

4

The Foundations of Two-Dimensional
Semantics*David J. Chalmers*

1. Meaning, Reason, and Modality

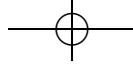
Why is two-dimensional semantics important? One can think of it as the most recent act in a drama involving three of the central concepts of philosophy: meaning, reason, and modality. First, Kant linked reason and modality, by suggesting that what is necessary is knowable a priori, and vice versa. Second, Frege linked reason and meaning, by proposing an aspect of meaning (sense) that is constitutively tied to cognitive significance. Third, Carnap linked meaning and modality, by proposing an aspect of meaning (intension) that is constitutively tied to possibility and necessity.

Carnap's proposal was intended as something of a vindication of Frege's. Frege's notion of sense is somewhat obscure, but Carnap's notion of intension is more clearly defined. And given the Kantian connection between reason and modality, it follows that intensions have many of the properties of Fregean senses. In effect, Carnap's link between meaning and modality, in conjunction with Kant's link between modality and reason, could be seen as building a Fregean link between meaning and reason. The result was a golden triangle of constitutive connections between meaning, reason, and modality.

Some years later, Kripke severed the Kantian link between apriority and necessity, thus severing the link between reason and modality. Carnap's link between meaning and modality was left intact, but it no longer grounded a Fregean link between meaning and reason. In this way the golden triangle was broken: meaning and modality were dissociated from reason.

Two-dimensional semantics promises to restore the golden triangle. While acknowledging the aspects of meaning and modality that derive from Kripke, it promises to explicate further aspects of meaning and modality that are more closely tied to the

An abridged version of this paper appeared under the title 'Epistemic Two-Dimensional Semantics' in a special issue of *Philosophical Studies* in 2004. Portions of this paper have been presented at the conferences on Two-Dimensionalism in Barcelona and ANU, at the Pacific Division meeting of the APA, and at UC Berkeley and the University of North Carolina. I am grateful to the audiences on those occasions for comments. I am especially grateful to Manuel Garcia-Carpintero and Daniel Stoljar for detailed comments on the paper.



rational domain. In particular it promises to look at the space of possibilities in a different way, and to erect a notion of meaning on that basis. In this way, we might once again have a grip on an aspect of meaning that is constitutively tied to reason.

To date, this restoration has been incomplete. Many different ways of understanding two-dimensional semantics have been proposed, and many of them restore the triangle at best partially. It is controversial whether two-dimensional semantics can be understood in such a way that the triangle is fully restored. To see this is possible, we need to investigate the foundations of two-dimensional semantics, and explore the many different ways in which the framework can be understood. I think that when the framework is understood in the right way, it can reconstitute the links between meaning, reason, and modality.

1.1 Frege, Carnap, and Kripke

We can begin with some more detailed background. If we squint at history from just the right angle, focusing on one strand of thought and setting aside others, we obtain a simplified rational reconstruction that brings out the key points.

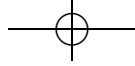
It is useful to start with Frege. Frege held that an expression in a language typically has a *referent*—or what I will here call an *extension*. The extension of a singular term is an individual: for example, the extension of the name ‘Hesperus’ is the planet Venus, and the extension of the description ‘the teacher of Aristotle’ is Plato. The extension of a general term is a class. And the extension of a sentence is its truth-value.

Frege noted that the extension of an expression does not in general determine its cognitive significance: the role it plays in reasoning and in knowledge. For example, ‘Hesperus’ (the name used for the evening star) and ‘Phosphorus’ (the name used for the morning star) have the same referent but have different cognitive significance, as witnessed by the fact that ‘Hesperus is Hesperus’ is cognitively trivial, while ‘Hesperus is Phosphorus’ is nontrivial. The same goes for many other pairs of expressions: perhaps ‘renate’ (creature with a kidney) and ‘cordate’ (creature with a heart), or ‘water’ and ‘H₂O’, or ‘I’ (as used by me) and ‘David Chalmers’. In each pair, the members are co-extensive (they have the same extension), but they are cognitively and rationally distinct.

Frege held that meaning is tied constitutively to cognitive significance, so that if two expressions have different cognitive significance, they have different meaning. It follows that there must be more to meaning than extension. Frege postulated a second aspect to meaning: *sense*. When two expressions are cognitively distinct, they have different senses. For example, the nontriviality of ‘Hesperus is Phosphorus’ entails that although ‘Hesperus’ and ‘Phosphorus’ have the same extension, they have a different sense. We can put the general idea as follows:

Fregean Thesis: Two expressions ‘A’ and ‘B’ have the same sense iff ‘ $A \equiv B$ ’ is cognitively insignificant.

Here, ‘ $A \equiv B$ ’ is a claim that is true if and only if ‘A’ and ‘B’ have the same extension. Where ‘A’ and ‘B’ are singular terms, this will be the identity ‘ $A = B$ ’; where ‘A’ and ‘B’ are sentences, this will be the material biconditional ‘A iff B’; and so on. As for cognitive significance, we can say at a first approximation that a claim is cognitively



insignificant when it can be known trivially by a rational being. As such, we can see this characterization of sense as providing a first bridge between meaning and reason.

The idea that expressions have senses is attractive, but senses are nevertheless elusive. What exactly is a sense? What exactly is cognitive significance? How does one analyze meanings beyond extensions? In the middle part of the 20th century, a number of philosophers, notably Carnap, had an insight. We can use the notions of *possibility* and *necessity* to help understand meaning, and in particular to help understand sense.

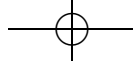
There are many possible ways the world might be; and we can use language to describe these possibilities. An expression can be applied to the actual state of the world, yielding an actual extension, or it can be applied to alternative possible states of the world, yielding alternative possible extensions. Take expressions such as ‘renate’ and ‘cordate’. In the world as it actually is, all renates are cordates, so these terms have the same extension. But it is not *necessary* that all renates are cordates: if the world had been different, some renates might have failed to be cordates. Applied to such an alternative possibility, the two terms have a different extension. We can say: ‘renate’ and ‘cordate’ are co-extensive, but they are not *necessarily* co-extensive. Carnap suggested that we say two expressions have the same *intension* if and only if they are necessarily co-extensive.¹ So ‘renate’ and ‘cordate’ have the same extension, but different intensions. We can put the general claim as follows:

Carnapian Thesis: ‘A’ and ‘B’ have the same intension iff ‘ $A \equiv B$ ’ is necessary.

What exactly is an intension? Carnap’s characterization suggests a natural definition: an intension is a function from possibilities to extensions.² The possibilities here correspond to different possible states of the world. Relative to any possibility, an expression has an extension: for example, a sentence (e.g. ‘All renates are cordates’) can be true or false relative to a possibility, and a singular term (e.g. ‘the teacher of Aristotle’) picks out an individual relative to a possibility. An expression’s intension is the function that maps a possibility to the expression’s extension relative to that possibility. When two expressions are necessarily co-extensive, they will pick out the same extension relative to all possibilities, so they will have the same intension. When two expressions are not necessarily co-extensive, they will not pick out the same extension relative to all possibilities, so they will have different intensions. So intensions behave just as Carnap suggests they should.

¹ See Carnap (1947). The idea is also present in Lewis (1944).

² This definition of an intension is often attributed to Carnap, but in Carnap (1947) it plays at most a minor role. He proposes something like this (in section 40, p. 181) as a way of understanding individual concepts, which are the intensions of names, but then moves to a slightly different understanding. Earlier in the book, he characterizes necessity (“L-truth”) in terms of state-descriptions, which are akin to possible worlds. But state-descriptions soon drop out of the discussion, so that intensions are treated in effect as something of a primitive semantic value. This sort of construction is also discussed in Carnap (1963), 892–94. (Thanks to Wolfgang Schwartz for pointers here.) A proposal close to the definition above is present in C. I. Lewis’s suggestion that an intension “comprises whatever must be true of any possible world in order that the proposition should apply to it or be true of it” (Lewis 1944).



Seen this way, the notion of an intension provides a bridge between meaning and modality. Just as a sense can be seen as a sort of meaning that is constitutively tied to reason, an intension can be seen as a sort of meaning that is constitutively tied to modality. Furthermore, intensions seem to behave very much as senses are supposed to behave. Just as two expressions can have the same extension but different senses, two expressions can have the same extension but different intensions. And just as sense was supposed to determine extension, intension seems to determine extension, at least relative to a world.

One can make a direct connection by adding an additional claim connecting modality and reason. It has often been held that a proposition is necessary if and only if it is a priori (knowable independently of experience) or trivial (yields no substantive knowledge of the world). The notions of apriority and triviality are essentially rational notions, defined in epistemic terms. Carnap himself held a version of the thesis involving triviality, but it is more useful for our purposes to focus on the version involving apriority. In this form, the relevant thesis goes back at least to Kant, so we can call it:

Kantian Thesis: A sentence S is necessary iff S is a priori.

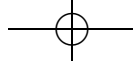
If we combine the Carnapian Thesis with the Kantian Thesis, we obtain the following:

Neo-Fregean Thesis: Two expressions 'A' and 'B' have the same intension iff ' $A \equiv B$ ' is a priori.

If this claim is accepted, then one has recaptured something that is at least close to the Fregean thesis. For apriority is at least closely related to cognitive insignificance. When a proposition is cognitively insignificant, it is plausibly a priori. The reverse is not the case, on Frege's understanding of cognitive significance: many logical and mathematical propositions are cognitively significant, even though they are a priori. But in any case, apriority and cognitive insignificance are at least closely related rational notions. Typical cognitively significant identities, such as 'Hesperus is Phosphorus', 'Water is H_2O ', and 'I am David Chalmers' are all a posteriori. If the Neo-Fregean Thesis is correct, it follows that 'Hesperus' and 'Phosphorus' have different intensions, as do 'water' and ' H_2O ', and 'I' and 'David Chalmers'. So intensions behave quite like Fregean senses.

In effect, modality serves as a bridge in explicating the tie between meaning and reason. One constructs a notion of meaning using modal notions, combines this with the claim that modality is constitutively tied to reason, and ends with a link between all three. The central connection between meaning, reason, and modality is captured within the Neo-Fregean thesis: intension is a notion of meaning, defined in terms of modality, that is constitutively connected to reason.

This golden triangle was shattered by Kripke, who cut the connection between reason and modality. Kripke argued that the Kantian Thesis is false: there are many sentences that are necessarily true but whose truth is not knowable a priori. For example, Kripke argued that given that Hesperus is actually Phosphorus, it could not have been that Hesperus was not Phosphorus: Hesperus is necessarily the planet



Venus, and so is Phosphorus. So although ‘Hesperus is Phosphorus’ is not knowable a priori, it is nevertheless necessary. More generally, Kripke argued that names and natural kind terms are rigid designators, picking out the same extension in all possible worlds. It follows that any true identity involving such terms is necessary. For example, ‘Water is H₂O’ is necessary, even though it is a posteriori. The same goes for claims involving indexicals: ‘I am David Chalmers’ (as used by me) is another a posteriori necessity.

If Kripke is right about the Kantian Thesis, then the Neo-Fregean Thesis is also false. Since ‘Hesperus is Phosphorus’ is necessary, ‘Hesperus’ and ‘Phosphorus’ have the same intension, picking out the planet Venus in all possibilities. But the equivalence between ‘Hesperus’ and ‘Phosphorus’ is nevertheless a posteriori and cognitively significant. So cognitively and rationally distinct pairs of expressions can have the same intension: witness ‘Hesperus’ and ‘Phosphorus’, ‘water’ and ‘H₂O’, ‘I’ and ‘David Chalmers’. So the Neo-Fregean Thesis fails, and intensions no longer behave like Fregean senses.

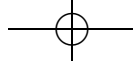
In effect, Kripke leaves intact the Carnapian link between meaning and modality, but in severing the Kantian link between reason and modality, he also severs the Fregean link between meaning and reason. This is roughly the received view in contemporary analytic philosophy: meaning and modality are connected, but both are disconnected from reason.

1.2 Two-dimensional semantics

Although most contemporary analytic philosophers accept Kripke’s arguments against the Kantian thesis, many would still like to hold that Frege was right about *something*. There remains an intuition that ‘Hesperus’ and ‘Phosphorus’ (or ‘water’ and ‘H₂O’, or ‘I’ and ‘David Chalmers’) differ in at least some dimension of their meaning, corresponding to the difference in their cognitive and rational roles. One might try to do this by breaking the Carnapian connection between meaning and modality. Two-dimensional semantics takes another strategy: in effect, it finds another way of looking at modality that yields a Fregean aspect of meaning.

The core idea of two-dimensional semantics is that there are two different ways in which the extension of an expression depends on possible states of the world. First, the actual extension of an expression depends on the character of the actual world in which an expression is uttered. Second, the counterfactual extension of an expression depends on the character of the counterfactual world in which the expression is evaluated. Corresponding to these two sorts of dependence, expressions correspondingly have two sorts of intensions, associating possible states of the world with extensions in different ways. On the two-dimensional framework, these two intensions can be seen as capturing two dimensions of meaning.

These two intensions correspond to two different ways of thinking of possibilities. In the first case, one thinks of a possibility as representing a way the actual world might turn out to be: or as it is sometimes put, one *considers a possibility as actual*. In the second case, one acknowledges that the actual world is fixed, and thinks of a possibility as a way the world might have been but is not: or as it is sometimes put, one *considers a possibility as counterfactual*. When one evaluates an expression relative



to a possible world, one may get different results, depending on whether one considers the possible world as actual or as counterfactual.

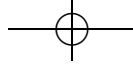
The second way of thinking about possibilities is the more familiar in contemporary philosophy. Kripke's arguments rely on viewing possibilities in this way. Take a possibility in which the bright object in the evening sky is a satellite around the earth, and in which Venus is visible and bright only in the morning. When we think of this possibility as a counterfactual way things might have been, we do not describe it as a possibility in which Hesperus is Mars, but as one in which Hesperus (and Phosphorus) is invisible in the evening. So relative to this possibility considered as counterfactual, 'Hesperus' picks out Venus. Correspondingly, the second-dimensional intensions of both 'Hesperus' and 'Phosphorus' both pick out Venus in this possibility, and in all possibilities in which Venus exists. It is this familiar sort of intension that yields the Kripkean gap between intension and cognitive significance.

The first way of thinking about possibilities is the less familiar in contemporary philosophy. If we take the possibility described above, and think of it as a way the world might actually be, we can say: if the world really is that way, then 'Hesperus' picks out a satellite. So relative to this possibility considered as actual, 'Hesperus' picks out not Venus but the satellite. Correspondingly, the first-dimensional intension of 'Hesperus' picks out the satellite in this possibility, while that of 'Phosphorus' picks out Venus. So 'Hesperus' and 'Phosphorus' have different first-dimensional intensions. This difference is tied to the fact that the actual-world reference of 'Hesperus' and 'Phosphorus' is fixed in quite different ways, although as things turn out, their referents coincide. Because of this, it seems that the first dimension may be better suited than the second for a link to reason and to cognitive significance.

The possibilities evaluated in the second dimension are usually thought of as possible worlds. The possibilities evaluated in the first dimension are a little different, as they reflect the nature of a world from the point of view of a speaker using an expression within a world. It is useful for many purposes to see these possibilities as *centered worlds*: worlds marked with a "center", which is an ordered pair of an individual and a time. We can think of the center of the world as representing the perspective of the speaker within the world.

I have been deliberately vague about just how the relevant intensions are to be defined, since as we will see, there are many different ways to define them. Because of this, giving detailed examples is tricky, because different frameworks treat cases differently. Nevertheless, it is useful to go through some examples, giving an intuitive analysis of the results that two-dimensional semantics might be expected to give *if* it is to yield something like a Fregean sense in the first dimension. We will later see how this can be cashed out in detail. For now, I will use "1-intension" as a generic name for a first-dimensional intension, and "2-intension" as a generic name for a second-dimensional intension.

First, 'Hesperus is Phosphorus'. In a centered world considered as actual, this is true roughly when the morning star visible from the center of that world is the same as the evening star. In a world considered as counterfactual, it is true when Venus is Venus. 'Hesperus' functions roughly to pick out the evening star in the actual world, so the 1-intension of 'Hesperus' picks out the evening star in a given centered



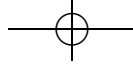
world. Likewise, the 1-intension of 'Phosphorus' picks out the morning star in a centered world. Both of these terms behave rigidly in counterfactual evaluation, so their 2-intensions pick out their actual referents in all worlds. So the 2-intensions of both 'Hesperus' and 'Phosphorus' pick out Venus in all worlds.

Second, 'Water is H_2O '. In a centered world considered as actual, this is true roughly when the clear, drinkable liquid around the center of that world has a certain pattern of chemical structure. In a world considered as counterfactual, it is true when H_2O is H_2O . The reference of 'water' is fixed roughly by picking out the substance with certain superficial properties and a certain connection to the speaker in the actual world, so its 1-intension picks out roughly the substance with those properties connected to the center of a given world. Similarly, the 1-intension of ' H_2O ' picks out the substance with the right sort of chemical structure in a centered world. As in the first case, both expressions behave rigidly in counterfactual evaluation, so their 2-intensions pick out H_2O in all worlds.

Third, 'I am a philosopher'. In a centered world considered as actual, this sentence is true when the being at the center of the world is a philosopher. In a world considered as counterfactual, this sentence (or at least my utterance of it) is true if David Chalmers is a philosopher in that world. The actual-world reference of 'I' is fixed by picking out the subject who utters the token; so the 1-intension of 'I' picks out the subject at the center of a given world. 'I' behaves as a rigid designator in counterfactual evaluation, so its 2-intension picks out the actual referent (in this case, David Chalmers) in all possible worlds. 'Philosopher', by contrast, is a broadly descriptive term: both its 1-intension and its 2-intension function to pick out beings with certain characteristic attributes. Certain patterns seem to emerge. The first two sentences are necessary (at least if Kripke is right), and both of them have a 2-intension that is true in all worlds. The third sentence is contingent, and its 2-intension is false in some worlds. So it seems that a sentence is necessary precisely when it has a necessary 2-intension. This corresponds directly to the Carnapian thesis: 2-intensions, in effect, are defined so that two expressions will have the same 2-intensions when they are necessarily equivalent.

On the other hand, all three of these sentences are a posteriori, and all of them appear to have a 1-intension that is false in some centered worlds. At the same time, a priori sentences such as (perhaps) 'All bachelors are unmarried males' or 'Hesperus (if it exists) has been visible in the evening' can plausibly be seen as having a 1-intension that is true in all centered worlds. So it is at least tempting to say that a sentence is a priori precisely when it has a necessary 1-intension. This corresponds to the neo-Fregean thesis: one might naturally suggest that two expressions have the same 1-intension precisely when they are a priori equivalent. To illustrate, one can note that the difference in the 1-intensions of 'Hesperus' and 'Phosphorus', or of 'water' and ' H_2O ', seems to be closely tied to their a priori inequivalence. All this needs to be analyzed in more depth, but one might at least characterize the general sort of behavior suggested in the examples above, where differences in 1-intensions go along at least roughly with differences in cognitive significance, as *quasi-Fregean*.

Along with the 1-intension and the 2-intension of a given expression, one can also define a *two-dimensional intension*. In many cases, just as an expression's extension



depends on how the actual world turns out, an expression's 2-intension depends on how the actual world turns out. The expression's two-dimensional intension captures this dependence: it can be seen as a function from centered worlds to 2-intensions, or equivalently as a function from pairs of centered worlds and worlds to truth-values. In the case of 'Hesperus', for example, the two-dimensional intension maps a centered world V to the 2-intension that picks out V 's evening star (if it exists) in any worlds W . The actual 2-intension of an expression corresponds to the two-dimensional intension evaluated at the actual centered world of the speaker: given that Venus is the actual world's evening star, the 2-intension of 'Hesperus' picks out Venus in all worlds. The 1-intension of an expression can be reconstructed by "diagonalizing" the two-dimensional intension: one evaluates the two-dimensional intension at a centered world W , yielding a 2-intension, and then one evaluates this 2-intension at the same world (stripped of its center). One might think of the two-dimensional intension as representing the way that an expression can be used to evaluate counterfactual worlds, depending on which world turns out to be actual.

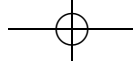
1.3 Varieties of two-dimensional semantics

I will return to these themes later, but for now it must be acknowledged that the situation is much more complicated than I have made things sound. A number of different two-dimensional systems have been introduced, and many of these give different results. A partial list of proponents of these systems, along with the names they give to their two-dimensional notions, includes:

Kaplan (1978; 1989): character and content
 Stalnaker (1978): diagonal proposition and proposition expressed
 Evans (1979): deep necessity and superficial necessity
 Davies and Humberstone (1981): "fixedly actually" truth and necessary truth
 Chalmers (1996): primary intension and secondary intension
 Jackson (1998a): A-intension and C-intension

There are many differences between these systems, some on the surface, and some quite deep. Surface differences include the fact that where Chalmers and Jackson speak of two sorts of intensions, Evans and Davies and Humberstone speak of two sorts of necessity, while Kaplan and Stalnaker speak of propositions. This sort of difference is mostly intertranslatable. Given a notion of necessity and a corresponding way of evaluating possibilities (as with Evans and Davies and Humberstone), one can define a corresponding sort of intension, and vice versa. Stalnaker's propositional content is just a set of possible worlds, which is equivalent to the intension of a sentence, and Kaplan's content is closely related.³ Kaplan's

³ Kaplan's content is strictly speaking a singular proposition rather than a set of worlds, but it immediately determines a set of worlds. For our purposes, the difference between singular propositions, other structured propositions, and sets of worlds in analyzing the second dimension of content will not be crucial, so for simplicity I will speak as if the relevant second-dimensional contents are intensions. The discussion can be straightforwardly adapted to other views.



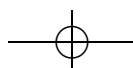
and Stalnaker's first-dimensional notions are defined over contexts (which are at least closely related to centered worlds), and initially involve a two-dimensional intension: a function from contexts to 2-intensions. Stalnaker diagonalizes this function, yielding a function from contexts to truth-values, or a 1-intension. Kaplan leaves his character as a two-dimensional function from contexts to 2-intensions, but a corresponding step could straightforwardly be taken. So in all these cases, there is a similar formal structure.

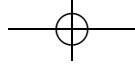
At a conceptual level, the systems have something further in common. In each case, the first-dimensional notion is put forward at least in part as a way of better capturing the cognitive or rational significance of an expression than the second dimension. And in each case, at least some sort of link between the first-dimensional notion and apriority has been claimed. In Kaplan's and Stalnaker's original publications, it is held that character and diagonal propositions closely reflect matters of apriority, at least in some cases. For Evans and Davies and Humberstone, when a statement of a certain sort is knowable a priori, it is deeply necessary, or true fixedly actually. And for Chalmers and Jackson, whenever a sentence is a priori, it has a necessary primary intension or 1-intension.

But these similarities mask deep underlying conceptual differences. These systems are defined in quite different ways, and apply to quite different items of language, yielding quite different results. Correspondingly, proponents of these systems differ greatly in the scope and strength of their claims. Kaplan's analysis is restricted to just a few linguistic expressions: indexicals and demonstratives. He explicitly resists an extension of his system to other expressions, such as names and natural kind terms. Evans and Davies and Humberstone develop their analysis for a different narrow class of expressions: descriptive names, and perhaps (in the case of Davies and Humberstone) some natural kind terms. Stalnaker's analysis applies in principle to any sentence, but in more recent work, he has explicitly disavowed any strong connection with apriority, and has been skeptical about applications of two-dimensional semantics in that direction. By contrast, Chalmers and Jackson suggest that their notions are defined for a very wide class of expressions, and make strong claims about the connection between these notions and apriority. (The current paper might be viewed in part as a defense of these strong claims.)

These differences arise from different *interpretations* of the formal two-dimensional framework. The framework of worlds and intensions, taken alone, is simply an abstract structure in need of content. Different interpretations flesh out this content in different ways. The interpretations are not necessarily incompatible, although it is possible that some are ill-defined, or rest on false presuppositions. The relations between these interpretations, however, are not well-understood.

The main project of this paper is to explore the different ways in which a two-dimensional framework can be understood. What are the fundamental concepts underlying different interpretations of the framework? How are these related? How do the differences between these interpretations explain the differences in the scope and strength of the claims that are made for them? Which interpretations of the framework yield the strongest connections between the first dimension and the rational domain?





1.4 The Core Thesis

The central question on which I will focus is the following. Is there an interpretation of the two-dimensional framework that yields constitutive connections between meaning, reason, and modality? That is, is there an interpretation on which the first dimension is tied universally to the rational domain? On this way of thinking, the ideal form of the two-dimensional framework will recapture something like the neo-Fregean thesis: two terms will have the same 1-intension if and only if they are equivalent a priori. To get at this question, we can focus on the following core thesis:

Core Thesis: For any sentence S , S is a priori iff S has a necessary 1-intension.

