107: 1–32.

—— (1999), 'A Theory of the A Priori', *Philosophical Perspectives*, 13: 29–55.

—— (2001), 'Rationalism, Concept Identity, and the Solution to Frege's Puzzle', MS.

—— (forthcoming), *Philosophical Limits of Science* (New York: Oxford University Press).

Burge, Tyler (1979), 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4: 73–122.

Chalmers, David (1995), 'Facing Up to the Problem of Consciousness', *Journal of Consciousness Studies*, 2: 200–19.

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

Church, Alonzo (1950), 'On Carnap's Analysis of Statements of Assertion and Belief', *Analysis*, 10: 97–9.

Fine, Kit (1994), 'Essence and Modality', *Philosophical Perspectives*, 8 (Atascadero, Calif.: Ridgeview), 1–16.

Hill, Christopher (1997), 'Imaginability, Conceivability, Possibility and the Mind–Body Problem', *Philosophical Studies*, 87: 61–85.

Hirsch, Eli (1986), 'Metaphysical Necessity and Conceptual Truth', *Midwest Studies in Philosophy*, 11: 243–56.

Jackson, Frank (1993), 'Armchair Metaphysics', in Michaelis Michael and John O'Leary-Hawthorne (eds.), *Philosophy in Mind* (Dordrecht: Kluwer), 23–42.

—— (1998), *From Metaphysics to Ethics* (Oxford: Oxford University Press).

Jubien, Michael (2001), 'Propositions and the Objects of Thought', *Philosophical Studies*, 104: 47–62.

Kaplan, David (1989), 'Afterthoughts', in Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan* (New York: Oxford University Press), 565–614.

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

Putnam, Hilary (1975), 'The Meaning of "Meaning" ', in *Language, Mind, and Knowledge*, Minnesota Studies in the Philosophy of Science, 8 (Minneapolis: University of Minnesota Press), 131–91.

Schiffer, Stephen (1987), *Remnants of Meaning* (Cambridge, Mass.: MIT Press).
Yablo, Stephen (1990), 'The Real Distinction between Mind and Body', *Canadian Journal of Philosophy*, supp. vol. 16: 149–201.
—— (1992), 'Mental Causation', *Philosophical Review*, 101: 245–80.
—— (1993), 'Is Conceivability a Guide to Possibility?', *Philosophy and Phenomenological Research*, 53: 1–42.