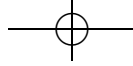
Here, S should be understood as a sentence token (such as an utterance) rather than a sentence type, to accommodate the possibility that different tokens of the same expression type may have different 1-intensions. Correspondingly, we should understand apriority as a property of sentence tokens. I will say more about the relevant notion of apriority and the type/token distinction in Section 3.8. But for now, to a first approximation, we can say that a sentence token S is a priori when S expresses actual or potential a priori knowledge (for the subject who utters S). And I will take it that the intuitive judgments about apriority above are correct: a typical utterance of ‘Hesperus is Phosphorus’ is not a priori, in this sense, while a typical utterance of ‘All bachelors are unmarried’ is a priori in this sense.

The Core Thesis links the rational notion of apriority, the modal notion of necessity, and the semantic notion of intension. If the Core Thesis is true, it restores a golden triangle of connections between meaning, reason, and possibility. It also immediately entails a version of the Neo-Fregean Thesis (given plausible principles about compositionality).

Neo-Fregean Thesis (2D Version): Two expressions ‘ A ’ and ‘ B ’ have the same 1-intension iff ‘ $A \equiv B$ ’ is a priori.

If the two-dimensional framework can be understood in such a way that the Core Thesis is true, it promises an account of a broadly Fregean aspect of meaning, tied constitutively to the epistemic domain. It also promises further rewards: perhaps an account of the contents of thought on which content is tied deeply to a thought’s rational role (potentially yielding an account of so-called “narrow content” and “modes of presentation” in thought), and perhaps a view of modality on which there are deep links between the rational and modal domains (potentially grounding a connection between notions of conceivability and possibility). So the key question in what follows will be: can we define 1-intensions so that the Core Thesis is true?

To anticipate, my answer will be as follows. There are two quite different ways of understanding the two-dimensional framework: the *contextual* understanding and the *epistemic* understanding. The contextual understanding uses the first dimension to capture *context-dependence*. The epistemic understanding uses the first dimension to capture *epistemic dependence*. The contextual understanding is more familiar, but it cannot satisfy the Core Thesis. The epistemic understanding is less familiar, but it



can satisfy the Core Thesis. The reason is that only on the epistemic understanding is the first dimension constitutively tied to the epistemic domain.

Within each of these general understandings of the framework, there are various possible specific interpretations. In what follows, I will first explore contextual interpretations (Section 2), and then epistemic interpretations (Section 3). Some of these interpretations are closely related to existing proposals, but rather than working directly with existing proposals, I will characterize these interpretations from first principles. This allows us to examine the properties of these interpretations in a clear light, free of problems of textual exegesis. Later in the paper, I will examine how existing proposals fit into this scheme.

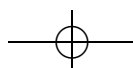
A methodological note: in this paper I will adopt the approach of *semantic pluralism*, according to which expressions can be associated with semantic values in many different ways. Expression types and expression tokens can be associated (via different semantic relations) with extensions, various different sorts of intensions, and with many other entities: structured propositions, conventionally implied contents, and so on. On this approach, there is no claim that any given semantic value exhausts the meaning of an expression, and I will not claim that the semantic values that I focus on are exhaustive. I think that such claims are almost always implausible.

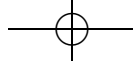
Likewise, this approach gives little weight to disputes over whether a given (purported) semantic value is “the” meaning of an expression, or even whether it is truly a “semantic” value at all. Such disputes will be largely terminological, depending on the criteria one takes to be crucial in one’s prior notion of “meaning” or “semantics”. On the pluralist approach, the substantive questions are: can expressions (whether types or tokens) be associated with values that have such-and-such properties? If so, what is the nature of the association and of the values? What aspects of language and thought can this association help us to analyze and explain?

My focus in this paper will be almost wholly on whether there is an association between expression tokens and 1-intensions that satisfies the Core Thesis, and on how this association can be understood. I will not say much more about the motivations for this sort of approach, about the broader shape of the resulting semantic theory, or about applications. Motivation and broader questions are discussed in “On Sense and Intension” (Chalmers 2002b), which gives a gentler introduction to these issues. Applications are discussed in “The Components of Content” (Chalmers 2002c) and in “Does Conceivability Entail Possibility?” (Chalmers 2002a).

2. The Contextual Understanding

On the contextual understanding of two-dimensional semantics, the possibilities involved in the first dimension represent possible *contexts of utterance*, and the intension involved in the first dependence represents the *context-dependence* of an expression’s extension. There are many ways in which the extension of an expression can depend on the context in which it is uttered. On the contextual understanding, a 1-intension captures the way in which an expression’s extension depends on its context. As we will see, this sort of context-dependence can itself be understood in a number of different ways.





To formalize this, we can start by focusing on expression *tokens*: spoken or written tokens of words, sentences, and other expressions. We can take it that any expression token has an extension. In cases where a token “aims” to have an extension but fails, as with an empty name, we can say that it has a null extension. If there are some expression tokens that do not even aim to have an extension (as perhaps with some exclamations), they are outside the scope of our discussion. A token of a sentence corresponds to an utterance; its extension is a truth-value.

Any expression token falls under a number of different expression *types*. A token may fall under an orthographic type (corresponding to its form), a semantic type (corresponding to its meaning), a linguistic type (corresponding to its identity within a language), and various other types. Different tokens of the same expression type will often have different extensions. When two tokens of the same expression type have different extensions, this reflects a difference in the *context* in which the tokens are embedded.

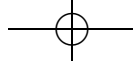
For our purposes, contexts can be modeled as centered worlds. The context in which an expression token is uttered will be a centered world containing the token. This can be modeled as a world centered on the speaker making the utterance, at the time of utterance. It is also possible to model a context by a different sort of centered world with just an expression token marked at the center. The previous version will work for most purposes, however, as long as we assume that a subject makes at most one utterance at a given time.

One can now define the *contextual intension* of an expression type. This is a function from centered worlds to extensions. It is defined at worlds centered on a subject uttering a token of the expression type. At such a world, the contextual intension returns the extension of the expression token at the center.

One can also define the contextual intension of an expression token, relative to a type of which it is a token. This is also a function from centered worlds to extensions. It is defined at worlds centered on a token of the same type, and returns the extension of the token at the center. This contextual intension is the same as the contextual intension of the relevant expression type.

The first-dimensional intensions in the two-dimensional framework are often understood as contextual intensions of some sort. On this way of seeing things, a 1-intension mirrors the evaluation of certain metalinguistic subjunctive conditionals: if a token of the relevant type were uttered in the relevant context, what would its extension be? Of course, for every different way of classing expression tokens under types, there will be a different sort of contextual intension. In what follows I examine some of the relevant varieties of contextual intension.⁴

⁴ Constructs akin to contextual intensions have been stressed by Robert Stalnaker in a number of writings (e.g. Stalnaker 1978, 1999). At the same time, Stalnaker and Ned Block have both been active critics of the overextension of this framework (e.g. Stalnaker 1990, 2001; Block 1991; Block and Stalnaker 1999). The discussion in this section owes a significant debt to Stalnaker and Block. Although I carve up the territory in a different way, a number of the varieties of contextual intension that I mention are touched on explicitly or implicitly by Stalnaker and Block at various points, and some of my critical points echo points made by them in criticizing certain applications of the two-dimensional framework.



2.1 Orthographic contextual intensions

We can say that two tokens are tokens of the same *orthographic type* when they have the same orthography. This holds roughly when they are made up of the same letters or sounds, regardless of their meaning, and regardless of the language in which they are uttered. The exact details of what counts as the same orthography can be understood in different ways, but these differences will not matter for our purposes.

The *orthographic contextual intension* of an expression token T is defined at centered worlds with a token of T 's orthographic type at the center. It maps such a world to the extension of the relevant token in that world.

(The orthographic contextual intension of a sentence token is closely related to its *diagonal proposition*, as defined by Stalnaker (1978). I will return to this matter later.)

As an example, let S be Oscar's utterance of 'Water is H_2O '. Let W_1 be Oscar's world (Earth), centered on Oscar making this utterance. Oscar's utterance is true, so S 's orthographic contextual intension is true at W_1 . Let W_2 be a universe containing Twin Earth (where everything is just as on earth except that the watery liquid is XYZ), centered on Twin Oscar uttering 'Water is H_2O '. Twin Oscar's utterance is false (his word 'water' refers to XYZ), so S 's orthographic contextual intension is false at W_2 . Let W_3 be a universe containing Steel Earth, where the word 'water' refers to steel but chemical terms are the same, centered on Steel Oscar uttering 'Water is H_2O '. Steel Oscar's utterance is false, so S 's orthographic contextual intension is false at W_3 .

It is clear that orthographic contextual intensions do not satisfy the Core Thesis. For *every* orthographic type, some possible token of that type expresses a falsehood. For example, there are worlds in which the string 'bachelors are unmarried' means that horses are cows. In such a centered world, the orthographic contextual intension of 'bachelors are unmarried' is false. The same goes for any sentence. So no truth has a necessary contextual intension, and in particular no a priori truth has a necessary contextual intension. So if 1-intensions are understood as orthographic contextual intensions, the Core Thesis is obviously false.

2.2 Linguistic contextual intensions

We can say that two expression tokens are tokens of the same *linguistic type* when they are tokens of the same expression in a language. This assumes that expression tokens belong to languages, and that languages involve expressions such as words, phrases, and sentences. So any two tokens of the English word 'water' share a linguistic type, as do any two utterances of the French sentence 'C'est la vie'.

The *linguistic contextual intension* of an expression token T is defined at centered worlds with a token of T 's linguistic type at the center. It maps such a world to the extension of the relevant token in that world.

(The linguistic contextual intension of an expression is in some respects like its *character*, as defined by Kaplan. I will return to this matter later.) As before, let S be Oscar's utterance of 'Water is H_2O '. If W_1 is Oscar's own centered world (Earth): S 's linguistic contextual intension is true at W_1 . If W_2 is Twin Oscar's centered world (Twin Earth): it is arguable that Twin Oscar's word 'water' is a *different* word from

Oscar's word 'water'. Certainly *if* the referent of 'water' is essential to the word, as many theorists hold, then Twin Oscar's 'water' is a different word. If so, S's linguistic contextual intension is not defined at W_2 . If W_3 is Steel Oscar's centered world (where 'water' means steel): here it is reasonably clear that Steel Oscar's 'water' is a different word that has the same orthography. If so, S's linguistic contextual intension is not defined at W_3 . Applying this sort of reasoning, one reaches the conclusion that S's contextual intension is true at every world in which it is defined, since the English word 'water' refers to H_2O in every world in which it exists, and so does the English expression ' H_2O '.

If this is right, then linguistic contextual intensions do not satisfy the core thesis. 'Water is H_2O ' is a posteriori, but it seems to have a necessary contextual intension, true at every world at which it is defined. The same goes even more clearly for sentences involving names, such as 'Cicero is Tully'. It is widely held that names have their referents essentially; if so, the linguistic contextual intensions of true identities of this sort will be true at all worlds at which they are defined. As such, linguistic contextual intensions do not behave at all like Fregean senses. If 1-intensions are understood as linguistic contextual intensions, the Core Thesis is false.

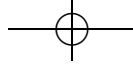
There are some expressions for which linguistic contextual intensions behave more like Fregean senses. One such is 'I': setting certain odd cases aside, any token of the English word 'I' picks out the utterer of that token. So the linguistic contextual intension of 'I' picks out the speaker at the center of any centered world at which it is defined. In this way, it behaves much as we earlier suggested the 1-intension of 'I' should behave. Something similar applies to other indexicals, such as 'today', and to some broadly descriptive terms, such as 'philosopher'. It is in the case of names and natural kind terms that the fit seems to be worst.

2.3 Semantic contextual intensions

We can say that two expression tokens are tokens of the same *semantic type* when they have the same semantic value. An expression token's semantic value is its meaning or content, or some aspect of its meaning or content. There are many different ways of assigning semantic values to expression tokens, so there are correspondingly many different ways of classing expression tokens under semantic types.

The *semantic contextual intension* of an expression token T is defined at centered worlds with a token of T's linguistic type at the center. It maps such a world to the extension of the relevant token in that world.

As before, let S be Oscar's utterance of 'Water is H_2O '. If W_1 is Oscar's own centered world (Earth): S's semantic contextual intension is true at W_1 . If W_2 is Twin Oscar's centered world (Twin Earth): at least on many ways of assigning semantic value, Twin Oscar's term 'water' has a different semantic value from Oscar's, so S's semantic contextual intension (for this sort of semantic type) is undefined at W_2 . If W_3 is Steel Oscar's centered world, then Steel Oscar's term 'water' clearly has a different semantic value from Oscar's, so S's semantic contextual intension is undefined at W_3 . If W_4 is a world centered on French Oscar, a counterpart of Oscar who speaks French and is uttering 'eau est H_2O ': then it is plausible that this



utterance has the same semantic value as Oscar's, so S's semantic contextual intension is defined at \mathbb{W}_1 and is true there.

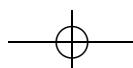
Of course the behavior of a semantic contextual intension will depend on our choice of semantic value. For example, if we stipulate that the relevant semantic value of an expression is its *extension*, then any two co-extensive expressions will have the same semantic contextual intension, and there is no chance that the Core Thesis will be true. There are two choices of semantic value that are somewhat more interesting, however.

We might stipulate that the relevant semantic value of an expression is its *standing meaning*: roughly, the aspect of meaning that is common to all tokens of the expression's linguistic type. If we do this, then an expression's semantic contextual intension will be an extension of its linguistic contextual intension to a broader space of worlds. At worlds centered on a token of the same linguistic type, the intensions will give the same results. But the semantic contextual intensions will also be defined at other worlds, centered on synonyms and translations of the original expression. Nevertheless, if 1-intensions are understood as these semantic contextual intensions, the Core Thesis will be false for the same reasons as in the case of linguistic contextual intension. For example, if the extension of 'water' is essential to the word, then it is part of the word's standing meaning. So the semantic contextual of 'Water is H_2O ' will be true at every world where it is defined, and the Core Thesis is false.

Alternatively, we might stipulate that the relevant semantic value of an expression token is its *Fregean or descriptive content*, corresponding roughly to the expression's cognitive significance for the subject. On this reading, the Core Thesis may be more plausible. For example, one might argue that Oscar's and Twin Oscar's terms 'water' have the same descriptive content. If so, then the semantic contextual intension of Oscar's utterance 'Water is H_2O ' is defined at \mathbb{W}_2 and is false there. On the other hand, Steel Oscar's term 'water' plausibly has a different descriptive content, so the semantic contextual intension of Oscar's utterance is not defined at \mathbb{W}_3 .

Understood this way, semantic contextual intensions behave as we might expect a Fregean 1-intension to behave, at least to some extent. One can argue that when a statement is a priori, any possible statement with the same descriptive content will be a priori and so will be true, so that the expression's semantic contextual intension will be necessary, as the Core Thesis requires. Correspondingly, one might suggest that when a statement is not a priori, then there will be possible statements with the same descriptive content that are false, so that the statement's semantic contextual intension will not be necessary, as the Core Thesis requires.

I will argue shortly that this is not quite right. But even if it were right, it is clear that this sort of 1-intension cannot underwrite the full ambitions of the Fregean two-dimensionalists. The Fregean two-dimensionalist, as sketched previously, intends to use the two-dimensional framework to *ground* an aspect of meaning that is constitutively tied to meaning. But semantic contextual intensions as defined here *presuppose* such a Fregean semantic value, and so cannot independently ground such an account. If this is the best a two-dimensionalist can do, then if someone is independently doubtful about a Fregean aspect of meaning, two-dimensionalism



cannot help. At best, two-dimensionalism will be a helpful tool in analyzing such a notion of meaning, given an independent grounding for the notion.⁵

2.4 A further problem

We have seen that orthographic contextual intensions are far from satisfying the Core Thesis, while linguistic contextual intensions are closer at least in some cases, and some sort of semantic contextual intensions may be closer still. But there is a further problem that arises for any sort of linguistic or semantic contextual intension, suggesting that no such contextual intension can satisfy the Core Thesis.

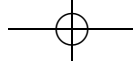
Let S be a token of 'A sentence token exists' (where a sentence token is understood to be a concrete entity produced by speech, writing, or a similar process). Then S is true. Furthermore, any token of the linguistic item 'A sentence token exists' is true. Any token that *means* the same thing as 'A sentence token exists' is true. So it seems that S will have a necessary linguistic contextual intension, and a necessary semantic contextual intension, under any reasonable way of classifying linguistic and semantic types. But S is clearly a posteriori: it expresses empirical knowledge of the world, which could not be justified independently of experience. So S is a counterexample to the Core Thesis. So the Core Thesis is false for any sort of semantic or linguistic contextual intension.

The same goes for a number of other sentences. If S_1 is 'Language exists' (where a language is understood to be a spoken or written language, not just an abstract language), then any utterance of the same expression or with the same meaning will be true. So S_1 has a necessary linguistic and contextual intension, despite being a posteriori. If S_2 is 'I exist', then any utterance of the same expression with the same meaning will be true, so S_2 has a necessary linguistic and semantic contextual intension. But (somewhat controversially) S_2 is a posteriori, justifiable only on the basis of experience. If S_3 is 'I am uttering now', then any utterance of the same expression or with the same meaning will be true. S_3 is clearly a posteriori, but has a necessary linguistic and semantic contextual intension.

All these cases are counterexamples to the Core Thesis. All of them are a posteriori and cognitively significant, and many of them seem to be as cognitively significant as paradigmatic expressions of empirical knowledge. But all have necessary semantic and linguistic contextual intensions. So the Core Thesis is false for all such intensions.

The trouble is that apriority and being true whenever uttered are fundamentally different notions. The first builds in an epistemic or rational element, but the second builds in no such element. The second notion builds in a metalinguistic element, but the first builds in no such element. It is possible to understand the second in a way that makes it coincide with the first in many cases, in effect by building in an epistemic element into the individuation of the relevant linguistic types. But it is impossible to do so in all such cases, since the second has an ineliminable metalinguistic element that goes beyond the epistemic domain.

⁵ This sort of point is made quite clearly, in the context of discussing narrow content, by Stalnaker (1991), Block (1991), and Block and Stalnaker (1999).



I think the moral is that to satisfy the Core Thesis, we must understand the two-dimensional framework in a quite different, non-contextual way. But before doing so, I will more briefly examine some further ways in which one might define a contextual intension.

2.5 Hybrid contextual intensions

Given orthographic, linguistic, and semantic types for expression tokens, it is possible to define *hybrid types* corresponding to conjunctions of two or more of these types. One can then define corresponding *hybrid contextual intensions*.

For example, one might say that two expressions share the same *orthographic/semantic type* when they share the same orthographic type and the same semantic type. One can then define the *orthographic/semantic contextual intension* of an expression as the function that maps a world centered on a token of the appropriate orthographic/semantic type to the extension of that token.

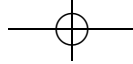
Hybrid contextual intensions may be useful for some purposes, but it is clear that they will not satisfy the Core Thesis any better than non-hybrid contextual intensions. So I will set them aside here.

2.6 Token-reflexive contextual intensions

It is possible to define a slightly different sort of contextual intension for an expression token by focusing not on the types that the token falls under, but on the token itself. Let us assume that expression tokens are not tied to their context essentially: a given token might have been uttered in another context. Then we can say that the *token-reflexive* contextual intension of an expression token T is a function that maps a centered world containing T to the extension of T in that world.

The precise behavior of a token-reflexive contextual intension will depend on what properties an expression token has necessarily. It is plausible that such a token has any properties necessarily, it has its orthographic properties necessarily. If so, its token-reflexive contextual intension will be a restriction of its orthographic contextual intension, obtained by eliminating worlds centered on a *different* token of the same orthographic type. One might also hold that a token has some semantic value necessarily, or that it has its linguistic type necessarily. If so, its token-reflexive contextual intension will be a restriction of its semantic or linguistic contextual intension. If an expression has more than one of these things necessarily, its token-reflexive contextual intension will be a restriction of a hybrid contextual intension. If it has further properties necessarily (e.g. its speaker), it will be a further restriction of the relevant contextual intension.

It is not obvious how to decide exactly which properties an expression token has necessarily. But however we do this, it is clear that token-reflexive contextual intensions cannot satisfy the Core Thesis. The counterexamples discussed above, such as ‘I am uttering now’, will apply equally to token-reflexive contextual intensions. Furthermore: insofar as tokens have any properties necessarily, one can likely construct sentence tokens attributing these properties that are true whenever uttered, but not a priori (e.g. ‘This token has four words’; ‘David Chalmers is speaking now’). And



insofar as tokens have few properties necessarily, one can likely construct sentences that are a priori but that are not true whenever uttered (e.g. 'All bachelors are unmarried'). So if I-intensions are understood as token-reflexive contextual intensions, the Core Thesis is false.

2.7 Extended contextual intensions

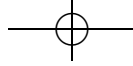
In an attempt to get around the problems posed by sentences such as 'I am uttering now', one might attempt to construct contextual intensions that are defined at centered worlds that do not contain a token of the relevant expression type. The most obvious way to do this is to appeal to certain counterfactual conditionals. Let us say that the *extended contextual intension* is defined at any centered world, independently of whether a token of the type is present there. At a given centered world, the extended contextual intension returns what the extension of a token of that type *would be*, if it *were* uttered at the center of that world.

One can then say that the extended contextual intension of an expression token (relative to a type) maps a centered world to what the extension of a token of that type would be, if it were uttered at the center of the world. So in principle, one might have extended linguistic contextual intensions, extended semantic contextual intensions, and so on. One could define an extended token-reflexive contextual intension in an analogous way.

An obvious problem here is that in many cases, it is unclear how to evaluate the counterfactual. It may be reasonably straightforward in some cases, such as 'I am a philosopher': true just when an utterance of 'I am a philosopher' by the subject at the center would be true, so true just when the person at the center is a philosopher. But how is one to evaluate what a token of 'water' would refer to if it were used in a world where there is no liquid, and in which nobody speaks a language? How does one evaluate whether an utterance of 'I am speaking loudly' would be true if it were uttered, in a world where the subject at the center is not in fact speaking? In some cases, it seems impossible for a token of the relevant type to be uttered in the relevant context. In other cases, it may be possible, but it is possible in many different ways, yielding many different results. So the truth of the relevant counterfactuals seems to be underdetermined, and an expression's extended contextual intensions seems to be ill-defined.⁶

Another problem: even if extended contextual intensions behave coherently, they give results that are different from what we need. For example, let $S =$ 'I am uttering now'. S is a posteriori, so the Core Thesis requires that its I-intension be false at some worlds. For example, it is desirable that S 's I-intension be false at an utterance-free world. Let \mathbb{W} be such an utterance-free centered world. To evaluate S 's contextual intension at \mathbb{W} , we ask: if S were uttered at the center of \mathbb{W} , what would its extension be? It is clear that if S *were* uttered in \mathbb{W} , it would be true. So S 's extended contextual intension is true at \mathbb{W} , and indeed is true at all worlds. So the Core Thesis is still false for extended contextual intensions.

⁶ A point of this sort is made by Stalnaker (1990).



To get anything like the result that is needed, we would need to evaluate S 's extension in \mathbb{W} *without* S being present in \mathbb{W} . But it is very hard to do that on the contextual model. On the contextual understanding, 1-intensions are derivative on facts about the extensions of various possible tokens, as uttered in various possible contexts. It seems clear that on such an understanding, the 1-intension of a sentence such as 'There are sentence tokens' will never be false.

I think that the idea of an extended contextual intension is getting at something important: that we need to be able to evaluate an expression's 1-intension in centered worlds that do not contain a token of the expression. But this is the wrong way to achieve the goal. To do this properly, I think we need to go beyond the contextual understanding of 1-intensions.

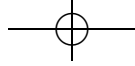
2.8 Cognitive contextual intensions

One might suggest that to capture a token's cognitive significance, we should not focus on a token's broadly *linguistic* properties, such as its orthography, its semantic value, and its language. Instead, we need to focus on its *cognitive* properties, which correspond to mental features of the subject that produces the token. Some such features include: the concept or belief that the token expresses; the cognitive role associated with the token; and the intentions associated with the token. Assuming that we have a way of individuating the mental types in question, we can then classify expression tokens under corresponding cognitive types.

For a given scheme of cognitive typing, one can then define the *cognitive contextual intension* of an expression token as the intension that maps a world centered on a token of the same cognitive type to the extension of that token. In the three cases above: a *conceptual* contextual intension will be defined at worlds centered on a token expressing the same concept or belief; a *cognitive-role* contextual intension will be defined at worlds centered on a token associated with the same cognitive role; and an *intention-based* contextual intension will be defined at worlds centered on a token associated with the same intentions.

Assuming that one can make sense of the relevant typing, there is a natural extension of this idea. One could define a sort of *extended* cognitive contextual intension, defined at worlds that do not contain the token at all, but merely contain the relevant mental feature. For example, the extended conceptual contextual intension will be defined at any world that contains the relevant *concept* at its center, irrespective of whether it contains any token, and will return the extension of the concept. (This assumes that concepts have extensions, which seems reasonable enough.) The extended cognitive-role contextual intension might be defined at any world centered on a concept that plays the relevant cognitive role, returning the concept's extension; and the extended intention-based contextual intension will be defined at any world centered on a concept that is associated with the same intentions.

This sort of intension has some promise of dealing with the central problems raised so far. In the case of 'A sentence token exists': one can make a case that the extended conceptual contextual intension of this expression is *false* at some centered worlds: those in which a subject has the relevant concepts and the relevant thought, but in



which there are no sentence tokens. So the intension is not necessary, reflecting the aposteriority of the sentence. The same goes for 'Language exists', and for 'I am uttering now'. By allowing intensions to be evaluated without relying on language, the metalinguistic element of contextual intensions has been reduced or eliminated.

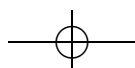
Still, analogous problems arise. 'I am thinking now' will plausibly have a necessary conceptual contextual intension, but it is plausibly a posteriori: the thought itself is justified only by experience, albeit by introspective experience. The same goes for 'I exist'. And the same will apply to specific attributions of mental features: a thought such as 'I have the concept *concept*' will be true whenever it is thought, but it is not justifiable a priori. Something similar applies to thoughts attributing certain cognitive roles or certain intentions. So even here, some a posteriori sentences and thoughts will have a necessary I-intension.

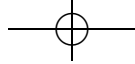
As for the other main sort of problem discussed so far, that associated with 'Water is H₂O': a proponent might hold that although Oscar and Twin Oscar do not have the same *word* 'water', their words express the same *concept*, at least under one reasonable way of individuating concept types. If so, then the conceptual contextual intension of Oscar's token 'Water is H₂O' will be false at the world centered on Twin Oscar, as the Fregean conception requires. At the same time, it might be undefined at the world centered on Steel Oscar (since he seems to have a different concept), as required.

It is controversial, however, whether concepts (or roles or intensions) can be individuated in a way that yields these results. Many theorists hold that even a token concept expressed by 'water' has its extension essentially, and that all concepts of the same type have the same extension. If so, then a statement such as 'Water is H₂O' will have a necessary intension. They might concede that concepts or thoughts can also be individuated syntactically or formally; but on this way of doing things, 'All bachelors are unmarried' will have a contingent intension. So either way, the Core Thesis is false.

One might argue that there is an intermediate way of individuating concept types that yields the right results. But many will deny this. It might be objected that this requires individuating concepts by their *narrow content* (that aspect of their content that is determined by a subject's intrinsic properties), and it is highly controversial whether narrow content exists. Some think that the two-dimensional framework can be used to give an account of narrow content; but in this context, it seems illegitimate for the framework to presuppose narrow content. This is a precise analog of the problem that arose for the Fregean version of semantic contextual intensions above.

I think that the situation here is not entirely clear. One could argue with some plausibility that there is an *intuitive* sense in which Oscar and Twin Oscar have the same concept, where there is no corresponding intuitive sense that they have the same word. If so, one could appeal to this intuitive sort of concept individuation to ground some sort of conceptual contextual intension here. One might arguably be able to do the same sort of thing with cognitive roles, or intentions. But the intuitions in question are likely to be disputed by many, so this approach will be at best weakly grounded, unless one can give some sort of independent account of the relevant concept types.





On my view, (extended) cognitive contextual intensions are the sort of contextual intensions that are closest to satisfying the Core Thesis. But ultimately, the central problems arise for them too. One might try appealing to related notions that carry features of the subject across worlds: for example, an *evidential* contextual intension, requiring sameness of evidence; a *fixing* contextual intension, requiring sameness of reference-fixing procedures or intentions; a *physical* contextual intension, requiring that subjects be physical duplicates; *functional*, *phenomenal*, *physical-phenomenal* contextual intensions, which require that subjects be functional, phenomenal, and physical-phenomenal duplicates; and so on. But it is not hard to see that all of these suggestions are subject to versions of the problems mentioned above. So we still need an account of the relevant intensions.

2.9 Summary

Overall, it seems that there is no way to define contextual intensions so that they satisfy the Core Thesis. Two central problems have arisen repeatedly. First, by building in a token of the relevant mental or linguistic type into the world of evaluation, the constitutive connection with the a priori is lost. Second, for a contextual intension to behave in a quasi-Fregean manner, we need to antecedently classify tokens under some sort of quasi-Fregean type, so that the framework cannot independently ground quasi-Fregean notions, as was originally hoped.

Contextual intensions may still be useful for many purposes. But they do not yield any restoration of the golden triangle, and in particular they do not deliver a notion of meaning that is deeply tied to reason. The fundamental problem is that although some contextual intensions yield a reasonably strong correlation with the epistemic domain, none is *constitutively* connected to the epistemic domain. To restore the connection between meaning and reason, we need to approach the two-dimensional framework in epistemic terms.

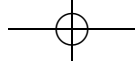
3. The Epistemic Understanding

3.1 Epistemic dependence

On the epistemic understanding of two-dimensional semantics, the possibilities involved in the first dimension are understood as *epistemic possibilities*, and the intensions involved in the first dimension represent the *epistemic dependence* of the extension of our expressions on the state of the world.

There are two key ideas here. The first is the idea of *epistemic space*: there are many ways the world might turn out to be, and there is a corresponding space of epistemic possibilities. The second is the idea of *scrutability*: once we know how the world has turned out, or once we know which epistemic possibility is actual, we are in a position to determine the extensions of our expressions. Together, these two ideas suggest that an expression can be associated with a function from epistemic possibilities to extensions: an *epistemic intension*.

Take the first idea first. There are many ways the world might be, for all we know. And there are even more ways the world might be, for all we know a priori. The



oceans might contain H_2O or they might contain XYZ; the evening star might be identical to the morning star or it might not. These ways the world might be correspond to epistemically possible hypotheses, in a broad sense. Let us say that a claim is *epistemically possible* (in the broad sense) when it is not ruled out a priori. Then it is epistemically possible that water is H_2O , and it is epistemically possible that water is XYZ. It is epistemically possible that Hesperus is Phosphorus, and epistemically possible that Hesperus is not Phosphorus.

Just as one can think of metaphysically possible hypotheses as corresponding to an overarching space of metaphysical possibilities, one can think of epistemically possible hypotheses as corresponding to an overarching space of epistemic possibilities. Some possibilities in the space of metaphysical possibilities are maximally specific: these can be thought of as *maximal metaphysical possibilities*, or as they are often known, possible worlds. In a similar way, some possibilities in the space of epistemic possibilities are maximally specific: these can be thought of as *maximal epistemic possibilities*, or as I will call them, *scenarios*.

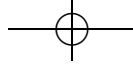
A scenario corresponds, intuitively, to a maximally specific way the world might be, for all one can know a priori. Scenarios stand to epistemic possibility as possible worlds stand to metaphysical possibility. Indeed, it is natural to think of a scenario as a sort of possible world, or better, as a *centered* possible world. There are some complications here, but for the moment it is helpful to think of scenarios intuitively in such terms.

For any scenario, it is epistemically possible that the scenario is actual. Intuitively speaking, for any qualitatively specified centered world W , it is epistemically possible that W is actual. Here the center represents a hypothesis about my own location within the world. In entertaining the hypothesis that W is actual, I entertain the hypothesis that the actual world is qualitatively just like W , that I am the subject at the center of W , and that now is the time at the center of W .

For example, let the XYZ-world be a specific centered “Twin Earth” world, in which the subject at the center is surrounded by XYZ in the oceans and lakes. Then no amount of a priori reasoning can rule out the hypothesis that the XYZ-world is my actual world: i.e., that I am in fact living in such a world, where the liquid in the oceans and lakes around me is XYZ. So the XYZ-world represents a highly specific epistemic possibility.

When we think of a world as an epistemic possibility in this way, we are *considering it as actual*. On the epistemic understanding, to consider a world W as actual is to consider the hypothesis that W is one’s own world. When one considers such a hypothesis, in effect one considers the hypothesis that D is the case, where D is a statement giving an appropriate description of W . One can think of D , intuitively, as a description of W in neutral qualitative terms, along with a specification in indexical terms of a center’s location in W . I will return to this matter later.

The second key idea is that of scrutability: the idea that there is a strong epistemic dependence of an expression’s extension on the state of the world. If we come to know that the world has a certain character, we are in a position to conclude that the expression has a certain extension. And if we were to learn that the world has a different character, we would be in a position to conclude the expression has a different



extension. That is: we are in a position to come to know the extension of an expression, depending on which epistemic possibility turns out to be actual.

If we take the case of ‘Water is H_2O ’: we can say that given that the world turns out as it actually has, with H_2O in the oceans and lakes, then it turns out that water is H_2O . So if the H_2O -world is actual, water is H_2O . But if we were to discover that the oceans and lakes in the actual world contained XYZ, we would judge that water is XYZ. And even now, we can judge: *if* it turns out that the liquid in the oceans and lakes is XYZ, it will turn out that water is XYZ. Or we can simply say: if the XYZ-world is actual, then water is XYZ.

The same goes more generally. If W_1 is a specific scenario in which the morning and evening stars are the same, and W_2 is a scenario in which the morning and evening stars are different, then we can say: if W_1 is actual, then Hesperus is Phosphorus; if W_2 is actual, then Hesperus is not Phosphorus. The same goes, in principle, for a very wide range of scenarios and statements. Given a statement S, and given enough information about an epistemically possible state of the world, we are in a position to judge whether, *if* that state of the world obtains, S is the case.

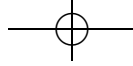
All this is reflected in the way we use language to describe and evaluate epistemic possibilities. It is epistemically possible that water is XYZ. It is also epistemically possible that the XYZ-world is actual. And intuitively speaking, the epistemic possibility that the XYZ-world is actual is an *instance* of the epistemic possibility that water is XYZ. We can say as above: *if* the XYZ-world turns out to be actual, it will turn out that water is XYZ. We might also use a straightforward indicative conditional: if the XYZ-world is actual, then water is XYZ. Or we can use the Ramsey test, commonly used to evaluate indicative conditionals: if I hypothetically accept that the XYZ-world is actual, I should hypothetically conclude that water is XYZ.

We can put all this by saying that the XYZ-world *verifies* ‘Water is XYZ’, where verification is a way of expressing the intuitive relation between scenarios and sentences described above.⁷ Intuitively, a scenario W verifies a sentence S when the epistemic possibility that W is actual is an instance of the epistemic possibility that S is the case; or when we judge that if W turns out to be actual, it will turn out that S is the case; or if the indicative conditional ‘if W is actual, then S is the case’ is rationally assertible, or if hypothetically accepting that W is actual leads to hypothetically concluding that S is the case. We can also say that when W verifies S, W makes S true when it is *considered as actual*. Verification captures the way that we use language to describe and evaluate epistemic possibilities.

This dependence can be represented by the *epistemic intension* of a sentence S. This is a function from scenarios to truth-values. If a scenario W verifies S, then S’s epistemic intension is true at W; if W verifies $\sim S$, then S’s epistemic intension is false at W; otherwise, S’s epistemic intension is indeterminate at W. So the epistemic intension of ‘Water is XYZ’ is true at the XYZ-world.

Given this intuitive conception of epistemic intensions, there is a strong *prima facie* case that they satisfy the Core Thesis. When S is a priori, we would expect that

⁷ The term ‘verify’ is used for a related idea in Evans (1979). See also Yablo (1999).



every scenario verifies S . And when S is not a priori, $\sim S$ is epistemically possible, so we would expect that there is a scenario that verifies $\sim S$. If these claims hold true, then S is a priori iff S has a necessary epistemic intension (one that is true at all scenarios).

Epistemic intensions resemble contextual intensions in some superficial respects, but they are fundamentally quite different. The central difference, as we will see, is that epistemic intensions are defined in epistemic terms. From what we have seen so far, epistemic intensions behave at least somewhat as one would like a quasi-Fregean 1-intension to behave. But to investigate this matter, we must define the relevant notions more precisely.

3.2 Epistemic intensions

The intuitive picture of the epistemic understanding above can be regarded as capturing what is essential to an epistemic understanding. To fill in the picture, however, a more precise analysis is required. What follows is one way to flesh out these details. Not all of the details that follow are essential to an epistemic account *per se*, but they provide a natural way of elaborating such an account.

Starting with the intuitive picture, we can say that the epistemic intension of a sentence token is a function from a space of scenarios to the set of truth-values, such that:

The epistemic intension of a sentence token S is true at a scenario W iff the hypothesis that W is actual epistemically necessitates S .

When the conditions specified here obtain, we can also say that W *verifies* S . The epistemic intension of S will be false at W when W verifies $\sim S$, and it will be indeterminate at W when W verifies neither S nor $\sim S$.

Rather than leaving the notion of “the hypothesis that W is actual” as primitive, it is useful (although not mandatory) to invoke the notion of a *canonical description* of a scenario. We can then characterize an epistemic intension as follows.

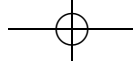
The epistemic intension of a sentence token S is true at a scenario W iff D epistemically necessitates S , where D is a canonical description of W .

It remains to clarify three notions: the notion of a scenario, that of a canonical description, and that of epistemic necessitation. I investigate each of these in what follows.⁸

3.3 Epistemic necessitation

First, we need to say more about epistemic possibility and necessity. The epistemic understanding of two-dimensional semantics is grounded in a notion of *deep epistemic possibility*, or equivalently, of *deep epistemic necessity*. In the ordinary sense, we say

⁸ Note that some of these details are necessarily complex, and some readers may prefer to skim the remainder of this section or skip ahead to Section 4 on a first reading. Some other papers cover some of this material in more depth: notably, “The Nature of Epistemic Space” (Chalmers forthcoming), which covers the issues in 3.4 in more detail; “Conceptual Analysis and Reductive Explanation” (Chalmers and Jackson 2001), which is especially relevant to the issues in 3.6; and “Does Conceivability Entail Possibility?” and “On Sense and Intension” (Chalmers 2002a and 2002b), which discuss a number of aspects of these issues that are not discussed here.



that S is epistemically possible roughly when S may be the case for all we know, and that S is epistemically necessary roughly when we are in a position to know that S is the case. A notion of deep epistemic necessity goes beyond this sort of dependence on the shifting state of an individual's knowledge, to capture some sort of rational *must*: a statement is deeply epistemically necessary when in some sense, it rationally must be true.

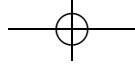
Such a notion can be understood in various ways, but for our purposes there is a natural candidate. We can say that S is deeply epistemically necessary when it is *a priori*: that is, when the thought expressed by S expresses actual or potential a priori knowledge. (I say more about the notion of apriority in Section 3.9.) Then S is deeply epistemically possible when the negation of S is not epistemically necessary: that is, when the thought that S expresses cannot be ruled out a priori. Henceforth, I will usually drop the modifiers “deep” and “deeply”, and speak simply of epistemic possibility and necessity.

In this sense, ‘Water is XYZ’ is epistemically possible: one cannot know a priori that water is not XYZ. In the same way, ‘Hesperus is not Phosphorus’ is epistemically possible, as is ‘I am not a philosopher’. On the other hand, ‘Some bachelors are married’ is not epistemically possible, and ‘All bachelors are married’ is epistemically necessary. Similarly, one can argue that ‘Hesperus is not Hesperus’ is epistemically impossible, and that its negation is epistemically necessary.

A claim is deeply epistemically possible, intuitively speaking, when it expresses a rationally coherent hypothesis about the actual world. The standards of rational coherence here are in one sense weaker than usual: if a hypothesis conflicts with empirical knowledge, it may still be deeply epistemically possible. The standards are in another sense stronger than usual: if a hypothesis can be ruled out only by a great amount of a priori reasoning, it is nevertheless deeply epistemically impossible. It is possible to define notions of possibility that meet different standards, but the current standards are best for our current purposes.

The epistemic necessity operator applies to both sentence types and sentence tokens. We require this as the sentences S whose epistemic intensions we are defining are tokens, and it is possible for two sentence tokens of the same linguistic type to have different epistemic properties (for the reasons, see Section 3.8). The canonical descriptions D of scenarios, on the other hand, are sentence types, using expressions whose epistemic properties are fixed by the language. We also need an epistemic necessitation operator between sentence types of this sort and sentence tokens.

An epistemic necessity operator of this sort can be seen as a primitive of the system I am developing. On the picture where epistemic necessity corresponds to apriority, we can characterize its properties intuitively as follows. Let us say that *thoughts* are the sort of occurrent propositional attitudes expressed by assertive sentences. Then a sentence token S is epistemically necessary when the thought expressed by S can be justified independently of experience, yielding a priori knowledge. A sentence type D is a priori when it is possible for a token of S to be epistemically necessary. A sentence type D epistemically necessitates a sentence token S when a material conditional ‘ $D \supset S$ ’ is epistemically necessary, where this is understood as a possible token material conditional whose constituent token of S expresses the same thought as the original



token. I will say more about the characterization of epistemic necessity in Section 3.9, but this understanding will suffice for present purposes.

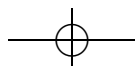
We can now say that a scenario W verifies a sentence token S when a material conditional ' $D \supset S$ ' is epistemically necessary, where D is a canonical description of W . If epistemic necessity is understood as apriority, then on this model a scenario W verifies a sentence S when one could in principle rule out a priori the hypothesis that W is actual but S is not the case.

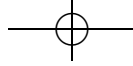
This definition works naturally with the characterizations we will give of scenarios and of canonical descriptions, but it should be noted that this is not the only possible definition. There are various ways in which an epistemic framework might characterize the required relationship between D and S in other terms, which need not appeal directly to notions such as apriority.

For example, one might appeal to the intuitive heuristics described earlier. One could say that W verifies S when the epistemic possibility that W is actual is an instance of the epistemic possibility that S is the case. Or appealing to canonical descriptions, one could say that W verifies S when the epistemic possibility that D is the case is an instance of the epistemic possibility that S is the case. Here one might leave this intuitive evaluation of epistemic possibilities as a primitive, much as the intuitive evaluation of counterfactual possibilities is often taken as a primitive.

Alternatively, one could ground epistemic necessitation in indicative conditionals: D epistemically necessitates S when the indicative conditional "if D is the case, then S is the case" is intuitively acceptable on rational reflection. (See Chalmers (1998) for a discussion of this approach.) In a closely related idea, one could ground epistemic necessitation in the Ramsey test: D epistemically necessitates S (relative to a subject) when if the subject hypothetically accepts that D is the case, the subject should rationally conclude that S is the case. The latter approach yields what we might call the *Ramsey intension* of an expression: the Ramsey intension of a subject's expression S is true at W when if the subject hypothetically accepts that D is the case (where D is a canonical description of W), the subject should rationally conclude that S is the case.

Ramsey intensions behave very much like epistemic intensions as defined above. It is plausible they often yield the same results: for example, both the epistemic intension and the Ramsey intension of "water is H_2O " are plausibly false at the XYZ-world. There are arguably some cases where they yield different results. For example, Yablo (2002) has argued that the indicative conditional "if 'tail' means leg, then tails are legs" is acceptable. If so, then the Ramsey intension of 'tails are legs' may be true in a world where 'tail' means legs, but the epistemic intension will not. (See Chalmers (2002a) for discussion.) Likewise, if I accept that I have recently been given a drug that corrupts my adding abilities, then I should arguably suspend judgment about whether 57 plus 46 is 103. If so, the Ramsey intension of " $57 + 46 = 103$ " will plausibly be indeterminate in a scenario where the subject at the center has been given such a drug, but the epistemic intension will not. It may be that the Ramsey test can be understood in a way that handles the cases above differently, so that Ramsey intensions behave in the way that a Fregean intension should, but the matter is not entirely clear.





Ramsey intensions are a sort of epistemic intension in the general sense, as they are defined in epistemic terms. But where epistemic intensions as defined above are grounded in the notion of apriority, Ramsey intensions are grounded in the notion of rational inference. This has certain advantages: for example, those who are skeptical about apriority usually still accept that there is a coherent notion of rational inference. In what follows I will usually stay with epistemic intensions grounded in a notion of apriority, but the possibility of alternative understandings should be kept in mind.

These alternative understandings suggest that the epistemic understanding of the two-dimensional framework is not entirely beholden to the notion of apriority. Even if one rejects apriority, or if one rejects the application of apriority in this context, one should not reject the epistemic understanding. It is a *prima facie* datum that there is an epistemic dependence between epistemic possibilities and sentence tokens of the sort that was intuitively characterized earlier. One who rejects apriority will simply need to capture this dependence in other ways. My own view is that the understanding in terms of apriority runs the deepest, but the alternatives deserve exploration.

We can here note a fundamental difference between all of these sorts of epistemic evaluation and contextual evaluation. To evaluate a sentence *S* in a scenario *W*, there is no requirement that *W* contain a token of *S*. Even if *W* contains such a token, the definition gives it no special role to play. All that matters is the first-order epistemic relation between *D* and *S*, not whether *D* says something metalinguistic about a token of *S*. More generally, metalinguistic facts about how a token of *S* would behave in certain possible circumstances play no role in defining epistemic intensions. This enables us to deal straightforwardly with the problem cases for contextual intensions.

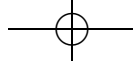
3.4 Scenarios

Scenarios are intended to stand to epistemic possibility as possible worlds stand to metaphysical possibility. This claim can be expressed by the following:

Plenitude Principle: For all *S*, *S* is epistemically possible if and only if there is a scenario that verifies *S*.

In effect, the Plenitude Principle says that there are enough scenarios to verify every epistemically possible claim, and that no scenario verifies an epistemically impossible claim. It is easy to see that if we understand epistemic necessity as apriority, the Plenitude Principle is equivalent to the Core Thesis. (I give it a different name to leave open the option of understanding epistemic necessity in different terms.) So the only question is whether we can understand scenarios and verification so that the Plenitude Principle is true.

Intuitively, a scenario should correspond to a maximally specific epistemically possible hypothesis, or (for short) a maximal hypothesis: a hypothesis such that if one knew that it were true, one would be in a position to know any truth by reasoning alone. (Note that talk of “hypotheses” here is intuitive; formalizations of the relevant notions will follow.) We might say that a hypothesis *H*₁ leaves another hypothesis *H*₂ open if the conjunctions of *H*₁ with both *H*₂ and its negation are epistemically possible. A maximal hypothesis is one that leaves no possible hypothesis open. To every scenario, there should correspond a maximal hypothesis, and vice versa.



3.4.1 Scenarios as centered worlds

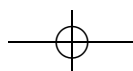
There are two concrete ways in which we might understand scenarios. The first is the way we have already sketched: as centered possible worlds. The uncentered part of the world corresponds to a hypothesis about the objective character of one's world. The centered part is needed to handle indexical claims, such as "I am in Australia". If we are given only a full objective description of a world, numerous indexical hypotheses will be left open, so such a description does not correspond to a maximal hypothesis. Correspondingly, there are numerous epistemically possible (but incompatible) objective-indexical claims: for example "the world is objectively thus and I am a philosopher" and "the world is objectively thus and I am not a philosopher". We need distinct scenarios to verify these claims: hence centered worlds.

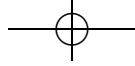
There is good reason to believe that for every centered world, there is a corresponding maximal hypothesis, at least if we describe worlds under the right sort of canonical description. (It is arguable that for certain indexical hypotheses involving demonstratives, one may need further information in the center of the world: marked experiences, as well as a marked subject and time. But I will leave this matter to one side.) And one can easily make the case that an epistemically impossible sentence will be verified by no centered world (if it were so verified, it would not be epistemically impossible). The residual question is whether there are enough centered worlds to correspond to *all* maximal hypotheses, and to verify all epistemically possible statements. This matters turns on the following thesis:

Metaphysical Plenitude: For all S, if S is epistemically possible, there is a centered metaphysically possible world that verifies S.

The standard Kripkean cases of statements that are epistemically possible but metaphysically impossible are straightforwardly compatible with this thesis. For each such statement S, there is *some* way the world could turn out such that if things turn out that way, it will turn out that S is the case; and each of these ways the world could turn out can be seen as a centered world. In the case of 'Water is XYZ', the XYZ-world is such a world; something similar applies to other cases. One might worry about how a metaphysically possible world (the XYZ-world) can verify a metaphysically impossible statement ('Water is XYZ'). But two-dimensional evaluation makes this straightforward: 'Water is XYZ' is true at the XYZ-world considered as actual, but false at the XYZ-world considered as counterfactual. The metaphysical impossibility of 'Water is XYZ' reflects the fact that it is false at all worlds considered as counterfactual. But this is quite compatible with its being true at some worlds considered as actual.

Are there any counterexamples to the Metaphysical Plenitude thesis? I have argued elsewhere (Chalmers 2002a) that there are no such counterexamples. Certainly, there are no *clear* cases of epistemically possible claims that are verified by no centered world. Still, some controversial philosophical views entail that there are such cases. For example, some theists hold that it is necessary that an omniscient being exists, while also holding that it is not a priori that an omniscient being exists. If so, "No omniscient being exists" will be a counterexample to Metaphysical Plenitude: it will





be an epistemically possible statement that is verified by no possible world. In effect, on this view the space of metaphysical possibilities is *smaller* in some respects than the space of epistemic possibilities.

The same goes for some other philosophical views. On some views on which the laws of nature of our world are the laws of all worlds, for example, the negation of a law of nature will be a counterexample to Metaphysical Plenitude. On views on which a mathematical claim (such as the Continuum Hypothesis) can be necessarily true but not knowable a priori, the negation of such a claim will be a counterexample to Metaphysical Plenitude. On some versions of the epistemic theory of vagueness, some claims involving vague terms (e.g. the statement that someone of a certain height is tall) may be a counterexample to Metaphysical Plenitude. On some materialist views about consciousness, the claim that there are zombies (unconscious physical duplicates of conscious beings) may be a counterexample to Metaphysical Plenitude. If these views are correct, there will be epistemically possible claims that are not verified by any centered metaphysically possible worlds. If so, Metaphysical Plenitude (and the Core Thesis for epistemic intensions over centered metaphysically possible worlds) will be false.

All of these views are highly controversial, and I have argued elsewhere (Chalmers 2002a) that all of them are incorrect. One can plausibly argue in reverse: the Metaphysical Plenitude thesis, which appears to fit all standard cases, gives us reason to reject these controversial views. More deeply, one can argue that these views rest on a mistaken conception of metaphysical possibility and necessity. My own view is that a careful analysis of the roots of our modal concepts supports constitutive links between epistemic and metaphysical modal notions, and thereby grounds the Metaphysical Plenitude thesis. If this is correct, then understanding scenarios in terms of centered worlds yields epistemic intensions that satisfy the Core Thesis.

It is nevertheless useful to have an approach to the space of epistemic possibilities that is neutral on these substantive questions about metaphysical possibility. This allows even those philosophers who deny Metaphysical Plenitude to make use of the notion of an epistemic intension, and allows a maximally general defense of the epistemic understanding of two-dimensional semantics.

3.4.2 *Scenarios as maximal hypotheses*

The alternative is to understand scenarios in purely epistemic terms from the start. One might reasonably hold that since we want epistemic intensions to be constitutively connected to the epistemic realm, we need not invoke the metaphysical modality at all. Instead, we can do things wholly in terms of the epistemic modality. There are a couple of ways one might proceed here. One could introduce the notion of a scenario (a maximal epistemic possibility) as a modal primitive, in the same way that some philosophers introduce the notion of a world (a maximal metaphysical possibility) as a modal primitive. Or one could try to *construct* scenarios directly out of materials that are already at hand.

I take the second course in Chalmers (forthcoming), examining a detailed construction. I do not have space to do that here, but I can give a brief idea of how one might proceed. The idea I will outline is a linguistic construction of scenarios,

constructed out of linguistic expressions in an idealized language, along with a basic operator of epistemic possibility.

Let us say that a sentence D of a language L is *epistemically complete* when (i) D is epistemically possible, and (ii) there is no sentence S of L such that both $D \& S$ and $D \& \sim S$ are epistemically possible. When D is epistemically complete, it is in effect as specific as an epistemically possible sentence can be. Let us say that D is *compatible* with H when $D \& H$ is epistemically possible, and D *implies* H when $D \& \sim H$ is epistemically impossible (that is, when there is an a priori entailment from D to H). Then if D is epistemically incomplete, it leaves questions open: there will be H such that D is compatible with H but D does not imply H . If D is epistemically complete, D leaves no questions open: if D is compatible with H , D implies H . Note that D need not explicitly include every such hypothesis as a conjunct; these hypotheses need only be implied.

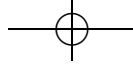
Intuitively, scenarios should correspond to epistemically complete *hypotheses*, whether or not they are expressible in a language such as English. It is likely that actual languages do not have the expressive resources to express an epistemically complete hypothesis, as they are restricted to finite sentences and have a limited lexicon. So for the purposes of this construction, we need to presuppose an idealized language that can express arbitrary hypotheses. In particular, our language L should allow infinitary sentences (at least infinitary conjunctions) and should have terms that express every possible concept, or at least every concept of a certain sort. It is also important that expressions in L are *epistemically invariant*, so that there cannot be two tokens S_1 and S_2 of the same sentence type (used with full competence) such that S_1 is epistemically necessary and S_2 is not. The exact requirements for L raise subtle issues, but we can pass over them here.

We can then focus on epistemically complete sentences of L . By the idealization, every such sentence will express a maximally specific hypothesis, and vice versa. So scenarios should correspond to epistemically complete sentences in L , although perhaps with more than one such sentence per scenario. We can say that two sentences S and T are *equivalent* when S implies T and T implies S (that is, when $S \& \sim T$ and $T \& \sim S$ are epistemically impossible). Any epistemically complete sentences in L will then fall into an equivalence class. We can now identify scenarios with equivalence classes of epistemically complete sentences in L . To anticipate the definition of verification: we can also say that a scenario verifies a sentence S (of an arbitrary language) when D implies S , where D is an epistemically complete sentence of L in the scenario's equivalence class.

Defined this way, scenarios are tailor-made to satisfy the Plenitude Principle. This principle requires the following:

Epistemic Plenitude: For all S , if S is epistemically possible, then some epistemically complete sentence of L implies S .

Here S may be a sentence token in any language (not necessarily in L). To see the plausibility of this thesis, first note that because L has unlimited expressive power, some epistemically possible sentence S_1 of L will imply S . Second, it is plausible that any epistemically possible sentence S_1 of L is implied by some epistemically complete



sentence D of L. Intuitively, to obtain D from S_1 , one simply conjoins arbitrary sentences that are epistemically compatible with S_1 (and other conjoined sentences) until one can conjoin no more. The issue is not completely trivial, as there might be endless infinitary conjunction with no maximal point, but under certain reasonable assumptions, such a sentence will exist. If so, then every epistemically possible sentence is verified by some scenario. In reverse, it is clear that any sentence verified by a scenario is epistemically possible. So the corresponding version of the Plenitude Principle is plausibly true.

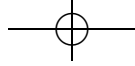
In effect, this construction formalizes the intuitive idea of a maximal hypothesis: a maximal hypothesis is equivalent to an equivalence class of epistemically complete sentences in an idealized language. We might say that where the first approach takes a *metaphysical* approach to scenarios, on which they correspond to centered metaphysically possible worlds, the second approach takes an *epistemic* approach to scenarios, on which they correspond to maximal hypotheses.

What is the relationship between the two constructions? My own view is that there is a close correspondence: every centered world corresponds to a maximal hypothesis, and every maximal hypothesis corresponds to a centered world. (Not quite one-to-one: in certain cases there may be more than one centered world per maximal hypothesis, for example when there are symmetrical worlds with symmetrically corresponding centers.⁹) If so, then the Plenitude Principle will plausibly be satisfied either way. But philosophers who deny Metaphysical Plenitude will deny the close correspondence, holding that there are maximal hypotheses that correspond to no centered world. For example, a philosopher who holds that ‘There is an omniscient being’ is necessary but not a priori will hold that there is a maximal hypothesis that verifies the negation of the sentence in question, but that there is no centered metaphysically possible world in the vicinity. Such a philosopher should embrace the epistemic approach to scenarios.

The epistemic approach to scenarios is grounded more purely in the epistemic realm, and its central theses require fewer commitments than the metaphysical approach. For this reason, one can argue that the epistemic approach to scenarios is more basic. Centered worlds are more familiar and are useful for various applications, however, so I will use both understandings of scenarios in what follows.

On either understanding, one scenario will be privileged with respect to any statement token as the *actualized* scenario at that token. On the world-based view, this will be the world centered on the speaker and the time of utterance. On the epistemic view, this will correspond to the maximal hypothesis that is true of the world from the speaker’s perspective at the time of utterance. In general, we expect that when an

⁹ See Chalmers (forthcoming), section 4(4) for more on ways in which there could be more than one centered world per maximal hypothesis. Schroeter (2004) raises the possibility that there are intrinsic properties for which there is no semantically neutral conception. If there are such properties, then this is another source of a many-to-one correspondence. If such properties exist, then an epistemically complete description of a centered world may not need to specify their precise distribution. If so, then an epistemically complete description need not be ontologically complete, and more than one centered world (with different but isomorphic distributions of intrinsic properties) may correspond to the same maximal hypothesis.



expression token's epistemic intension is evaluated at the scenario that is actualized at that token, the result will be the token's extension.

3.5 Canonical descriptions

When we consider a scenario as actual, in order to evaluate an expression, we always grasp it under a description. This raises an issue. A scenario can be described in multiple ways, and it is not obvious that all such descriptions will give equivalent results. So we have to isolate a special class of *canonical descriptions* of scenarios under which they must be considered.

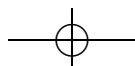
If we take the epistemic approach to scenarios by the second construction above, the choice will be straightforward. A scenario will correspond to an equivalence class of epistemically complete sentences. Here, we can say that a canonical description of the scenario is any sentence in the corresponding equivalence class. Because all of these sentences are equivalent under implication, they will all give the same results under verification.

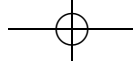
If we take the metaphysical approach to scenarios, things are more complicated. Here, we require that a canonical description be a *complete neutral description* of the world. Both neutrality and completeness need explanation.

First, neutrality. To describe a world, we must choose sentences that are true of it. But will these be sentences true of the world considered as actual, or of the world considered as counterfactual? If we choose the first, there is a danger of circularity: evaluation of a world considered as actual will be defined in terms of canonical descriptions, which will be defined in terms of evaluation of worlds considered as actual. If we choose the second, there is a danger of incoherence: the framework requires that the XYZ-world verifies 'water is not H₂O', but 'Water is H₂O' is true of the XYZ-world considered as counterfactual. Either way, we need to ensure that sentences such as 'Water is H₂O' are not present within canonical descriptions of the XYZ-world.

The solution is to restrict canonical descriptions to *semantically neutral* expressions. Intuitively, a semantically neutral expression is one that behaves the same whether one considers a world as actual or as counterfactual. We cannot simply *define* a semantically neutral expression in this way, since the definition presupposes evaluation in a world considered as actual, and this evaluation (as developed here) presupposes the notion of a canonical description. But nevertheless we have a good grasp on the notion. For example, 'water' and 'Hesperus' are not semantically neutral; but 'and', 'philosopher', 'friend', 'consciousness', and 'cause' plausibly are. One could rely on our intuitive grasp of this notion for current purposes, or one could seek to define it.

One promising approach is to define such an expression as one that is not "Twin-Earthable". We can say that two possible individuals (at times) are twins if they are physical and phenomenal duplicates; we can say that two possible expression tokens are twins if they are produced by corresponding acts of twin speakers. Then a token is Twin-Earthable if it has a twin with a different 2-intension. This test works for many purposes. A semantically neutral term (in the intuitive sense) is never Twin-Earthable. But the reverse is not quite the case. For example, let L be an expression that functions to rigidly designate the speaker's height. Then any twin of L will have





the same 2-intension (since a twin speaker will have the same height), but L is not semantically neutral. One might respond by watering down the requirements of physical and phenomenal duplication (perhaps to some sort of mental duplication), but similar cases will still arise: e.g. if M is an expression that rigidly picks out 1 if the speaker has visual experience, and 0 if not, then M is not Twin-Earthable even by this sort of standard, but it is nevertheless not semantically neutral.¹⁰

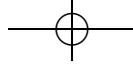
A better characterization might be as follows: a semantically neutral expression is one whose extension in counterfactual worlds does not depend on how the actual world turns out (that is, on which epistemically possible scenario turns out to be actual). This is an intuitive characterization rather than a formal characterization: it invokes the intuitive idea of dependence of counterfactual extensions on the actual world, and formalizing this idea would require something equivalent to the two-dimensional framework (with ensuing circularity). But nevertheless, we have a good grip on the notion. In this sense, it is clear that most names, natural kind terms, and indexicals are not semantically neutral (and neither are L or M above), while numerous other terms (such as those listed above) are plausibly semantically neutral.

A precise formal characterization of semantic neutrality remains an open question for future research. One might try a characterization wholly in terms of our modal operators of epistemic and metaphysical necessity (that is, apriority and necessity), but it is not entirely clear how this would work. In the meantime, the intuitive characterization suffices for our purposes. It is also useful to stipulate that terms with context-dependent behavior, such as “heavy”, are not semantically neutral. This allows us to describe worlds using expression types and not just expression tokens.

To characterize a centered world, semantically neutral terms must be supplemented by some indexical terms, to characterize the location of a center. The best way to do this is the following. We can say that a statement is in canonical form when it has the form $D \ \& \ 'I \text{ am } D_1' \ \& \ 'now \text{ is } D_2'$, where D , D_1 , and D_2 are all semantically neutral, and D_1 and D_2 are identifying predicates relative to the information in D (that is: D implies ‘Exactly one individual is D_1 ’ and ‘Exactly one time is D_2 ’). We can say that a *neutral description* of a centered world is a statement in canonical form such that D is true of the world, D_1 is true of the subject at the center, and D_2 is true of the time at the center. (If the center of a centered world includes entities other than an individual and a time, then one can extend similar treatment to these entities.)

In a few cases involving completely symmetrical worlds, there may be no identifying predicates available: that is, there may be no semantically neutral predicates true only of the individual (or time) at the center. In that case, one can invoke a maximally specific predicate instead: a predicate D_1 such that for all D_2 true of the center, D entails ‘everything that is D_1 is D_2 ’. Here, two centered worlds that differ only in

¹⁰ Non-Twin-Earthability is related to Bealer’s (1996) notion of semantic stability: “an expression is semantically stable iff, necessarily, in any language group in an epistemic situation qualitatively identical to ours, the expression would mean the same thing” (Bealer 1996, 134). It is clear that semantic stability cannot be used to characterize semantic neutrality, for the same reasons as in the case of non-Twin-Earthability. For example, the expression M in the text is semantically stable but not semantically neutral.



symmetrical placement of the center may yield the same canonical description. This is reasonable, as intuitively both worlds correspond to the same maximal hypothesis.

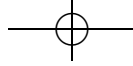
Second, completeness. We require that a canonical description be a *complete* neutral description of a centered world. There are two possibilities here. First, we can appeal to a criterion in terms of (metaphysical) necessity. Let us say that a semantically neutral description of a world is ontologically full when it (metaphysically) necessitates all semantically neutral truths about that world, and is minimal among the class of descriptions with this property. For example, if physicalism is true, a full semantically neutral specification of fundamental physical truths will be ontologically full. Then an ontologically complete neutral description of a centered world is a neutral description where the first (non-indexical) component of the description is ontologically full.

Alternatively, we can appeal to epistemic completeness. In this sense, a complete neutral description of a centered world is simply a neutral description that is epistemically complete. This requires the claim that for any centered world, there exists an epistemically complete neutral description. This claim is nontrivial, but there are good grounds to accept it. One can argue that although non-neutral terms are modally distinctive, they do not add fundamentally new *epistemic* power to a language, so that neutral terms constitute what I call an *epistemic basis* (see Section 3.6) for the space of epistemic possibilities.

It is not hard to see that if Metaphysical Plenitude is correct, then an ontologically complete neutral description will also be an epistemically complete neutral description.¹¹ If so, we can then use either criterion for a canonical description. There will arguably be more explanatory power, however, in using a complete description in the ontological sense, and then allowing this description to epistemically determine all truths about a world considered as actual.

If Metaphysical Plenitude is false, then the two criteria will not coincide. An ontologically complete neutral description will not be epistemically complete, and it will leave some hypotheses unsettled (e.g. the complete physical truth about the world may leave the Continuum Hypothesis unsettled, even if it is necessarily true). If we require that canonical descriptions be ontologically complete, the epistemic intensions of these hypotheses will have an indeterminate truth-value. A consequence may be that when an expression's epistemic intension is evaluated at the actual centered world of the expression, it does not yield the expression's extension (e.g., the epistemic intension of CH may be indeterminate at the actual world, even if CH is true). If, on the other hand, we require epistemic completeness, then the epistemic intensions of the relevant claims will have a determinate truth-value (e.g. the epistemic intension of CH will be true or false at the world according to whether CH itself is true or false there). One might do things either way, depending on

¹¹ The statement of Metaphysical Plenitude uses the notion of verification, which in turn requires the notion of a canonical description. For the purposes of interpreting Metaphysical Plenitude, we can assume that the canonical descriptions are required to be epistemically complete. *If* Metaphysical Plenitude formulated this way is correct, ontologically complete descriptions will give the same results as epistemically complete descriptions.



one's purposes, although for most purposes it is probably best to require epistemic completeness overall. In any case, this situation will not matter much for our purposes, since we already know that if Metaphysical Plenitude is false, then the Core Thesis will be false when scenarios are understood as centered worlds.

(A third alternative is to require "qualitative completeness", where this is characterized as in Chalmers (2002a) in terms of a notion of positive conceivability. This yields a notion that is usefully intermediate between epistemic completeness and ontological completeness. But I will leave this option aside here.)

It is clear that if scenarios are understood as centered worlds, the characterization of canonical descriptions is significantly more complicated than if scenarios are understood in wholly epistemic terms. This may be another point in favor of the purely epistemic understanding of scenarios.

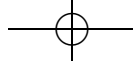
3.6 Scrutability

Given the epistemic understanding of scenarios, one might have the following worry: the epistemic intension of a sentence may be well-defined, but it is *trivial*. The triviality comes from the requirement that descriptions be epistemically complete. One may worry that in order for a description to be epistemically complete, it will need to specify the truth or falsity of most sentences *S* explicitly. For example, 'Water is H₂O' will be true precisely in those scenarios that have 'Water is H₂O' in their canonical description, and it will be false precisely in those scenarios that have 'Water is not H₂O' in their canonical description. If this sort of thing is typical, then epistemic evaluation as defined will have an uninteresting structure.

A related worry arises on the metaphysical understanding of scenarios. Here, the issue concerns the thesis (mentioned in the previous section) that there is an epistemically complete neutral description of any centered world. If one had the worry just mentioned about 'Water is H₂O', one might worry that an epistemically complete description of a centered world requires non-neutral terms, such as 'water'. The key question is whether the truth-value of all sentences *S* is epistemically necessitated by a description of a centered world in terms of semantically neutral expressions plus indexicals. If this is not the case, then as defined, the epistemic intension of the relevant sentences will be indeterminate at the relevant centered worlds.

These worries are reasonable enough, but I think that they are ultimately unfounded. In what follows, I will concentrate on the worry that applies to the epistemic understanding, but similar considerations also apply to the metaphysical understanding. To answer the worry, one needs to make the case that epistemically complete descriptions do not need to specify the truth or falsity of most statements explicitly, so that epistemic evaluation does not have a trivial structure. To see this, it is useful to focus on the *actual* world, and consider what an epistemically complete description of this world must contain. The sort of argument I give here is presented in much more depth by Chalmers and Jackson (2001) and Chalmers (2002a); but here I will give the basic idea.¹²

¹² Chalmers and Jackson (2001) can be seen as providing a crucial part of the foundation for the two-dimensional framework as it is understood here, even though the framework is hardly



The second principle underlying the epistemic understanding of the two-dimensional framework was what we might call the *scrutability of truth*. This can be put informally as the thesis that once we know enough about the state of the world, we are in a position to know the truth-values of our sentences. Furthermore, we usually need not be informed about a sentence explicitly in order to know whether it is true. We could put this somewhat more precisely as follows:

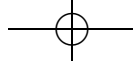
Scrutability of Truth: For most terms T used by a speaker, then for any truth S involving T , there exists a truth D such that D is independent of T , and such that knowing that D is the case puts the speaker in a position to know (without further empirical information, on idealized rational reflection) that S is the case.

Here, we can say that D is independent of T when D does not contain T or any close cognates. Of course this notion is somewhat vague, as is the notion of “most” above, but this does not matter for our purposes. To save breath, we can abbreviate “knowing that D is the case puts the speaker in a position to know (without further empirical information, on idealized rational reflection) that S is the case” as “ D is epistemically sufficient for S ”.

Take the case of ‘water’. Here, we can let D be a truth specifying an appropriate amount of information about the appearance, behavior, composition, and distribution of objects and substances in one’s environment, as well as information about their relationship to oneself. D need not contain the term ‘water’ at any point: appearance can be specified in phenomenal terms, behavior and distribution in spatiotemporal terms, composition in microphysical or chemical terms. Then D is epistemically sufficient for ‘Water is H_2O ’. When one knows that D is the case, one will be in a position to know all about the chemical makeup of various liquids with various superficial properties in one’s environment, and will thereby be able to infer that water is H_2O . After all, this information about appearance, behavior, composition, and distribution is roughly what we need in the case of ordinary knowledge, to determine that water is H_2O . And there is no need for further empirical information to play a role here: even if we suspend all other empirical beliefs, we can know that *if* D is the case, then water is H_2O .

The same goes for terms such as ‘Hesperus’. Once again, if D contains appropriate information about the appearance, behavior, composition, and distribution of various objects in the world, then D is epistemically sufficient for ‘Hesperus is Venus’, for ‘All renates are cordates’, and so on. The information in D enables one to know that the object that presents a certain appearance in the evening is the same as the object that presents a certain appearance in the morning, and so enables us to know that Hesperus is Phosphorus. Something similar applies to ‘heat’, ‘renate’, and so on.

mentioned in the paper (which is packaged as a response to Block and Stalnaker on the explanatory gap). Section 3 of the paper in effect argues for the scrutability thesis in a general form, and sections 4 and 5 defend a specific version of the thesis. The reply to objection 6 in section 5 is particularly important in defending the a priori entailment version of the scrutability thesis. Sections 8 and 9 of Chalmers (2002a) provide a further defense of a version of the thesis.



Here, the base information need not contain terms such as ‘Hesperus’ or ‘renate’, or any cognates. And no further empirical information is required: the information in the base is all that is needed.

Something similar applies for terms like ‘philosopher’, or even names like ‘Gödel’ or ‘Feynman’. Here, the base information D may need more than in the cases above: for example, it may need to include information about people and their mental states, and the use of certain names, and so on. For example, once I know enough about the history of the use of the name ‘Gödel’ by others in my community, about the properties of relevant individuals, and so on, then I will be in a position to know that Gödel was a mathematician, even if I had no substantive knowledge of Gödel beforehand. And again, my information need not use the terms ‘Gödel’ or ‘mathematician’ to do this. It might use the quite different term “‘Gödel’”, in order for me to track down the referent via those from who I obtained the name, but that is legitimate in this context.

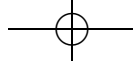
This pattern may not apply to *all* expressions. There are plausibly some primitive terms (perhaps ‘and’, ‘cause’, and ‘conscious’, for example) such that to know whether a sentence involving these terms is true, one needs a base that includes those terms or relevant cognates that invoke them implicitly. But as long as the principle applies reasonably widely, it is good enough.

By the sort of reasoning above, one can infer a slightly stronger claim. Let us say that a vocabulary is a set of terms, and that a V-truth is a truth that uses only terms in V. Then we can say: there is a relatively limited vocabulary V such that for any truth S, there is a V-truth D such that D is epistemically sufficient for S. To arrive at V, intuitively, we might simply eliminate terms one by one from the language according to the scrutability principle laid out above, until we cannot eliminate any further. Exactly how limited V must be is an open question, but I think the sort of reasoning above gives good ground to accept that it will involve only a small fraction of the original language. One can put the claim in a slightly stronger form:

Scrutability of Truth II: There is a relatively limited vocabulary V such that for any truth S, there is a V-truth D such that D implies S.

Here, we have moved from “D is epistemically sufficient for S” to “D implies S”: that is, that the material conditional ‘ $D \supset S$ ’ is a priori. This is a stronger but not a vastly stronger claim, given that epistemic sufficiency involved “no further empirical information”. One can argue for it along much the same lines as above, suggesting that even a speaker who suspends all empirical beliefs can know that *if* D is the case, then S is the case. Chalmers and Jackson (2001) argue in much more depth that this sort of conditional is a priori (for a specific choice of V). A point made there is worth noting here: this sort of a priori entailment does not require that there is an explicit definition of the terms in S using the terms in V.¹³

¹³ For this reason, the epistemic two-dimensional framework set out here does not require or entail that the epistemic intension of an expression be analyzable in terms of some explicit description: for example, it is not required that a name or a natural kind term N be analyzable



(Note that even if one is skeptical about apriority, the general point about epistemic sufficiency is still plausible. Such a skeptic can instead appeal to an alternative notion of epistemic necessitation, such as one understood in terms of rational inference. Corresponding theses about scrutability and nontriviality will remain plausible given such a notion.)

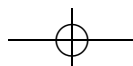
It is also plausible that there is some V-truth D that implies all V-truths. Of course D may need to be an infinitary conjunction, but we may as well stipulate that V is part of our idealized language, so this is no problem. We can think of D as a conjunction of the simple V-truths about the world, or as a conjunction of all V-truths of up to a certain level of complexity. There is plausibly a level such that any more complex V-truth will be implied by this sort of conjunction. If so, it follows that D implies all truths about the world. It follows plausibly that D is epistemically complete (if D is compatible with H and $\sim H$, then all truths about the world are compatible with H and with $\sim H$, which is plausibly impossible.)

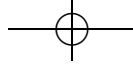
Exactly what is required for the vocabulary V and the description D is an open question. Chalmers and Jackson (2001) and Chalmers (2002a) argue that a specific description D will work here: PQTI, the conjunction of microphysical and phenomenal truths with certain indexical truths and a “that’s-all” truth. If this is right, then V requires only the vocabulary required for PQTI. It is possible that the vocabulary might be stripped down further, if Q is implied by P (as some physicalists hold), or if P is implied by a description in a more limited vocabulary, such as one in terms of space, time, and causal connections (an appropriate Ramsey sentence, for example). But in any case, this specific claim is not required here. The only claim required is that *some* limited vocabulary V suffices for this purpose.

What goes for the actual world goes also for any epistemic possibility. There is nothing special about the actual world here. Given any class of epistemically compatible sentences in our idealized language, one can strip down the vocabulary involved in it in the same sort of way as before, until one has a limited vocabulary V' such that each of the original sentences is implied by a V' -sentence. It follows by similar reasoning to the above that for any scenario W , there will be a limited vocabulary V' such that there is an epistemically complete V-truth that corresponds to the scenario. Of course the vocabulary may differ between scenarios. For example, there are presumably epistemically possible scenarios that involve conceptually basic kinds that are alien to our worlds. If so, the vocabulary required to describe our world must be expanded to describe this scenario. But the resulting vocabulary will still be limited.

as ‘the actual D ’ for some description D . Likewise, an expression’s epistemic intension need not correspond directly to any descriptive belief of the speaker: for example, it is not required that one who uses a term N has a priori “identifying knowledge” to the effect the referent of N is φ , for some property φ . All that is required is that certain conditionals be epistemically necessary.

This bears on criticisms of two-dimensionalism raised by Soames (2004) and by Byrne and Pryor this volume. A number of Soames’s arguments rest on criticizing the thesis that names are analyzable as rigidified descriptions. The central arguments of Byrne and Pryor rest on criticizing the thesis that users of names and natural kind terms have a priori identifying knowledge. The framework I have outlined is not committed to these theses (in fact, I think that the theses are probably false), so the corresponding arguments do nothing to undermine the framework I have outlined.





Let us say that a *basic* vocabulary is a minimal vocabulary V' such that every epistemically possible sentence is implied by some V' -sentence. We can think of such a vocabulary as providing an *epistemic basis*: the terms in it express a set of concepts sufficient to cover all of epistemic space. Given the reasoning above, there is reason to believe that a basic vocabulary will be a relatively limited vocabulary. Exactly how small a basic vocabulary can be is again an open question, but it may well involve only a very small fraction of the terms of the original language. With such a vocabulary in place, we can think of a scenario as corresponding to an equivalence class of epistemically complete V' -sentences, rather than of arbitrary epistemically complete sentences.

(Note that there is no need to appeal to a basic vocabulary for the *definition* of epistemic intensions. The canonical descriptions invoked in the definition are not restricted to a basic vocabulary, although it is easy to see that any such description will be epistemically equivalent to a description in a basic vocabulary.)

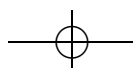
If a reasonably limited basic vocabulary exists, it follows that epistemic intensions are nontrivial. An epistemically complete description need not specify the status of most sentences explicitly. Most terms, such as 'water' and 'H₂O', will plausibly not be required in a basic vocabulary, so sentences involving these terms will be *nontrivially* true or false in scenarios. For all we have said here, it may be that some claims (for example 'there is space') are in a sense trivially true in some scenarios and trivially false in others, but this is only to be expected: it is analogous to the trivial truth or falsity (in an analogous sense) of claims about ontologically fundamental properties in metaphysically possible worlds. So there will be plenty of interesting structure to epistemic intensions in general.

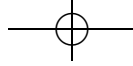
3.7 Subsentential epistemic intensions

So far I have defined epistemic intensions only for sentences. It is not too hard to define them for subsentential expressions, such as singular and general terms, kind terms, and predicates, but there are a few complexities. I will take it that we have already decided on independent grounds what sort of extensions these expressions should have: e.g. individuals, classes, kinds, and properties. Differences choices could be made here, but the same sort of treatment will work.

The details depend to some extent on whether we take the metaphysical or the epistemic approach to scenarios. The difference is that centered worlds already come populated with individuals and the like (or at least we are familiar with how to regard them as so populated), whereas maximal hypotheses do not (or at least we are less familiar with how to populate them).

If we take the metaphysical approach to scenarios: let W be a centered world with canonical description D , and let T be a singular term. In most cases, D will imply a claim of the form ' $T = T^*$ ', where T^* is a semantically neutral singular term. (Here I include definite descriptions as singular terms.) If so, the epistemic intension of T picks out the referent of T^* in W (that is, it picks out the individual that T^* picks out when W is considered as counterfactual). In some symmetrical worlds, it may be that there is no such semantically neutral T^* , but there is a T^* that involves semantically neutral terms plus 'I' and 'now' (plus other basic indexicals, if any). In this case, one





can replace the indexicals in T^* by labels for the entities at the center of the world, yielding an expression T^{**} such that the epistemic intension of T picks out the referent of T^{**} in W . If there is no such T^* , then the epistemic intension of T is null in W .

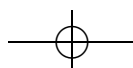
One can do the same for general terms, appealing to claims of the form ‘For all x , x is a T iff x is a T^* ’, and holding that the epistemic intension of T in W picks out the referent of T^* , for a T^* that is semantically neutral (perhaps plus indexicals). For kind terms, we again appeal to identities ‘ T is T^* ’. For predicates, we can appeal to claims of the form ‘For all x , x is T iff x is T^* ’. This delivers extensions for the epistemic intensions straightforwardly.

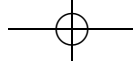
If we take the epistemic view of scenarios, then we need to populate scenarios with individuals and the like. If we simply admit scenarios as a basic sort of abstract object with certain properties, one could simply stipulate that they contain individuals that can serve as the extensions of relevant expressions—much as many of those who introduce possible worlds simply stipulate something similar. But it is useful to go through an explicit construction.

Let W be a scenario with canonical description D . Let us say that two singular terms T_1 and T_2 are equivalent under W if D implies ‘ T_1 is T_2 ’. Then we can identify every equivalence class of singular terms under W with an individual in W , and hold that the epistemic intension of T in W picks out the individual corresponding to T ’s equivalence class in W . As for general terms: every general term G will pick out a class of individuals. One of the individuals defined above will be in G precisely when D implies ‘ T is a G ’, for some T that picks out the individual. One can do something similar for predicates and kind terms: the details will depend on the precise view one takes of properties and kinds and their relation to individuals, so I will not go into them here.

There is one worry: what if the truth of certain existentially quantified claims in a scenario requires individuals that are not the referent of any singular term? For example, there may be a predicate φ such that D implies ‘ $\exists x \varphi(x)$ ’, and D does not imply any claim of the form ‘ $\varphi(T)$ ’, where T is a singular term. Of course since D is epistemically complete, it will tell us exactly how many individuals have φ , whether some individuals with φ also have ψ and some do not, and so on. It is not hard to see that this sort of case will ultimately require predicates φ (perhaps an infinitely conjunctive predicate) such that D implies that there exists more than one individual with ψ , and such that for all predicates ψ , D implies that these individuals are indistinguishable with respect to ψ . In this case, the individuals will be indistinguishable even in our idealized language, presumably because of deep symmetries in the world. In such a case, if D implies that there are n individuals with φ , one can arbitrarily construct n individuals, perhaps as ordered pairs $(\varphi', 1) \dots (\varphi', n)$, where φ' is the equivalence class containing φ , and stipulate that all of these individuals fall under the extension of φ , and of other predicates and general terms as specified by the relevant D -implied universally quantified truths about individuals with φ .

In this way, we can construct the relevant classes of individuals and the like, and specify the extensions of various expressions’ epistemic intensions. The construction ensures that where the extension of a complex expression is a compositional function





of the extensions of its parts, then the same will be true of the extension of a complex expression relative to a scenario. For an identity (e.g. ‘ $T_1 = T_2$ ’), compositionality will be ensured by the equivalence class construction. For a predication (e.g. ‘ T is a G ’, or $\checkmark(T)$), this will be ensured by the appropriate construction of extensions for general terms (as above) or predicates. The machinations two paragraphs above ensure that existential quantification will work straightforwardly, and universal quantification is guaranteed to work (if D implies $\forall x \varphi(x)$, then every individual constructed above will have φ). Logical compositionality is guaranteed at the sentential level (if D implies both S and T , D will imply $S\&T$, and so on). So the epistemic intension of a complex expression will be a compositional function of the epistemic intension of its parts.

Of course once one has engaged in this sort of construction, one need not usually bother with the details again. It is perfectly reasonable thereafter to speak of a scenario as containing individuals and the like, and to speak about terms as picking out various individuals in a scenario, quite independently of the details of the construction. On the epistemic approach to scenarios, for most purposes one can think of them as abstract objects that may behave somewhat differently from possible worlds, but that have the same sort of status in our ontology.

3.8 Tokens and types

As I have approached things, epistemic intensions have been assigned to expression tokens rather than expression types (such as linguistic types). The reason for this is straightforward. It is often the case that two tokens of the same linguistic type can have *different* epistemic intensions. This difference arises from the fact that different speakers may use the same expression so that it applies to epistemic possibilities in different ways. And this difference arises in turn from the fact that different speakers may use the same term with different a priori connections.

For example, it is often the case that two speakers will use the same *name* with different a priori connections. The canonical case is that of Leverrier’s use of ‘Neptune’, which he introduced as a name for (roughly) whatever perturbed the orbit of Uranus. For Leverrier, ‘If Neptune exists, it perturbs the orbit of Uranus’ was a priori. On the other hand, later speakers used the term (and still do) so that this sentence is not a priori for them: it is epistemically possible for me that Neptune does not perturb the orbit of Uranus. We can even imagine that when Leverrier’s wife acquired the name, she did not acquire the association with Uranus, so that she is in no position to know the truth of this sentence a priori.

How can we characterize the epistemic intension of Leverrier’s tokens of ‘Neptune’? To a first approximation, we can say that in any scenario, Neptune picks out whatever perturbs the orbit of Uranus in that scenario. How can we characterize the epistemic intension for Leverrier’s wife? This is a bit trickier, but we can assume that for his wife to determine the reference of ‘Neptune’, she would examine Leverrier’s own use and see what satisfies it. So to a first approximation, his wife’s epistemic intension picks out whatever Leverrier refers to as ‘Neptune’ in a given scenario. One can find a similar (although less stark) variation in the epistemic intensions of many names, and perhaps natural kind terms.

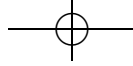
Something similar applies to many uses of context-dependent terms, such as 'heavy'. What I count as heavy varies with different uses of the term. In some contexts, 'My computer is heavy' may be true, and in other contexts it may be false, even though it is the same computer with the same weight. Correspondingly, the way I apply a term across epistemic possibilities will vary with these uses: if I *suppose* that my computer weighs such-and-such, I may hold the utterance true in the first case but not the second.

As we have defined epistemic intensions, they are grounded in the behavior of sentences under an epistemic necessity operator. So the variation in epistemic intensions of two expressions of the same type is traceable to variations in the epistemic necessity of two type-identical sentences. In particular, it will be traceable to variations in the apriority of two type-identical sentences. And this variation is traceable to variations in the apriority of the thoughts that the two sentences express.

Here we need to say a little more about thoughts. A thought is understood here as a token mental state, and in particular as a sort of occurrent propositional attitude: roughly, an entertaining of a content. The idea is that this is the sort of propositional attitude that is generally expressed by utterances of assertive sentences. Such utterances typically express occurrent beliefs, but they do not always express occurrent beliefs, as subjects do not always believe what they say. Even in these cases, however, the subject *entertains* the relevant content: a thought is an entertaining of this sort. Like beliefs, thoughts are assessable for truth. Thoughts can come to be *accepted*, yielding beliefs, and thoughts can come to be *justified*, often yielding knowledge. When an utterance expresses a thought, the truth-values of the utterance and the thought always coincide.

On this way of approaching things, we assume a relation of expression between statements and thoughts, and we assume a notion of epistemic necessity as applied to thoughts. The latter notion might be seen as the true conceptual primitive of the approach. On the account where epistemic necessity is tied to apriority, we can characterize it further by saying: a thought is epistemically necessary when it can be justified independently of experience, yielding a priori knowledge. We can then say that a thought is epistemically possible when its negation is not epistemically necessary. Two thoughts are epistemically *compatible* when their conjunction is epistemically possible. One thought *implies* another when the first is epistemically incompatible with the negation of the second. Here we assume that two thoughts of the same subject can stand in a relation of negation, and that a thought can stand in a relation of conjunction or disjunction to a set of two or more other thoughts of the same subject.

With these notions in hand, we can characterize epistemic necessity and necessitation as applied to sentences. A sentence token is epistemically necessary iff it expresses an epistemically necessary thought. A sentence type is epistemically necessary iff any token of the type (used competently and literally) is epistemically necessary. If D and E are sentence types, we can say that D epistemically necessitates E when $D \& \sim E$ is epistemically impossible. If D is a sentence type and S is a sentence token: let us say that a thought is a D-thought if it is the sort apt to be expressed by D. Then D epistemically necessitates S when a possible D-thought of the subject will imply the thought expressed by S. Equivalently, D epistemically necessitates S when, if the



thought expressed by S were to be disjoined with a \sim D-thought, the resulting thought would be epistemically necessary.

We can also use this framework to directly define epistemic intensions for thoughts as well as utterances. Much as above, we can say that a scenario verifies a thought when disjunction of the thought with a \sim D-thought is epistemically necessary, where D is a canonical description of the scenario. This yields a notion of mental content that can be applied to beliefs, thoughts, and any propositional attitude with a mind-to-world direction of fit (see Chalmers 2002c). Using the definitions above, we can see that when an utterance expresses a thought, the epistemic intension of the utterance will be identical to the epistemic intension of the thought.

This framework enables us to see how two tokens of the same type can differ in apriority. When Leverrier says ‘If Neptune exists, it perturbs the orbit of Uranus’, his statement presumably expresses a priori knowledge, and certainly expresses a thought that can be justified a priori. If his wife utters the same sentence, no amount of a priori rational reflection alone could justify the thought she expresses. Similarly, it is possible that two names for a single individual (‘Bill Smith’ and ‘William Smith’) are used completely interchangeably by one person, so that an utterance of an identity statement involving the names expresses a trivial thought. Such an utterance will then be a priori. But there clearly may be others for whom a corresponding utterance expresses a nontrivial thought and is a posteriori.¹⁴ Finally, if I say ‘someone with 1,000 hairs on their head is bald’ on one occasion, it may express an a priori false thought (one whose negation is a priori justifiable), while if I say it on another occasion, it may express a thought that is not a priori false, and may be plausibly true.

In a similar way, this framework enables us to see how two tokens of the same type can have different epistemic intensions. Let S be the sentence ‘Neptune is an asteroid’, and let D be a canonical description of a scenario W in which the orbit of Uranus is perturbed by an asteroid and in which no-one has ever used the term ‘Neptune’. (We can abstract away from complications involving the intension of ‘Uranus’ and ‘asteroid’.) Then D epistemically necessitates Leverrier’s utterance of S: a thought that D obtains would imply the thought Leverrier expresses with S. But D does not epistemically necessitate Leverrier’s wife’s utterance of S: a thought that D obtains would not imply the thought that his wife expresses with S. (Note that D itself will not exhibit this sort of variation, as expressions in the idealized language are required to be epistemically invariant.) So Leverrier’s utterance of S is verified by W, while his wife’s utterance is not. So the two utterances have different epistemic intensions.

One might reasonably ask: in languages such as English, what sorts of simple terms have epistemic intensions that vary between speakers and occasions of use? This happens most clearly for: (i) proper names (such as ‘Neptune’ and ‘Gödel’); (ii) ordinary natural kind terms (such as ‘water’ and ‘gold’); (iii) demonstratives (such as ‘that’ and ‘there’); and (iv) many context-dependent terms (such as ‘heavy’ and ‘bald’). For terms like this, it is clear that an epistemic intension is not part of a term’s “standing meaning”, where this is understood as the sort of meaning that

¹⁴ This sort of case is discussed in Chalmers (2002b), section 9.

is common to all tokens of a type in a language. Instead, it is a sort of “utterance meaning” or “utterance content”. Some theorists use the term “meaning” only for standing meaning, but this is a terminological matter. The substantive point is that the framework yields a useful and interesting sort of semantic value in the broad sense, one that can be associated with utterances and that can play a useful explanatory role. (There is more discussion of this matter in Chalmers 2002b, section 8.)

Are there terms for which an epistemic intension is common to all tokens of a type? This is perhaps most plausible for certain indexicals, such as ‘I’ and ‘today’ (at least setting aside unusual uses, and any context-dependence at the boundaries). It may also hold for some descriptive terms, such as ‘circle’. Most of these have some context-dependence, but this can be regimented out more straightforwardly than the epistemic variability of names and natural kind terms. Finally, it may hold for some descriptive names (e.g. ‘Jack the Ripper’), at least for a certain period of their existence. For terms like these, an epistemic intension might be seen as part of their standing meaning.

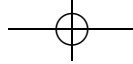
3.9 Apriority

On the main approach advocated here, epistemic necessity is regarded as a sort of apriority. This requires us to say a bit more about the notion of apriority. There are various ways in which apriority can be understood, but current purposes require a fairly specific understanding of it. A characterization of the relevant notion of apriority might run something like this. A sentence token is a priori when it expresses an a priori thought. A thought is a priori when it can be conclusively non-experientially justified on ideal rational reflection.

There are five distinctive features of this conception of apriority that deserve comment. The first feature is that the relevant sort of apriority is *token-relative*. The second is that apriority is *mode-of-presentation-sensitive*. The third is that apriority is *idealized*. The fourth is that apriority is *non-introspective*. The fifth is that apriority is *conclusive*. The first feature was been discussed in the previous section. Here I will say a little about the other four.

Mode-of-presentation sensitivity: Intuitively, sentences such as ‘Hesperus is Phosphorus’ are a posteriori. But some theorists (e.g. Salmon 1986; Soames 2002) hold that such sentences are a priori, on the grounds that they express trivial singular propositions that can be known a priori (e.g. by knowing that Venus is Venus). On the current definition of apriority, tokens of such a sentence are not a priori. The thought expressed by an utterance of ‘Hesperus is Phosphorus’ clearly cannot be justified independently of experience. At best, a different thought associated with the same singular proposition can be so justified. So on the current definition, the utterance is not a priori. On this approach, as on the intuitive understanding, apriority is sensitive to mode of presentation. The apriority of an utterance is grounded in the epistemic properties of a corresponding thought, which are tied to the inferential role of that thought in cognition.¹⁵

¹⁵ Note that to say that a token of a sentence S produced by speaker A is a priori is *not* to say that a knowledge ascription of the form ‘A knows a priori that S’ (or ‘A can know a priori that



Idealization: Here the notion of apriority is understood so that it idealizes away from a speaker's contingent cognitive limitations. A sentence token (of a complex mathematical sentence, for example) may be a priori even if the speaker's actual cognitive capacities are too limited to justify the corresponding thought a priori. What matters is that the thought could be justified a priori on idealized rational reflection.¹⁶

Non-introspectiveness: On some conceptions of apriority, introspective knowledge (for example my knowledge that I am thinking, or my knowledge that I believe I am Australian) qualifies as a priori. On the current conception, introspective knowledge does not qualify as a priori. We can stipulate that experiential justification should be understood in such a way to include both perceptual and introspective justification. It follows that in excluding experiential justification, apriority rules out both perceptual and introspective justification.

Conclusiveness: On the current conception, a priori justification must meet the sort of conclusive standard associated with proof and analysis, rather than the weaker standard associated with induction and abduction. On this conception, an inductive generalization from instances each of which is known a priori does not possess the relevant sort of a priori justification (even though it might be held to be a priori in some reasonable sense). Likewise, neither does an abductive conditional from total evidence to a conclusion that is grounded in and goes beyond the evidence. Intuitively, in such cases one may have non-experiential justification for believing a conclusion, but one is unable to conclusively rule out the possibility that the conclusion is false. As understood here, apriority is tied to the sort of justification that conclusively rules out the possibility that the relevant sentence is false.¹⁷

Of the five features above, the first two are necessary in order to capture the close tie between apriority and rational significance: it seems clear that rational significance is token-relative and mode-of-presentation sensitive. The last three are necessary in order to capture the idea that apriority should correspond to a sort of epistemic

S') is true. (Clearly a token of 'If I exist and am located, I am here' may be a priori for a speaker even if that speaker cannot know a priori that if I exist and am located, I am here. The criteria may also come apart in cases where ascriber and ascribee associate different modes of presentation with the expressions in S.) The current construal of apriority requires no commitment on the semantics of attitude ascriptions: what I have said here about the non-apriority of 'Hesperus is Phosphorus' is even consistent with Salmon and Soames' counterintuitive Millian semantics for attitude ascriptions, on which 'A knows a priori that Hesperus is Phosphorus' is true.

¹⁶ This characterization of the idealized a priori should be seen as an intuitive characterization of a notion that is being taken as primitive. It is best not to define idealized a priority in terms of possible justification, both because of the proliferation of primitive notions, and because it could lead to problems on views on which certain conceivable cognitive capacities are not metaphysically possible. For example, if it turns out that there are strong necessities entailing that no being can construct a proof with more than a million steps, then a statement whose proof requires more steps than this will not satisfy the putative definition, but it will still count as epistemically necessary in the idealized sense I am invoking here.

¹⁷ Again, this intuitive characterization should not be understood as an analysis of conclusive justification. It merely points to an intuitive distinction. A more detailed characterization might analyze conclusive justification of a belief in terms of the nonexistence of certain sorts of skeptical hypotheses under which the belief would be false.

necessity. We want epistemic necessity to capture the intuitive idea that some thoughts are true under all coherent hypotheses about the actual world. Inductive knowledge and introspective knowledge do not have this property (intuitively, they are false in some scenarios), while idealized mathematical knowledge does have this property. So our conception of apriority should exclude the first two and include the third.

This conception of apriority should be understood as stipulative. One can define ‘a priori’ in different ways, so that it is type-relative, or so that it is not sensitive to modes of presentation, or that it is not idealized, or so that it allows introspective knowledge or inductive knowledge. There is no need to adjudicate the terminological question of which of these conceptions is the “correct” one. In fact, nothing rests on the use of the term “a priori”: one could simply use the term “epistemically necessary” for the stipulated notion throughout.

3.10 The second dimension

I have concentrated almost wholly on the first dimension of the two-dimensional framework. This is because the second dimension is already well-understood. But I will say a few words about it here. It is worth examining how it can be understood in a way that is parallel to the way we have understood the first dimension.

Like the first dimension, the second dimension is founded on a certain sort of possibility and necessity. For the first dimension, this is epistemic possibility and necessity, tied to what *might be* the case. For the second dimension, this is what we might call *subjunctive* possibility and necessity, tied to what *might have been* the case.

We can say that S is subjunctively possible when it might have been the case that S (more strictly, when an utterance of ‘it might have been the case that S’ by the speaker, with the modal operator adjusted for the relevant language, would be true). Kripke is explicit that this is the basic notion of possibility and necessity with which he is working, and almost all of his modal arguments are directly grounded in intuitions about what might have been the case.

With this basic modal operator in hand, we can proceed as before. For example, one can define a *subjunctively complete* sentence parallel to the way we defined an epistemically complete sentence. One can construct equivalence classes of subjunctively complete sentences in an idealized language. One can identify these classes as *maximal metaphysical possibilities*, or as *possible worlds*. One can give possible worlds *canonical descriptions*, which will be subjunctively complete sentences in their equivalence class.

Just as one can consider a scenario as actual, by supposing that it actually obtains, one can consider a world as counterfactual, by supposing that it had obtained. That is, instead of thinking “if D is the case, then . . .”, one thinks “if D had been the case, then . . .” (where D is a canonical description of a world W). For example, for a given sentence S, one can entertain and evaluate the subjunctive conditional: “if D had been the case, would S have been the case?”. In some cases, the answer will intuitively be yes: in this case, we can say that W *satisfies* S. This is a distinctive sort of *counterfactual* evaluation. When W satisfies S, we can say the *subjunctive intension* of S is true at W.

For example, if we accept Kripke's intuitions, then we will say: if the bright object visible in the evening had been Mars, then it would not have been the case that Hesperus was Mars (Hesperus would still have been Venus). In this way, our subjunctive intuitions are quite different from our epistemic intuitions. Likewise, if we accept Putnam's intuitions, then we will say: if the clear liquid in the oceans and lakes had been XYZ, then it would not have been the case that water was XYZ (water would still have been H₂O). If we accept these intuitions, we will say that the subjunctive intension of 'Hesperus is Venus' is true at all worlds (or at all worlds where Venus exists), as is the subjunctive intension of 'Water is H₂O'. These intensions differ markedly from the epistemic intensions of 'Hesperus is Venus' and 'Water is H₂O', both of which are plausible false at many scenarios.

The subjunctive intension of a sentence *S* is a function from worlds to truth-values, true at *W* if and only if *W* satisfies *S*. Satisfaction can be intuitively characterized as above. Formally, we can say that *W* satisfies *S* when *D* subjunctively necessitates *S*, where *D* is a canonical description of *W*. We could define subjunctive necessitation by the subjunctive conditional heuristic above. Or more formally, one might say that *D* subjunctively necessitates *S* when *D* & ~*S* is subjunctively impossible.

With a possible world as constructed, we can construct a space of individuals much as we did with scenarios. We can then define subjunctive intensions for subsentential expressions straightforwardly. Subjunctive intensions are defined in the first instance for expression tokens, since subjunctive necessity judgments can vary between tokens of a type. For some expression types, all tokens of the type will have the same subjunctive intension: this is arguably so for names and natural kind terms (for example 'Hesperus' and 'water'), logical and mathematical terms, and some descriptive terms (for example 'circle'). For other expression types, subjunctive intensions will vary between tokens of the type: this is so for indexicals (for example 'I') and many context-dependent terms (for example 'heavy'). In the first case, subjunctive intension may be an aspect of linguistic meaning; in the second case, it is not.

The basic ideas here are parallel between the two cases. The explicit construction of possible worlds and the like may seem like unnecessarily heavy weather; but this seems so only because possible worlds are more familiar. Perhaps one does not really need any such construction to legitimize the appeal to possible worlds; but if so, the same applies to scenarios. In both cases, one takes a modal notion as basic, and invokes a corresponding modal space as a tool of analysis.

There is one important difference between worlds and scenarios. We have a means of reidentifying individuals across worlds, but in general there is no such means of reidentifying individuals across scenarios. In the case of worlds, these claims are grounded in *de re* subjunctive intuitions of the form 'x might have been F'—read so that they are distinct in their form from *de dicto* subjunctive intuitions such as 'it might have been that T was F' where T denotes x. We can use these claims in conjunction with the construction above to identify certain objects in alternative possible worlds as identical to certain objects in the actual world (or alternatively, to identify objects with equivalence classes across worlds; or at least to set up counterpart relations across worlds). There is no clear analog of a *de re* modal intuition in the epistemic case: 'Hesperus is the evening star' may be *a priori*, but it is not clear

what it means to say that Hesperus (i.e. Venus) is such that it is a priori that it is the evening star.

In the subjunctive case, one can also ground the reidentification of individuals across worlds in de dicto subjunctive intuitions involving a privileged class of designators, that is, names. Judgments of the form ‘it might have been that N was F’ where N is a name for the relevant object, arguably give the same result for any name of the object, and if so can ground a sort of crossworld identification. In the epistemic case, there is in general no analog to this privileged class of designators: different names for an individual are not generally a priori equivalent, so come apart in different scenarios, and there is no way in general to isolate a privileged class of epistemically equivalent designators here. At best this may be possible in special cases, such as canonical designators for phenomenal states and abstract entities. A consequence is that quantified modal claims will not generally be well-defined in the epistemic case, and quantified modal logic will be largely inapplicable in this domain.

In many cases, a term’s subjunctive intension will depend on its actual extension, or on other aspects of the actual world. This is particularly clear in the case of rigid designators such as names and indexicals. If Kripke is correct, these pick out the same individual in all possible worlds, and so pick out the term’s actual extension in all possible worlds (for example, the subjunctive intension of ‘Hesperus’ will pick out Venus in all worlds). In these cases, the subjunctive intension of a term itself depends on the character of the actual world. Here, in effect, a term’s subjunctive intension depends on which epistemic possibility turns out to be actual.

One can naturally encapsulate this behavior in a *two-dimensional intension*. This can be seen as a mapping from scenarios to subjunctive intensions, or equivalently as a mapping from (scenario, world) pairs to extensions. We can say: the two-dimensional intension of a statement S is true at (V, W) if V verifies the claim that W satisfies S. If D_1 and D_2 are canonical descriptions of V and W, we say that the two-dimensional intension is true at (V, W) if D_1 epistemically necessitates that D_2 subjunctively necessitates S. A good heuristic here is to ask “If D_1 is the case, then if D_2 had been the case, would S have been the case?”. Formally, we can say that the two-dimensional intension is true at iff (V, W) ‘ $\Box_1(D_1 \supset \Box_2(D_2 \supset S))$ ’ is true, where ‘ \Box_1 ’ and ‘ \Box_2 ’ express epistemic and subjunctive necessity respectively. One can define two-dimensional intensions for subsentential expressions by an extension of this idea.

One complication: the construction so far makes the space of possible world derive from subjunctive modal claims. The truth of some subjunctive modal claims depends on the character of the actual world, so one might think that the space of possible worlds will do so as well (that is, the space of W’s may depend on V). Whether this will be so depends on further substantive philosophical issues.

If (i) every possible world can be completely specified by semantically neutral terms (and if this is a priori), then one can require that canonical descriptions be given in these terms, and can use these descriptions to identify worlds across the spaces corresponding to each scenario. If also (ii) the truth of subjunctive modal claims in semantically neutral language is a priori, so that it does not depend on which scenario is actual, then one can identify the spaces themselves. If (i) holds but not (ii), then while we can identify worlds across spaces, some of these spaces will differ in their

extent. (Note that if Metaphysical Plenitude is a priori, as I hold, then this option is excluded.) If (i) does not hold (for example because there are pure haecceitistic differences between worlds, or because there are fundamental intrinsic properties that cannot be specified in semantically neutral terms), then it is probably best to see each scenario as being associated with a relativized space of possible worlds (putative worlds, in the case of non-actual scenarios). In this case, canonical descriptions of worlds on the second dimension will sometimes use non-neutral language, and the worlds will not always be identifiable across spaces.

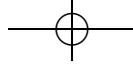
In effect, this two-dimensional structure will represent the space of epistemic possibilities concerning the space of metaphysical possibilities. If (i) and (ii) hold, the extent and nature of the space of metaphysical possibilities will be determined a priori, so that we will have the same space of worlds corresponding to every scenario. If (i) or (ii) fail to hold, then the space of metaphysical possibilities will depend to some extent on which epistemic possibility turns out to be actual, and we may have different spaces of (putative) worlds corresponding to different scenarios.

Note that this worry does not affect the earlier use of semantically neutral descriptions of centered worlds for epistemic purposes. The cases where semantically neutral resources do not fully describe a world will generally correspond to cases where two centered worlds have the same canonical description for epistemic purposes, so that they correspond to a single maximal hypothesis (and to a single scenario, on the epistemic construction). The case of two haecceitistically different but qualitatively identical worlds illustrates this: the haecceitistic differences are irrelevant for epistemic purposes.¹⁸

For every scenario, one world (in the scenario's space of worlds) will be the world *associated* with the scenario. Intuitively, this is the world that will be actual if the scenario obtains. If scenarios are centered worlds, a scenario's associated world will be the scenario stripped of its center. On the epistemic view of scenarios, we can say that (to a first approximation) W will be associated with V when canonical descriptions of V and W are epistemically compatible. (Note that this definition allows that in principle more than one world could be associated with a scenario, if scenarios are relevantly less fine-grained than worlds.¹⁹)

¹⁸ This bears on the issue about intrinsic properties raised by Schroeter (2004). I would like to think that there is a semantically neutral conception of fundamental intrinsic properties, but the framework is not committed to this. If there is no such conception, then one will have to use non-neutral language to fully characterize worlds on the second dimension. The first dimension will be unaffected, however: at worst, a single maximal hypothesis will correspond to an equivalence class of centered worlds (see footnote 9). At most, what is affected is the alignment between epistemic space and subjunctive space: epistemic space will be in this respect smaller than subjunctive space. There will still be a reasonably robust link between epistemic and metaphysical possibility: the resulting position will be what Chalmers (2002a) calls "strong modal rationalism" without "pure modal rationalism" (on this position, conceivability entails possibility, but possibility does not entail conceivability, due to the existence of "open inconceivabilities"). This is an instance of the general point that semantic neutrality is relevant to the alignment between the epistemic and the subjunctive, but is inessential to purely epistemic issues.

¹⁹ For example, if intrinsic properties operate as in the previous note, then two worlds with different but isomorphic distributions of intrinsic qualities may be associated with the same scenario.



Given the association relation between scenarios and worlds, one can define the *diagonal intension* of a sentence's two-dimensional intension. This will be a mapping from scenarios to truth-values, mapping V to the value of the two-dimensional intension at (V, W) , where W is associated with V . (If there is more than one such W for the reasons above, it is not hard to see that they will all give the same results.) The diagonal intension of a sentence will straightforwardly be equivalent to its epistemic intension. One can therefore reconstruct an expression's epistemic intension from its two-dimensional intension by diagonalizing, just as one can reconstruct its subjunctive intension by holding fixed the actualized scenario.

It should be clear, however, that this diagonal construction in no sense gives the *definition* of an epistemic intension. Epistemic intensions are defined in purely epistemic terms: they are in no sense derivative on subjunctive notions. The diagonal construction is conceptually much more complex, involving subjunctive evaluation, association of worlds with scenarios. In effect, the relation is akin to that between the functions $f(x) = x^3$, $g(x, y) = x^3 + \sin(x - y)$, and $g'(x) = g(x, x)$. Here, g' is the "diagonal" of g , and is the same function as f . But it would obviously be incorrect to hold that f is fundamentally the diagonal of g , or that it is derivative on trigonometric notions. For exactly the same reasons, it is incorrect to hold that an epistemic intension is fundamentally a diagonal intension, or that it is derivative on subjunctive notions.

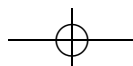
There is a sense in which the two-dimensional intension represents the full modal structure of an expression, capturing how it behaves under epistemic evaluation, modal evaluation, and combinations of the two. Just as an epistemic intension can be evaluated a priori, a two-dimensional intension can be evaluated a priori. A subjunctive intension cannot be evaluated a priori, but it can be evaluated when the actualized scenario is specified.

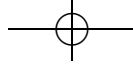
We can think of all of these intensions as aspects of the content of a sentence token. A sentence is in no sense ambiguous for having both epistemic intensions and subjunctive intensions; rather, it has a complex semantic value. Different aspects of this semantic value will be relevant to the evaluation of the sentence in different contexts. In certain epistemic contexts ('it is a priori that S '; 'it might turn out that S '; 'if S is the case, then T is the case'), the epistemic intension of S may play a key role in determining the truth-value of the complex sentence. In subjunctive contexts ('it might have been that S '; 'if it had been that S , it would have been that T '), the subjunctive intension of S may play the most important role. In combined epistemic-subjunctive contexts, truth-value may depend on the two-dimensional intension of S . As usual, there is no need to settle the question of which of these, if any, is *the* meaning or content of an expression.

3.11 The Core Thesis

Let me summarize where things stand with respect to the Core Thesis: that S is a priori iff S has a necessary 1-intension.

First, if S is a priori: then for all W with canonical description D , D implies S . (If S is a priori, D implies S for *any* D .) So S is verified by all W , and has a necessary 1-intension.





Second, if S is not a priori and we take the epistemic approach to scenarios: then $\sim S$ is epistemically possible. Under small assumptions (see 3.4.2), it follows that there is an epistemically complete D such that D implies $\sim S$. Any epistemically complete sentence describes a scenario, so there is a scenario W that verifies $\sim S$. So S does not have a necessary 1-intension.

Third, if S is not a priori and we take the metaphysical approach to scenarios: then if Metaphysical Plenitude is true, any epistemically complete D describes a scenario (a centered world), so S does not have a necessary 1-intension. If Metaphysical Plenitude is false, this does not follow: some epistemically possible statements will not be verified by any centered world, so the Core Thesis will be false.

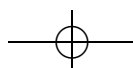
It follows that the Core Thesis is true on the epistemic approach to scenarios, and that it is true on the metaphysical approach iff Metaphysical Plenitude is true. I think there is good reason to hold that Metaphysical Plenitude is true; but even if it is not, we may simply adopt the epistemic approach to scenarios. Either way, the epistemic understanding of two-dimensional semantics plausibly yields an understanding of 1-intensions that satisfies the Core Thesis.

In effect, the epistemic understanding of two-dimensional semantics reconstructs the golden triangle by taking certain epistemic notions as basic and defining certain semantic notions in terms of them, with the aid of modal notions. On the epistemic approach to scenarios, the order of explication is as follows: we take an epistemic notion (such as apriority) as basic, use this to define a modal space (the space of epistemic possibilities), and use this to define corresponding semantic entities (epistemic intensions). On the metaphysical approach to scenarios, we take both an epistemic notion (such as apriority) and a modal notion (metaphysical possibility) as basic, and combine the two to define corresponding semantic entities. On the former approach, the strong connection to the epistemic domain is more or less guaranteed by the construction. On the latter approach, it is grounded in the construction along with the thesis of Metaphysical Plenitude, which articulates a connection between the epistemic and modal domains.

On this approach, the connection between meaning and reason is built in to a large extent by definition. This suggests that we should not make the claim embodied in the golden triangle too strong: a semantic pluralist should accept that there are many other aspects of meaning that are not connected in this way to the epistemic domain. But at the same time, it does not render the analysis trivial. Sense is definitionally connected to cognitive significance, and (subjunctive) modal intensions are definitionally connected to metaphysical possibility, but each of these semantic notions has a powerful role to play. Their cash value is grounded in the phenomena that they help us to analyze. Likewise, in the case of epistemic intensions, the fact that there is a semantic value that bears these connections to the epistemic and modal domains allows us to use semantic and modal tools to play an important role in analyzing the epistemic properties of language and thought.

3.12 Applications

This role for epistemic intensions can be brought out in a number of applications that I will simply summarize here.



- (i) *Fregean sense* (see Chalmers 2002b): Because they satisfy the Core Thesis, epistemic intensions also satisfy the Neo-Fregean Thesis: ‘A’ and ‘B’ have the same intension iff ‘A \equiv B’ is a priori. So epistemic intensions behave broadly like a sort of Fregean sense, tied to the rational notion of apriority. There are some differences. First, sentence-level Fregean senses are supposed to be true or false absolutely, but sentence-level epistemic intensions are true or false relative to a speaker and time (witness ‘I am hungry now’). Second, apriority is weaker than cognitive insignificance, so epistemic intensions are less fine-grained than Fregean senses. (One might adapt the current framework to yield a more fine-grained sort of epistemic intension, by starting with a less idealized notion of epistemic possibility; see Chalmers (forthcoming).) Nonetheless, epistemic intensions can serve as a broadly Fregean semantic value.
- (ii) *Narrow content* (see Chalmers 2002c): One can extend the current framework from language to thought in an obvious way. One can define epistemic intensions for beliefs and thoughts in the manner suggested in 3.8. The result can be seen as a sort of content of thought. It is very plausible that what results is a sort of *narrow* content, such that two physical and phenomenal duplicates will have thoughts with the same epistemic intension. (This narrowness is grounded in the narrowness of deep epistemic possibility: if a thought is epistemically necessary, then the corresponding thought of a physical and phenomenal duplicate will also be epistemically necessary.) This sort of content is much more closely tied to cognition and reasoning than “wide content”, and is well-suited to play a central role in explaining behavior.
- (iii) *Modes of presentation* (see Chalmers 2002c, section 8): In analyzing the behavior of belief ascription, it is common to appeal to a notion of “mode of presentation”, but there is little agreement on what sort of thing a mode of presentation is. Schiffer (1990) suggests that a mode of presentation must satisfy “Frege’s constraint”: roughly, that one cannot rationally believe and disbelieve something under the same mode of presentation. Because they satisfy the neo-Fregean thesis, epistemic intensions satisfy Frege’s constraint perfectly, at least if one invokes an idealized notion of rationality that builds in arbitrary a priori reasoning. So it is natural to suggest that modes of presentation are epistemic intensions. In this way, one can use epistemic intensions to analyze ascriptions of belief.

The current framework is compatible with a number of different proposals that give modes of presentation a role in belief ascription. A naive first account might suggest that ‘X believes that S’ is true if the subject specified has a belief whose epistemic intension is the epistemic intension of S (for the ascriber), but numerous counterexamples to this claim immediately present themselves.²⁰ A more plausible account follows the general shape of so-called “hidden-indexical”

²⁰ Soames (2004) attributes this naive account of belief ascriptions to “strong two-dimensionalism”, and criticizes the resulting view. These criticisms have no force against the view of belief ascriptions laid out in Chalmers (1995) and Chalmers (2002c).

accounts (Schiffer 1990). On such an account, at a first approximation, ‘X believes that S’ will be true if the subject specified has a belief whose subjunctive intension is that of S, and which has an appropriate epistemic intension, where the range of appropriate epistemic intensions may be contextually determined.

- (iv) *Indicative conditionals* (see Chalmers 1998 and Weatherson 2001): One can use epistemic intensions to give a semantics for indicative conditionals that parallels in certain respects the common possible-worlds semantics for subjunctive conditionals. As a first approximation, one can suggest that an indicative conditional ‘If S, then T’ uttered by a subject is correct if the epistemically closest scenario that verifies S also verifies T, where epistemic closeness will be defined in terms of the beliefs or knowledge of the subject. (Weatherson (2001) pursues a closely related idea.)
- (v) *Conceivability and possibility* (see Chalmers 2002a): The Core Thesis makes possible a certain sort of move from conceivability to possibility. If we say that S is conceivable when its negation is not a priori, then when S is conceivable, there will be a scenario verifying S. If we understand scenarios as possible worlds and if Metaphysical Plenitude is true, then when S is conceivable, there will be a centered possible world verifying S. This makes it possible to move from epistemic premises to modal conclusions, as is often done. Of course it is possible to embrace the current framework while rejecting Metaphysical Plenitude and so rejecting the relevant move from conceivability to possibility. But the current framework at least shows how a certain sort of link between conceivability and possibility is tenable in light of the Kripkean phenomena that are often thought to be the greatest threat to such a connection.

4. Epistemic Intensions and Contextual Intensions

We have seen that there are two very different ways of understanding two-dimensional semantics: the epistemic understanding and the contextual understanding. On the epistemic understanding, 1-intensions are constitutively tied to the epistemic domain and satisfy the Core Thesis. On the contextual understanding, 1-intensions are not constitutively tied to the epistemic domain and do not satisfy the Core Thesis.

It is useful to examine the relationship between the two in somewhat more depth. First I will examine how the epistemic understanding deals with the problems that arise for the contextual understanding. Then I will examine to what grounds the resemblance of certain contextual intensions to epistemic intensions.

4.1 Problem cases

The first main problem area for contextual intensions involved sentences such as ‘A sentence token exists’, which are a posteriori, but have a necessary contextual intension. These problems arose because contextual intensions require a token of the evaluated expression in the evaluated world. There is no such requirement for epistemic intensions, so the problem does not arise.

For example, there will be many language-free scenarios: there are many centered worlds with no sentence tokens, and there are many epistemically possible hypotheses according to which there are no sentence tokens. If D is a canonical description of such a scenario, D will verify ‘There are no sentence tokens’. Intuitively, if we consider such a scenario W as actual, we can say that *if* W is actual, then there are no sentence tokens. So the epistemic intension of ‘A sentence token exists’ will be contingent, as required.

The same goes for ‘words exist’, and something similar applies to ‘I am uttering now’. In the latter case, there will be many centered worlds in which the subject at the center is not uttering, and there will be many epistemically possible hypotheses (for me) under which I am not uttering. If D is a canonical description of such a scenario, D will verify ‘I am not uttering now’. So this expression will also have a contingent epistemic intension. The same applies even to ‘I am thinking now’.

‘I exist’ is a slightly trickier case. If ‘I exist’ is a priori, there is no problem. If ‘I exist’ is a posteriori (as I think is the case), then there will be various epistemically possible hypotheses for me under which I do not exist: for example, a hypothesis under which nothing exists (which is arguably itself not ruled out a priori). So on the epistemic view, there will be corresponding scenarios that verify ‘I do not exist’, and ‘I exist’ will have a contingent epistemic intension, as required.

On the world-based view, there is a worry: one might think that any centered world will verify ‘I exist’, since there is always a subject at the center. This raises a subtlety. In the general case, centering is *optional*: on the world-based view, the space of scenarios contains worlds without a marked subject and time, and perhaps worlds with only a marked subject or only a marked time. A world without a marked subject will then verify ‘I do not exist’. The exact choices here will depend on exactly which indexical claims one holds to be a priori, but it should be possible to arrange things so that there is a verifying centered world for every epistemically possible claim.

In any case, we see that problems that arise due to the required presence of a token do not arise here. At most there are problems due to the required presence of a *subject* in a centered world; but these will not arise on the epistemic view, and can be dealt with reasonably straightforwardly on the world-based view. So the epistemic understanding does not suffer from the problems of the contextual understanding here.

Another problem, at least for orthographic contextual intensions, concerned worlds where the subject at the center uses ‘bachelor’ to mean something different, such as horse, so that the 1-intension picks out horse there, which is not the desired result. Again, this problem will not arise for epistemic intensions. In general, to evaluate the epistemic intension of ‘bachelor’ at a scenario, the presence or absence of tokens of ‘bachelor’ in that scenario will be irrelevant (with one qualification to be outlined shortly). What the epistemic intension of ‘bachelor’ picks out in a Steel Earth scenario will depend on a number of other factors, especially the appearance and behavior of substances located around the center of the scenario, but there is no danger that it will pick out steel.

Note that this analysis requires that “My term ‘bachelor’ means bachelor” and similar claims are not a priori. If such a claim was a priori, then because a canonical description of the Steel Earth scenario will contain something like “My term

‘bachelor’ means horse”, the scenario would verify ‘bachelors are horses’, which is the wrong result. But it is independently plausible that these claims are not a priori—at least *if* “‘bachelor’” is understood in purely orthographic terms. It is a posteriori that the string has any meaning at all, and it is a posteriori that it means what it does. If “‘bachelor’” is understood in partly semantic terms, so that it is constitutively tied to a given meaning for ‘bachelor’, then the claim in question may be a priori; but this is not a problem, since in this sense the Steel Earth scenario will not verify “My term ‘bachelor’ refers to horse”. For more on this matter, see Chalmers (2002a) and Yablo (2002).

Note also that there is one case where evaluating an expression’s epistemic intension may turn on the presence of tokens of that expression in a world: expressions used *deferentially*. It may be that Leverrier’s wife uses ‘Neptune’ to (rigidly) pick out whatever her husband refers to as ‘Neptune’. If so, then in a given scenario, the epistemic intension of ‘Neptune’ will pick out roughly the referent of her husband’s term ‘Neptune’ in that scenario (abstracting away from issues about the epistemic intension of ‘my husband’, etc.). In this case, something like “If Neptune exists, my husband refers to it as ‘Neptune’” will be a priori for her. Something similar to this will apply to other terms used deferentially, such as a non-expert’s use of ‘arthritis’, although the details may be less clean. But here, we get only the results that would be expected. For example, if I use ‘water’ wholly deferentially, then if I consider as actual a Steel Earth scenario where those around me use ‘water’ for steel, then this scenario verifies ‘Water is steel’ for me. This seems correct: for a deferential user, although perhaps not for a nondeferential user, ‘Water is steel’ expresses an epistemically possible thought.

(Note that even in deferential cases, evaluation turns on the referent of *others’* use of the expression. It may be that evaluation could also turn on one’s own past use of the expression; but it cannot happen that evaluation will turn on the referent of one’s own current use of the expression, since such a circular criterion cannot secure a referent. (I set aside pathological nonreferring cases, such as ‘the referent of this expression’.) So even in a strongly deferential case, the epistemic intension of ‘water’ will not turn on the referent of a use of ‘water’ at the very center of a scenario.)

There is also the Twin Earth case, where Twin Oscar uses ‘water’ to refer to XYZ. This was a problem for linguistic and semantic contextual intensions, since these are arguably not defined at such a world, whereas we would like the 1-intension of Oscar’s term to return XYZ at this world. Again, the epistemic framework handles this unproblematically. The epistemic intension of ‘water’ returns XYZ at this world, *not* because Twin Oscar’s term ‘water’ refers to XYZ (Twin Oscar’s term is irrelevant), but because the scenario verifies the claim that XYZ has a certain appearance, behavior, relation to oneself, and so on, which in turn verifies ‘Water is XYZ’.

Finally, there was the problem of Fregean typing. It seemed that in order for contextual intensions to give roughly Fregean results, then one had to classify expression tokens under some sort of Fregean type. For a semantic contextual intension to give the right results, for example, one needed to appeal to some sort of prior Fregean semantic notion, which is unhelpful in the current context. No such problem applies

to epistemic intensions. Because these intensions do not rely on tokens of the same type being present within scenarios, there is no need to isolate the common type under which these tokens fall. All one needs is the expression token itself, and its epistemic properties. This approach may *ground* an account of a sort of Fregean semantic value, but it need not presuppose any such account.

These advantages of the epistemic account over the contextual account are all grounded in the fact that the contextual understanding is an essentially metalinguistic understanding, while the epistemic understanding is not. The contextual understanding concerns content that an expression might have had; but the epistemic understanding reveals aspects of the content that it has. Everything is grounded in certain first-order epistemic claims, which we use as tools to reveal an expression's content, just as in the familiar modal case, various first-order subjunctive claims are used as tools to reveal an expression's content. As before, the cases are parallel.

4.2 Semantic contextual intensions

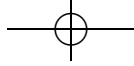
We saw earlier that some versions of a semantic contextual intension presupposed a quasi-Fregean notion of content. We can now turn the picture the other way around, and using the quasi-Fregean notion of content developed here to ground a semantic contextual intension.

Let us say that an *epistemic contextual intension* of an expression is the semantic contextual intension that derives from the use of epistemic intensions as the relevant semantic value. The epistemic intension of an expression token is a function from centered worlds to extensions, defined at worlds that have a token at the center with the same epistemic intension as the original token, and returning the extension of that token in that world.

It is easy to see that at the worlds where it is defined, an expression's epistemic contextual intension yields the same extension as its epistemic intension. If \mathbb{W} is a centered world containing a token S' with the same epistemic intension as the original token S : let E be the extension of S' . Then the epistemic contextual intension of S returns E at \mathbb{W} . Further, the epistemic intension of S' (on the world-based view of scenarios) returns E at \mathbb{W} , since \mathbb{W} is actualized at S' . By identity of epistemic intensions, the epistemic intension of S also returns E at \mathbb{W} . So S 's epistemic contextual intension and epistemic intension are coextensive at \mathbb{W} . Something similar applies on the epistemic view of scenarios, if we invoke the scenario corresponding to the centered world \mathbb{W} .

So an expression's epistemic contextual intension is a restriction of the term's epistemic intension. For this reason, it will give appropriate quasi-Fregean results in many cases. It will not satisfy the Core Thesis: it will have the usual problems with 'A sentence token exists' and other metalinguistic claims, as it will not be defined at scenarios without the token at the center, or where the token has a different content. But it will be reasonably close for many purposes.

One could also define a epistemic version of the *cognitive* contextual intension of a token, defined at all worlds centered on a *concept* or *thought* with the same epistemic intension as the token, and returning the extension of that concept or thought.



This would again be a restriction of the expression's epistemic intension, but it would be less of a restriction, since it would not require a linguistic token in the evaluated world. The Core Thesis will still be false due to various metacognitive claims and the like; but it will not be far off. I think this last notion is the best approximation that a contextual intension can yield to a quasi-Fregean content that satisfies the Core Thesis. It is clear, however, that this notion is essentially derivative of that of an epistemic intension.

This way of seeing things also helps to explain why some other contextual intensions give approximately Fregean results. It may be that there are various other features of type F of a subject or a token that at least correlate with an epistemic intension to some degree. We can then set up an F-based contextual intension, defined at worlds centered on a subject or token with the same F features as the original, and returning the extension of the relevant token. Then in each such world, the token at the center will have at least approximately the same epistemic intension as the original token, and so in most cases will return the same or similar extension at that world. So the F-based contextual intension will approximate the behavior of a restriction of the original epistemic intension.

This applies especially to some cognitive contextual intensions. It may be that possession of a concept with a given epistemic intension is itself determined by features such as a concept's cognitive role and/or associated phenomenology, or more deeply by the subject's physical state, or functional state, or physical/phenomenal state. To know exactly which features are crucial would require a solution to the problem of intentionality, which is not yet available. But one can say: *insofar* as epistemic intensions are determined by features such as cognitive role or physical/phenomenal state, then corresponding contextual intensions (here, cognitive-role contextual intensions or physical/phenomenal contextual intensions) will be restrictions of the original epistemic intension, and so will behave in a quasi-Fregean manner. Again, however, the epistemic intension is the more fundamental notion of content.

4.3 Linguistic contextual intensions

We saw earlier (in Section 2.2) that for some expressions, a linguistic contextual intension behaves in a quasi-Fregean manner. We are now in a position to see why this is.

For some expressions, their epistemic intension is part of (or determined by) their linguistic meaning. That is, some linguistic expression types are such that every token of that type has the same epistemic intension. As noted in Section 3.7, something like this appears to apply to some pure indexicals, such as 'I', 'now', and 'today', to some descriptive terms, such as 'circular', and to some descriptive names, such as 'Jack the Ripper'.

When an expression token's epistemic intension is part of its linguistic meaning, then the token's linguistic contextual intension will be a restriction of its epistemic intension. This can be seen by the same sort of reason as in the previous section. Or one can simply apply the point there directly: if any token of S's linguistic type has the same epistemic intension, then S's linguistic contextual intension will be a restriction of its epistemic contextual intension, which is a restriction of its epistemic intension.

It follows that in cases such as ‘I’, ‘now’, ‘circular’, and ‘Jack the Ripper’, the terms’ linguistic contextual intensions will be quasi-Fregean. They will not satisfy the Core Thesis because of the restriction to worlds containing relevant tokens, but they will be reasonably close. This explains the phenomenon noted in Section 2.2: the quasi-Fregean behavior is a direct consequence of the fact that for these tokens, epistemic intension is an aspect of linguistic meaning. Once again, a contextual intension is interesting largely because of the degree to which it approximates an epistemic intension.

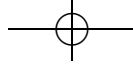
5. Other Varieties of Two-dimensionalism

With this analysis of the contextual and epistemic understandings on the table, we are now in a position to turn to existing two-dimensional proposals to see how they fit into this analysis, and to use this analysis to help understand their foundations. I should say at the start that although I will occasionally criticize these approaches and argue that the approach I have recommended has certain advantages, any advantages are due largely to building on the insights that these approaches embody.

5.1 Stalnaker’s diagonal

The diagonal proposition of Stalnaker (1978) is characterized as follows. We start with an understanding of propositions as sets of possible worlds, and with the idea that any utterance has a proposition as its content. (This propositional content coincides roughly with what I have called a subjunctive intension.) We can then say: the utterance could have had different propositional content. So there are worlds where the utterance has different propositional content. This allows us to define an utterance’s *propositional concept*, which is a function from possible worlds to propositions, defined at any world containing the utterance, returning the propositional content of the utterance at a world. We can then define the utterance’s *diagonal proposition* as the set of worlds such that the utterance’s propositional concept, evaluated at that world, yields a proposition that is true at that world.

As defined here, a diagonal proposition is much like a token-reflexive contextual intension. There are minor differences. A token-reflexive contextual intension was defined directly in terms of what an utterance’s truth-value would be at a world, rather than in terms whether the proposition it expresses would be true at that world, but it is clear that within the propositional framework, these yield the same results. A diagonal proposition is a set of possible worlds or equivalently a function from worlds to truth-values, whereas a token-reflexive proposition was a function from centered worlds to truth-values. But again, that is a minor difference: one can translate between the relevant worlds and centered worlds either by “marking” the location of the token as a center, or by removing the marked center from the location of the token. (Token-reflexive contextual intensions uniquely do not need a center to specify the relevant token, since the token is independently identified by transworld identity with the original token.) So diagonal propositions are equivalent to token-reflexive contextual intensions.



The behavior of a token-reflexive contextual intension is not clear unless we know which properties are essential to a token and which properties are inessential. It is clear that Stalnaker holds that a token's semantic properties are not essential to it, since he holds that a token could have had different semantic content. It seems plausible that on his picture, a token's orthographic properties are essential to it; at least, in all examples, a token's orthographic properties are held constant across its possible occurrences, so I will assume this in what follows. It is not entirely clear which other properties are essential (language? probably not; speaker? maybe), but we need not settle that issue for now.

From what we have said here, it appears that a token's token-reflexive contextual intension will be a restriction of its orthographic contextual intension, restricted to cases where the orthographically identical token *is* the original token. One might think that its general behavior will be very much like that of an orthographic contextual intension: for my utterance 'Water is H₂O', there will be worlds at which it means that steel is orange; if so, its token-reflexive contextual intension will be defined and presumably false there.

However, this sort of understanding is inconsistent with a central point in Stalnaker's 1978 paper, where he says that for a sentence to have a necessary diagonal proposition is for it to be an a priori truth. He also says that an official's utterance of 'this bar is one meter long' could not have expressed a false proposition. It is unclear how this can be justified. Why could not the utterance have meant something like "that boat is two miles long", and been false? It is natural to suppose that Stalnaker is holding fixed some intuitive sort of meaning and content (thus yielding something that behaves like a semantic/orthographic contextual intension). But it is not clear what aspect this could be: the only content that his account officially recognizes is propositional content, which is explicitly held to vary with possible occurrences of an utterance; and even if there were some further aspect of content, it is unclear why this sort of content should be essential to an utterance and the other sort inessential. Alternatively, it may be that Stalnaker is assuming that some sort of cognitive factor (for example, associated cognitive role?) is essential to an utterance (thus yielding something that behaves like a cognitive/orthographic contextual intension), but this is nowhere specified. In this article, the connection between apriority and diagonal propositions appears to be ungrounded.

In later work, Stalnaker does not repeat the claim about apriority, and he allows a much wider range of behavior for an utterance across possible worlds. For example, in Stalnaker (1999) he allows that there are worlds where 'Julius', used in the actual world as a descriptive name for the inventor of the zip, is used instead for the inventor of bifocals. And in Stalnaker (2001) he allows that there are worlds where our word 'tiger' refers to pieces of furniture. In both of these articles he explicitly denies that necessity of diagonal proposition corresponds to apriority, as seems reasonable. In effect, this diagonal proposition behaves very much like an orthographic contextual intension, or a straightforward restriction thereof.

As such, a diagonal proposition is clearly useful. For example, Stalnaker often uses diagonal propositions to model situations of communication, in which a hearer hears an utterance but is unsure what the speaker meant by it, or has false beliefs about

what the speaker meant by it. It seems clear that this sort of metalinguistic use requires something quite different from a quasi-Fregean notion, so this is reasonable. It is less clear that diagonal propositions are useful for addressing matters of cognitive significance, rational inference, apriority, and the like, especially when divorced from issues about communication. Stalnaker sometimes suggests this sort of use, but I think the grounds here are weaker.

For example, in a recent paper (2001), Stalnaker holds that the “metasemantic” framework with diagonal propositions can “provide an explanation for the phenomena that Kripke’s work brought to light”—where this phenomenon is the distinctive behavior of the class of a posteriori necessities such as ‘Hesperus is Phosphorus’, ‘Water is H₂O’, and so on. This is a surprising claim. What is most distinctive about these phenomena are the *differences* with standard necessities such as ‘All bachelors are unmarried’, ‘ $2 + 2 = 4$ ’, and so on. If diagonal propositions function as Stalnaker (1978) suggests, the distinction would be straightforwardly represented by the fact that the second class have necessary diagonal propositions and the first do not. But on the metasemantic understanding, there seems to be no way to draw the distinction using diagonal propositions alone. For both sorts of necessity, there will be many worlds at which the diagonal is false, and there are no clear patterns that are distinctive to the first class. So it is not clear how the “explanation” is supposed to work.

In a brief ensuing discussion of ‘Hesperus is Phosphorus’, Stalnaker appeals to the fact that there is a world where the diagonal proposition is false. But clearly this holds equally for ‘all bachelors are unmarried’. One might find some differences if one focuses on a *restriction* of the diagonal proposition. It is notable that in the worlds Stalnaker discusses, ‘Hesperus’ and the like appear to be used with the same reference-fixing intentions as the original term, for example. It may be that under this sort of restriction, the two sorts of necessities behave differently: there are counterexamples to the diagonal for one class but not the other. But the restriction is doing all the work: in effect, it invokes something more like a cognitive contextual intension, rather than a diagonal proposition per se. Such intensions may be able to model the Kripkean distinction at least approximately, if imperfectly. Stalnaker appears to use similar tacit restrictions in some other cases; in all these cases, it seems that in effect a restricted contextual intension does the explanatory work.

The explanatory power of restricted contextual intensions here itself plausibly derives from that of epistemic intensions. Epistemic intensions handle these phenomena straightforwardly: for Kripkean necessities, there is a falsifying scenario, and for standard necessities there is not. For reasons we have seen, certain restricted contextual intensions approximate epistemic intensions, and so approximate this behavior (with some exceptions). So it is plausible that the usefulness of diagonal propositions in this context derives indirectly from the usefulness of epistemic intensions.

Stalnaker contrasts his “metasemantic” version of the framework with “semantic” versions, on which 1-intensions are an aspect of semantic content, and suggest that the apparent attractions of the latter in explaining these phenomena derive from the attractions of the former. The above suggests that this is not quite right: the attractions of Stalnaker’s version of the framework in this domain derive from the

attractions of the epistemic version. As for whether the epistemic understanding is itself a “semantic” understanding: this matter depends on what is meant by “semantic”. Stalnaker mostly uses the term to contrast with “metasemantic”, indicating an aspect of first-order content rather than a metalinguistic notion: in this sense, epistemic intensions are semantic. Stalnaker also sometimes uses the term to indicate those aspects of content that are built into linguistic expression types, rather than varying across tokens: in this sense, epistemic intensions are not semantic. Stalnaker appears to assume that his opponent’s framework is semantic in both these senses;²¹ but these are very different distinctions. Epistemic intensions need not be built into linguistic meaning to be a sort of first-order content that does explanatory work.

In any case, I think it is clear that diagonal propositions and epistemic intensions both have useful roles to play. Diagonal propositions are best suited to analyzing matters of context-dependence, and epistemic intensions are best suited to analyzing the epistemic domain.

5.2 Kaplan’s character

Kaplan’s notion of character is set out as follows. We assume a prior notion of the proposition expressed by an utterance: such a proposition is something in the vicinity of a 2-intension, although it may be a singular proposition instead. For some linguistic expression types (e.g. ‘I’), utterances of the same type can express different propositions in different contexts. The character of an expression type is a function from contexts to propositions, returning the proposition that an utterance would express in a given context.

At first glance, it may seem that character is much like a linguistic contextual intension. There are some superficial differences. For example, Kaplan’s contexts are not quite centered worlds, but they include an “actual-world” and a few other parameters (speaker, time, etc), so they can be modeled by centered worlds. Also, character is a function from contexts to propositions, not to extensions. But one can diagonalize character by evaluating the proposition associated with a given context in the world of that context, yielding an associated function from contexts to extensions.

In many cases, (diagonalized) character behaves quite like a linguistic contextual intension. We have seen that the linguistic contextual intension of indexicals such as ‘I’, ‘now’, and ‘today’ pick out the speaker, time, day (and so on) of the center of all worlds at which they are defined. The same is true for (diagonalized) character, on Kaplan’s account. At the same time, we have seen that the linguistic contextual intension of a name arguably picks out the same individual at all worlds where it is defined. Again, the same applies to character, on Kaplan’s account.

²¹ The only opponent that Stalnaker cites is Chalmers (1996). I note that Chalmers (1996) explicitly leaves open (p. 58) the question of whether different speakers might associate different 1-intensions with the same word. Stalnaker also argues that the framework cannot yield an account of the a priori; I agree, and have not claimed that it can.

One case where the two apparently behave differently is for demonstratives such as ‘that’.²² If I use ‘that’ intending to refer to an object in front of me, then its character will pick out (roughly) an object in front of the speaker in all contexts. But the linguistic contextual intension will not: what it picks out in a context will depend on the underlying demonstration or intention of a speaker in that context. But in a way, this is the exception that proves the rule. In Kaplan’s formal analysis, he stipulates that different uses of ‘that’ (roughly, those corresponding to different demonstrations) are tokens of *different* words: ‘that₁’, ‘that₂’, and so on. Under this stipulation, it is plausible that a linguistic contextual intension for one of these instances of ‘that’ will behave as characterized above.

However, there are aspects of Kaplan’s discussion that make it clear that character is fundamentally different from a linguistic contextual intension. Kaplan stresses that when we evaluate a sentence’s character in a context, we do not evaluate an *utterance* of that sentence within the context. Rather, we evaluate an *occurrence* of the sentence at the context. An occurrence is in effect an ordered pair of a sentence and a context. And crucially, the context need not itself contain an utterance of the sentence. In effect, this is to allow that the character of an expression can be evaluated directly at a centered world, whether or not the world contains a token of the original expression.

Kaplan’s reason for doing this are largely tied to his desire for a *logic* of demonstratives. He suggested that arguments involving demonstratives should be *valid* in virtue of their character: that is, a conclusion should follow from premises in virtue of an appropriate relation among their characters. But if the character of a claim were restricted to contexts containing an utterance of that claim, then each premise and the conclusion would be defined across different contexts, so their characters could not stand in the right sort of relation. He also says that there are sentences that express a truth in certain contexts, but in no contexts in which they are uttered: for example, ‘I say nothing’. If so, contexts cannot be required to contain a token of the relevant utterance.

For these purposes, it is natural to suggest that character should be something more like an epistemic intension. Validity, at least as Kaplan uses it here, is a deeply epistemic notion, tied to apriority and to rationally compelling inferences (in Kaplan’s discussion, it is clear that validity is not tied constitutively to necessity). The sort of intension that is *constitutively* tied to validity and to apriority is an epistemic intension. Similarly, for the intension of ‘I say nothing’ to be false in the relevant contexts, the best candidate is something like an epistemic intension.

It is difficult to adjudicate what Kaplan intends, however, since he never specifies how to evaluate an expression’s character in a context. He simply stipulates that expressions have a character associated with them, and then discusses the character’s properties. He does say on some occasions that character picks out what the expression would pick out if uttered in that context, but he retracts this because of the point about occurrences. (It presumably remains the case that *when* a context contains the right sort of utterance, character returns what the utterance picks out).

²² I use “demonstrative” for expressions such as ‘that’, ‘he’, and ‘you’, while using “indexical” for expressions such as ‘I’, ‘here’, and ‘now’.

He also says (505) that character is set by linguistic conventions and determines the content in a context, and he suggests that character is determined by a demonstration (526–27) or a directing intention (587–8). But nothing here tells us how to evaluate character in contexts not containing the utterance. In some cases the matter seems reasonably straightforward: ‘I’ picks out the marked subject in a context, ‘you’ picks out a marked addressee. But there seems to be no general principle here for assigning an evaluation function to an expression type.

(Another complication is that it is not entirely clear what is built into the relevant context. On a couple of occasions (528, 588) Kaplan entertains the idea that a context explicitly contains a parameter for a demonstratum, which serves as referent for a demonstrative. If this is done, it renders the question about how to evaluate character trivial, but at the cost of trivializing many other aspects of the framework. It also removes any special role for demonstrations and directing intentions in contextual evaluation, and removes the deep connection with cognitive significance. Partly for these reasons, and partly because it eliminates the connection between demonstratives and indexicals, this seems not to be Kaplan’s considered view. The alternative is a view on which the referents of demonstratives are not explicitly specified within a context, but instead are picked out by a directing intention.)

One way to address the question is to ask: is it *constitutive* of character that validity and apriority are governed by character? Or is this merely a feature that character turns out to have? It seems that it cannot be constitutive, for the obvious reason that in the case of proper names, validity and apriority come apart from character. But now the question arises: what justifies Kaplan’s claim that the character of indexicals and demonstratives must be logically well-behaved? In effect, it is this claim that determines his treatment of occurrences. If character were definitionally connected to validity, the claim would be reasonable; but character is not definitionally connected to validity. If character is independently grounded, it seems that one might equally say: character is reasonably close to reflecting validity and the like, but unfortunately the correspondence is imperfect, even for indexicals and demonstratives.

Perhaps the most likely diagnosis is the following. The initial notion of character is not constitutively connected to the domain of validity and apriority. (Perhaps it is something like a linguistic contextual intension.) But at least in the case of indexicals and demonstratives, this notion of character turns out to come very close to reflecting this domain. It turns out that a slight modification makes the correspondence precise, so at least in this case we adopt the modified notion. The resulting notion appears to be something quite close to an epistemic intension. It is not exactly an epistemic intension, for example because of the use of further parameters in a context. But it seems to behave in a quite similar way.

This raises the question: why not do the same for names? If character is to be connected to apriority, why not understand the character of a name so that it behaves something like an epistemic intension? The initial answer is that Kaplan thinks that names do not behave this way: their contents are essential to them, so they do not pick out different contents in different contexts. In discussing the matter, Kaplan notes especially (562) that occurrences of ‘Aristotle’ that refer to different people are different words. One might respond that this would be relevant if we were defining

the *contextual* intension of names, but we are now dealing with a modified notion. The fact that a name has its referent essentially (and the point about 'Aristotle') is compatible with its *epistemic* intension picking out different referents in different contexts.²³ But the crucial point may in fact be something different: character is supposed to be a sort of linguistic meaning, but the names as linguistic types do not have epistemic intensions (at best, epistemic intensions vary between tokens).²⁴ So character cannot be epistemic intension.

Still, an obvious response is that the same holds for demonstratives. Kaplan's formal move of stipulating that different tokens of a demonstratives are different words is clearly something of a convenient trick: the force of this move is to suggest that character need not really be associated with a linguistic type, but with a token.²⁵ If so, then we could say the same for names, perhaps making a similar stipulation, or perhaps not. It is not clear exactly how the cases are relevantly different. One suggestion is that different tokens of a name seem to be more closely tied together than different tokens of a demonstrative, with some sort of associated assumption of communication, agreement and disagreement, and so on. If so, then assigning all these name tokens a different linguistic type might be even more counterintuitive than for demonstratives. But it is not clear exactly what the rules are here. One should arguably take the real moral of the demonstrative case to be that character is fundamentally a property of tokens rather than linguistic types, in which case it is no longer obvious that names must have trivial character.

In any case, my best guess as to what constitutes character is the following: character is something like an epistemic intension in cases where it is reasonable

²³ I have not mentioned Kaplan's 'Fregean theory of demonstrations', according to which a demonstration does not have its referent essentially. It seems that this point would be highly relevant to a contextual understanding of demonstratives, but it is not so clear that it is required for an epistemic understanding.

²⁴ This sort of point about the difference in cognitive significance between different tokens of a name is never mentioned explicitly in Kaplan's article, but it may be playing a role implicitly in his claim that names do not have nontrivial character semantically associated with them. A useful diagnostic question would be whether descriptive names (if there are any), such as 'Jack the Ripper', can have nontrivial character. If yes, then variability is plausibly the key reason that standard names have trivial character. If no, then essentiality of referent is plausibly the key reason.

²⁵ See Braun (1996) on this topic. Braun notes that Kaplan sometimes adopts the informal strategy of taking 'that' to be a single word (type), associating character not with the word but with a word-plus-demonstration pair. This raises the question: since the relevant demonstrations (especially according to the later Kaplan) are a sort of intention, why not analogously associate character with a name-plus-intention pair? Then one will in effect have character for utterances of names.

One might even note: for a use of a name, there can be a directly linked demonstrative. This can happen with an anaphoric use in 'John . . . he . . .', or better, with a simple non-anaphoric 'he' backed by the intention to refer to John (perhaps this can be a mild counterfactual variant on the original utterance). It does not seem entirely unreasonable to say that this last demonstrative has nontrivial character. If so, one could use this character to motivate a nontrivial character in the vicinity of any token of 'John'. If not, then it seems to follow that the Fregean theory of demonstrations is false in at least some cases, and one wants to know more about the rules for associating character with demonstratives and demonstrations.

(perhaps at a stretch) to assign an epistemic intension to a linguistic type. If not, because of variability of epistemic properties between tokens, then character is something else, perhaps more like a contextual intension, or perhaps stipulated to be a function that returns an expression's actual content (whatever it is) at all contexts. Alternatively, if character really is something like "epistemic intension insofar as it is associated with a linguistic type", one might equally simply say that names have no character, rather than saying that they have constant character. That is not Kaplan's official view, but it does not seem wholly contrary to the spirit of his discussion.

There are a couple of interesting diagnostic cases. First, what is the character of a descriptive name, such as 'Jack the Ripper'? Second, what is the character of a context-sensitive predicate such as 'heavy'? Kaplan is silent about these matters, but the behavior of character in these cases might help decide just how the notion of character is grounded.

The distinction between the contextual and epistemic understandings also helps bear on a recent controversy about occurrences. Garcia-Carpintero (1998) argues for the superiority of a Reichenbachian token-reflexive account of indexicals over an account that relies on Kaplan's occurrences. In effect, he suggests that a sort of token-reflexive contextual intension (requiring the token in a context) is truer to the data than an account that does not require tokens. As part of his argument, he denies that there is any reasonable intuition that there are contexts in which 'I am not uttering now' is true. Our discussion makes it possible to render a split verdict here. On a contextual understanding of evaluation in contexts, there is no such intuition. But on an epistemic understanding, there is such an intuition. The intuition, I think, is that 'I am not uttering now' is not false a priori, so that there are epistemic possibilities in which it is true. These epistemic possibilities are scenarios in which the subject at the center is not uttering. The difference between Kaplan's and Reichenbach's frameworks may then be grounded in the fact that Kaplan's semantic value for an indexical is constitutively tied to its epistemic properties, while Reichenbach's is tied to its contextual properties.²⁶

In any case, it seems plausible that there are elements of both the contextual understanding and the epistemic understanding in Kaplan's account, not always disentangled. Perhaps character is fundamentally an extension of a linguistic contextual intension; perhaps it is fundamentally a sort of epistemic intension; or perhaps there is no fact of the matter. But it is clear that much of the value of character in the case of demonstratives comes from the fact that in this case, character behaves much as an epistemic intension does (whereas in the case of names, it does not). It does not seem unreasonable to hold that character is useful for

²⁶ Note, though, that on the framework I have suggested, epistemic intensions are fundamentally assigned to utterances, not to occurrences. Occurrences play a role in that epistemic intensions are *evaluated* at scenarios that need not contain the relevant utterance. This suggests that there are two quite distinct issues dividing the Reichenbachian and the Kaplanian: the issue of whether semantic values should be assigned to utterances or occurrences, and the issue of whether these semantic values can be evaluated in worlds (or contexts) in which the utterance is absent.

epistemic purposes precisely to the extent that it approximates or coincides with an epistemic intension.

5.3 Evans' deep necessity

In addressing Kripke's problems of the contingent a priori, Evans (1979) focuses on the case of descriptive names. He introduces the descriptive name 'Julius', whose referent is fixed as being whoever invented the zip. Then 'Julius invented the zip' seems to be a priori. In analyzing the case, Evans distinguishes between two sorts of necessity: "deep necessity" and "superficial necessity". Instances of the "contingent a priori" (such as 'Julius invented the zip') are superficially contingent but deeply necessary; instances of the "necessary a posteriori" are superficially necessary but deeply contingent. Evans says that whether a statement is deeply necessary or contingent depends on what makes it true; and whether it is superficially contingent depends on how it embeds under modal operators.

Superficial necessity is defined as follows. A sentence Q is superficially contingent if ' $\diamond \sim Q$ ' is true, or equivalently, if there is some world W where Q is not true $_W$. Here, the possibility operator is clearly subjunctive possibility ("it might have been that"), and the possible-worlds evaluation is clearly subjunctive counterfactual evaluation of the Kripkean sort. So superficial necessity is a second-dimensional notion: S is superficially necessary when it has a necessary 2-intension or subjunctive intension.

Deep necessity and contingency are characterized in the following passage toward the end of Evans' article:

- Q1 We have the idea of a state of affairs, or a set of state of affairs, determines• by the content of a statement as rendering it true, so that one who understands the sentence and knows it to be true, thereby knows that such a verifying state of affairs exists. A deeply contingent statement is one for which there is no guarantee that there exists a verifying state of affairs. If a deeply contingent statement is true, there will exist some state of affairs of which we can say both that if it had not existed the sentence would not have been true, and that it might not have existed. The truth of the sentence will thus depend on some contingent feature of reality. (Evans 1979, 185)

This passage has a strong epistemic element in the first half; and a strong contextual element in the second half. To understand these we need to examine the discussion earlier in Evans' article.

Evans introduced the notion of the *content* of a sentence earlier as capturing an epistemic element. Evans says that when two sentences have the same content, they are *epistemically equivalent*: a person who understands both cannot believe what one says and disbelieve what the other says. Evans makes a distinction between the content of a sentence and the *proposition* expressed by a sentence, which is a function from possible worlds to truth-values of the sort associated with the modal contexts of superficial necessity. He notes that two sentences that express the same proposition can have different contents, and argues that two sentences with the same content can express different propositions: e.g. 'Julius is F' and 'The inventor of the zip is F'.

Evans holds that there is a notion of “making a sentence true” that is tied directly to content (which he distinguishes from an alternative sense tied to proposition expressed). He says:

... if two sentences are epistemically equivalent, they are verified by exactly the same state of affairs, and what one believes in understanding the sentence and accepting it as true is precisely that some verifying state of affairs obtains. On this conception, the same set of states of affairs makes the sentence ‘Julius is F’ true as makes the sentence ‘The inventor of the zip is F’ true. If x, y, z, \dots is a list of all objects, then any member of the set { x ’s being the inventor of the zip & x ’s being F; y ’s being the inventor of the zip and y ’s being F; z ’s being the inventor of the zip and z ’s being F ... } will suffice to make the sentence true. (Evans 1979, 180)

On this conception, making a sentence true, at least in the case of a descriptive name, seems to involve something like satisfying its epistemic intension. In the sense of ‘verify’ that I tied to epistemic evaluation, ‘Julius invented the zip’ will be verified precisely when the conditions that Evans suggests for “making the sentence true” obtains. The claim that epistemic equivalence entails verification (in Evans’ sense) by the same states of affairs also suggests a tie to epistemic intension, and suggests a link between the two notions of ‘verification’. If deep necessity is tied to ‘making true’ in this sense, then at least in the case of descriptive names it seems to be a sort of necessity of epistemic intension.

Evans also characterizes this notion of ‘making true’ in alternative terms:

But there is an ineliminable modal element in the notion of what makes a sentence true. For what can it mean to say that any one of a set of states of affairs renders a sentence true, other than to say that, if any one of them obtains, the sentence will be true, and if any of them *had* obtained, the sentence *would have been* true.

This characterization has a more contextual flavor: “making true” is characterized in terms of a metalinguistic subjunctive about truth-values the sentence *could* have had. This suggests something like a linguistic contextual intension. But such an understanding will yield quite different results from the understanding above. For example, if ‘L’ is a descriptive name for the number of sentence tokens ever produced, then no token of ‘ $L > 0$ ’ could have been false. If “making true” is understood in terms of the possible truth of tokens, this will entail that ‘ $L > 0$ ’ is deeply necessary, even though it is clearly a posteriori, and will be deeply contingent on the earlier understanding. It seems doubtful that Evans would allow that this sentence is deeply necessary.²⁷

Alternatively, we might understand the locution “the sentence would have been true” as invoking an abstract sentence, one that need not be uttered in the state of affairs in question. But we must tread carefully here. It is natural to hold that the abstract sentence ‘Julius invented the zip’ would have been true in a state of affairs if and only if Julius invented the zip in that state of affairs. But Evans accepts the Kripkean claim that there are states of affairs such that if they obtained, Julius did not invent the zip. It follows from these two claims that the abstract sentence ‘Julius

²⁷ Evans’ letter to Martin Davies (this volume) suggests very strongly that he would regard sentences such as ‘ $L > 0$ ’ as deeply contingent, and that he would reject a contextual interpretation of deep necessity.

invented the zip' could have been false, so that on this understanding, it is not deeply necessary. So for this strategy to work, Evans must reject the claim that the abstract sentence 'Julius invented the zip' is true in a state of affairs iff Julius invented the zip there, and must give some other account of the evaluation of abstract sentences in states of affairs.²⁸

The most natural way to reconcile all this is to interpret the locution "the sentence would have been true" as meaning that the *content* of the abstract sentence would have been true. Then the result above will follow, given Evans' view that the content of "Julius invented the zip" is a descriptive content that differs from the proposition that is contributed to modal contexts. The cost is that on this approach, we cannot use a prior notion of "making true" to *ground* the notion of content, as one can do on some other approaches. Rather, the notion of "making true" and the consequent notions of deep necessity and the like are defined in terms of a prior notion of content.

On Evans' view of content, descriptive names have a descriptive content, because of their epistemic equivalence to descriptions. In this case, "making true" behaves like an epistemic intension. In other cases, it may not. Elsewhere (Evans 1982), Evans rejects the claim that ordinary proper names have a quasi-descriptive content, suggests that the referent of an ordinary proper name is part of its content. On this sort of view, it appears that a sentence such as 'Cicero is Tully' is made true by all states of affairs, so that it is deeply necessary. If this is correct, then deep necessity can come apart from apriority, and Evans' notion of verification does not in general behave in the manner of an epistemic intension.

One source of this difference is that Evans associates content with expression types rather than expression tokens, and imposes a semantic constraint on content: for Evans, the content of an expression type is closely tied to what is required for a speaker of a language to understand it. Given that epistemic intensions are often variable across competently used tokens of an expression type, it follows that epistemic intensions cannot be content in Evans' sense. But in cases such as that of descriptive names, where a specific epistemic intension is required for competent use of an expression type, one can expect that Evans' notion of content will behave more like an epistemic intension, and that deep necessity will coincide more closely with apriority.

Because of all the dependence on a prior notion of content, Evans' account is not naturally assimilated to either an epistemic or contextual understanding of

²⁸ To support his claims about which sentences could have been true, Evans goes to some length to argue that if *y* had invented the zip and had been *F*, *y* would have been the referent of 'Julius', and 'Julius is *F*' would have been true as a *sentence of English*. He argues that there is no *semantic* connection between 'Julius' and a particular referent, so one can suppose that the term could have had a different referent without supposing a semantical change in English. He says: "exactly the same theory of meaning serves to describe the language which would be spoken had *y* invented the zip, as describes the language which is actually spoken" (Evans 1979, 182). These passages use claims about counterfactual *spoken* language to support claims about "making true", so one might initially read them as supporting a contextual interpretation of this notion. But they are arguably also compatible with the second understanding above, if one conjoins this understanding with the thesis that an abstract sentence is true in a state of affairs *if* (although not only if) a token of a semantically identical sentence is true in that state of affairs.

two-dimensionalism. We might think of it as a broadly “semantic” understanding: Evans’ first-dimensional modal notions are defined in terms of a prior notion of content, and their grounds depend on the grounds of that notion of content. Given Evans’ own distinctive understanding of content, the result is a first-dimensional modal notion that lines up with the epistemic understanding in some cases but not in all. The result is that deep necessity coincides with apriority in some cases (for example, cases involving descriptive names), but not in all cases.

Of course, if the invokes a different notion of content, one will get a different corresponding notion of deep necessity. For example, if one loosens Evans’ epistemic constraint on content and embraces a view on which the content of a name just involves its referent, then the corresponding notion of deep necessity may coincide with superficial necessity. And if one loosens Evans’ semantic constraint on content and embraces a view on which the content of a token is something like an epistemic intension, then the corresponding notion of deep necessity may coincide with apriority.

5.4 Davies and Humberstone’s “fixedly actually”

Davies and Humberstone (1981) give a “formal rendering” of Evans’ distinction between deep and superficial necessity, using independently motivated tools from modal logic developed in Crossley and Humberstone (1977). The formal framework starts with a necessity operator N , and supplements it with an “actually” operator A , meaning “it is actually the case that”. This allows one to represent claims such as ‘It is possible for everything which is in fact φ to be ψ ’, as $\Diamond \forall x (A\varphi(x) \supset \psi(x))$. A model theory for these operators requires supplementing the space of possible worlds needed for the necessity operator with a designated “actual world”, where $A\alpha$ is true at a possible world iff α is true at the actual world.

This framework naturally suggests a further idea: just as one can ask whether $\alpha \pm$ is true with respect to a possible world (holding the actual world fixed), one might ask whether $\alpha \pm$ would be true if a *different* world were designated in the actual world. This notion is modeled by adding a further “fixedly” operator F , where $F\alpha$ is true at a world W iff α is true at W no matter which world is designated as actual. Here, the model theory requires that we have a “floating” actual world, or alternatively, it can invoke double-indexed evaluation of sentences at worlds. On the double-indexing approach, we can say that α is true at (V, W) when α is true with respect to W , when V is designated as actual.²⁹

²⁹ What follows is a two-dimensional “reconstruction” of Davies and Humberstone’s framework. In their original paper, Davies and Humberstone’s official model theory for the system with F and A involves one-dimensional evaluation with a “floating” actual world, although they note the possibility of a model theory with two-dimensional evaluation.

It is worth observing that it is not obviously correct to say, as is often said, that the ideas of two-dimensional semantics are grounded in two-dimensional modal logic. Two-dimensional modal logic *per se* does not play a crucial role in grounding the frameworks of Kaplan, Stalnaker, and Evans; and even in the case of Davies and Humberstone, two-dimensional modal logic is presented merely as an optional means of representation. Of course many of these ideas can be naturally

The double-indexed evaluation can be formally defined in terms of its interaction with the relevant modal operators.³⁰ When α is a simple sentence, α is true at (V, W) iff α is true at W according to single-indexed evaluation. When α has the form $F\beta$, α is true at (V, W) when for all V' , β is true at (V', W) . When α has the form $A\beta$, α is true at (V, W) when α is true at (V, V) . When α has the form $\Box\beta$, α is true at (V, W) when for all W' , β is true at (V, W') . The truth of other complex sentences at (V, W) is the obvious function of the truth of their parts at (V, W) . It follows that when α does not contain F or A , the evaluation of α at (V, W) is independent of V . In these cases, the double-indexed evaluation of α at (V, W) is the same as the single-indexed evaluation of α at W . If α contains F or A , double-indexed and single-indexed evaluation may come apart.

One can then introduce the combined operator FA , which functions so that $FA\alpha$ is true at (V, W) iff for all worlds V' , α is true at (V', V') . Or more simply, $FA\alpha$ is true when α is true at all worlds when that world is designated as actual. Davies and Humberstone note that this operator can be seen as yielding a sort of necessity. We might say that a sentence α is FA -necessary when $FA\alpha$ is true. For sentences that do not contain F or A , FA -necessity and ordinary necessity coincide: for all such sentences α , $FA\alpha$ is equivalent to $\Box\alpha$. But for sentences containing F or A , FA -necessity and ordinary necessity behave differently. For example, if α has the form $A\beta$, then $FA\alpha$ is equivalent to $\Box\beta$.

The A operator can be used to represent some contingent a priori truths. For example, if φ is a contingent truth, then ' $A\varphi \leftrightarrow \varphi$ ' is contingent but a priori. Truths of this sort are not necessary, but they are FA -necessary: for example, $FA(A\varphi \leftrightarrow \varphi)$ is equivalent to $\Box(\varphi \leftrightarrow \varphi)$, which is true. This behavior parallels Evans' observation that contingent a priori sentences are not superficially necessary, but they are deeply necessary.

Davies and Humberstone extend the parallel between deep necessity and FA -necessity to the case of descriptive names, by suggesting that descriptive names are abbreviations of descriptions of the form 'the actual G ', for an appropriate G . On this view, 'Julius' abbreviates 'the actual inventor of the zip'. Then the contingent a priori sentence 'if anyone uniquely invented the zip, Julius did' (which Evans holds is superficially contingent but deeply necessary) is equivalent to 'if anything uniquely has G , the actual G has G '. In Davies and Humberstone's formal framework, this sentence is contingent but FA -necessary. In light of these parallels, Davies and Humberstone put forward the hypothesis that a sentence is deeply necessary in Evans' sense iff it is FA -necessary.

Using this framework, one can define a corresponding sort of 1-intension. We can say the FA -intension of a sentence α is true at W when α is true at (W, W) according to Davies and Humberstone's method of evaluation. In the case of sentences with descriptive names, FA -intensions behave something like a quasi-Fregean semantic

represented using the tools of two-dimensional modal logic, and there is plausibly a relationship between the conceptual bases of these frameworks and of two-dimensional modal logic.

³⁰ This definition of two-dimensional evaluation is not given explicitly by Davies and Humberstone, but it is easy to see that it gives the intended results.

value. Davies and Humberstone also suggest (without endorsing the suggestion) that one might extend this treatment to other terms, such as 'water', 'red', or 'good', analyzing these as equivalent to 'actually'-involving descriptions. This raises the question: do FA-intensions satisfy the Core Thesis? Or: is a sentence a priori iff it is FA-necessary?

Addressing this sort of question, Davies and Humberstone say that they have found no examples of FA-contingent a priori sentences, and they appear to be sympathetic with the claim that there are no such sentences. But they say that there are many FA-necessary a posteriori sentences, including identities between ordinary proper names ('Cicero = Tully'). This asymmetrical attitude toward the FA-necessary a posteriori and the FA-contingent a priori mirrors Evans' attitude concerning deep necessity. Still, it is *prima facie* surprising that a formally defined notion such as FA-necessity should yield this asymmetry. So it is worthwhile to assess these claims independently.

A simple approach to these questions runs as follows. Let us say that a sentence of natural language is A-involving if its logical form contains an occurrence of 'A' or an occurrence of 'F' (in practice there will be few occurrences of 'F', or at least few occurrences of 'F' unaccompanied by 'A'). If there exist non-A-involving sentences that are necessary a posteriori, then these sentences are FA-necessary a posteriori. If there exist non-A-involving sentences that are contingent a priori, then these sentences are FA-contingent a priori. If no non-A-involving sentences fall into either of these classes, then no A-involving sentences fall into either of these classes. So FA-necessity and apriority are co-extensive if and only if necessity and apriority are co-extensive for non-A-involving sentences.

Are there non-A-involving necessary a posteriori sentences? On the face of it, it seems so. For example, identities between ordinary proper names can plausibly be necessary and a posteriori, and such names are plausibly non-A-involving (as Davies and Humberstone themselves note). To resist this claim, one would need to maintain that ordinary proper names, like descriptive names, abbreviate (or are equivalent in logical form to) descriptions of the form 'the actual G'. But there are numerous reasons to doubt such a claim, even on the broadly Fregean view that I have outlined. For example, we have seen that the epistemic intension of a name can vary from speaker to speaker in ways that the epistemic intension of a description does not, so there can be no equivalence in standing meaning between names and descriptions. Even for a single speaker, there may be no expression in the language that encapsulates the epistemic intension of the name as used by that speaker. Further, it is plausible that names have their referents essentially, but descriptions of the form 'the actual G' do not. If this is correct, then ordinary proper names are not A-involving, and identities between them are examples of the FA-necessary a posteriori.

Are there non-A-involving contingent a priori sentences? On the face of it, it seems so. For example, ordinary indexicals such as 'I', 'here', and 'now' give rise to instances of the contingent a priori, and such indexicals are plausibly non-A-involving. For example, 'I am here now (if I exist and am spatiotemporally located)' appears to be both contingent and a priori. Perhaps one could hold that at least one of the indexicals is A-involving: for example, one could suggest that 'I' is equivalent in logical form

to ‘the actual speaker’, or that ‘here’ is equivalent in logical form to ‘the place where I actually am now’. But these suggestions are unappealing in a number of respects,³¹ and are widely rejected in semantics. If these indexicals are not A-involving, then the sentence is plausibly FA-contingent a priori.³²

These conclusions accord with Davies and Humberstone’s view of the FA-necessary a posteriori, but not with their view of the FA-contingent a priori. If correct, these conclusions cast doubt on Davies and Humberstone’s claim that FA-necessity is equivalent to Evans’ deep necessity. For Evans, the existence of a deeply contingent a priori sentence is “intolerable”: it appears to be a conceptual constraint on his notion of content that any a priori sentence has a content that is verified by any state of affairs. If so, and if there are FA-contingent a priori sentences, then it seems that deep necessity is not the same as FA-necessity.

Is the two-dimensional framework of Davies and Humberstone fundamentally a contextual approach or an epistemic approach? As Davies (2004) notes, it seems to be neither. It is clearly not a contextual approach: sentence tokens present in counterfactual worlds play no special role here.³³ And it seems not to be an epistemic approach: epistemic notions play no role in defining the key concepts. I think it is best regarded as a *formal* approach: FA-necessity is in effect defined in terms of its interaction with ‘A’ and ‘F’ operators in a sentence’s logical form. This formal definition yields results that are consonant with those of an epistemic interpretation in some cases, but not in all cases.

This consonance stems from the fact that where the ‘A’ and ‘F’ operators are concerned, FA-intensions behave very much like epistemic intensions. If the *only* source of a posteriori necessary and contingent a priori sentences were the ‘A’ and ‘F’ operators, then FA-intensions and epistemic intensions would coincide. But we have seen that these operators appear not to be the only source of these phenomena. Because of this, the two intensions do not coincide, and the Core Thesis fails for FA-necessity.

Of course one might hold that even in the cases of non-A-involving terms that generate a posteriori necessary and contingent a priori sentences, there is something relevant in common with the behavior of A-involving sentences. For example, one might hold that utterances of indexicals and ordinary proper names involve the rigidification

³¹ For example, the claim that ‘I’ is equivalent to ‘the actual speaker’ has the unappealing consequence that ‘if I exist now, I am speaking’ is a priori. The ‘claim’ that ‘here’ is equivalent in logical form to ‘the place where I actually am now’ introduces an unappealing asymmetry between the logical forms of ‘here’ and ‘now’ that appears to be ad hoc and otherwise unmotivated.

³² For other examples of non-A-involving contingent a priori sentences, one might try sentences with complex demonstratives or partially descriptive names: for example, ‘that picture (if it exists) is a picture’, or ‘Pine Street (if it exists) is a street’. These sentences are plausibly contingent and are not obviously A-involving. On some views (but not all), these sentences are a priori. If so, these sentences are plausibly FA-contingent a priori.

³³ Surprisingly, Evans seems to have understood Davies and Humberstone’s notions as broadly contextual notions. See his letter to Martin Davies (included in this volume), in which he raises “utterance difficulties” for the framework, involving sentences such as ‘I exist’ and ‘There are no speakers’. These parallel the issues raised concerning the contextual understanding in Section 2.4 of this paper.

of some sort of Fregean content, even if these expressions are not equivalent in logical form to corresponding A-involving descriptions. If so, one could use this behavior to define a broader sort of two-dimensional evaluation of sentences that does not turn entirely on the presence of ‘F’ and ‘A’ operators. If one generalized Davies and Humberstone’s framework in this way, the resulting framework would more closely resemble the epistemic framework that I have outlined.

5.5 Chalmers’ primary intensions

In *The Conscious Mind* (Chalmers (1996), 56–65), I present “a synthesis of ideas suggested by Kripke, Putnam, Kaplan, Stalnaker, Lewis, Evans, and others”.³⁴ I distinguish what the “primary intension” and the “secondary intension” of a concept (where a concept is understood as either a linguistic or a mental token). How are these intensions to be understood? Here I will examine the text from the outside, leaving autobiographical remarks until the end.

The two intensions are initially characterized as follows (p. 57):

There are two quite distinct patterns of dependence of the referent of a concept on the state of the world. First, there is the dependence by which reference is fixed in the actual world, depending on how the world turns out; if it turns out one way, a concept will pick out one thing, but if it turns out another way, the concept will pick out something else. Second, there is the dependence by which reference in *counterfactual* worlds is determined, given that reference in the actual world is already fixed. Corresponding to each of these dependencies is an intension, which I will call the *primary* and *secondary* intensions, respectively.

The secondary intension seems to be the familiar sort of intension (2-intension, subjunctive intension) across possible worlds. The nature of the primary intension is somewhat less clear. The characterization above has both contextual and epistemic elements. The reference to what “a concept will pick out” under certain circumstances suggests a sort of contextual intension; but reference to how the world “turns out” suggests an epistemic element.

I also say (p. 57) that a concept’s primary intension is “a function from worlds to extensions”, such that “in a given world, it picks out what the referent of the concept would be if that world turned out to be actual”. (The “worlds” are later refined to centered worlds.) This is similar to the characterization above, although the use of “turned out” and “would be” arguably has a slightly different (more subjunctive, less epistemic?) flavor than the use of “turns out” and “will”. Again, the referent to potential reference of a concept suggests some sort of contextual intension. This is also suggested by a later discussion (p. 60) which casts the worlds in the domain of a primary intension as Kaplanian “contexts of utterance”, and which asks “how things would be if the context of the expression turned out to be W.” And again (p. 63):

³⁴ In retrospect, the “synthesis” remark is unfortunate. As we have seen, the formal similarities between the different frameworks mask deep conceptual differences, which are largely ignored in Chalmers (1996). One moral of this paper is that a blanket citation of theorists who have worked on two-dimensional ideas has the potential to confuse more than it clarifies.

“The primary truth-conditions tell us how the actual world has to be for an utterance to be true in that world; that is, they specify those *contexts* in which the statement would turn out to be true.”

If a primary intension is a contextual intension, what sort of contextual intension is it? The discussion suggests an intension that exhibits the sort of quasi-Fregean behavior described in Section 1 of this paper. It seems clear that a primary intension is not intended to be an orthographic contextual intension: nothing in the discussion suggests that in a world where ‘water’ means steel, the primary intension of our term ‘water’ picks out steel. It may be intended to be a linguistic contextual intension: footnote 21 on p. 364 suggests sympathy with the view that the word ‘water’ as used on Twin Earth is of the same linguistic type as ours, in which case a linguistic contextual intension may give quasi-Fregean results. It may also be that some sort of cognitive contextual intension is intended, where one holds fixed the epistemic situation of the subject. But the matter is not clear.

There are also a number of elements in the discussion that suggest an epistemic understanding. The expression “what a concept will refer to if the world turns out” carries an epistemic flavor that is quite different from the subjunctive “what a concept would refer to if the world turned out”; there is arguably more plausibility in the idea that it could *turn out* that ‘water’ refers to XYZ than that it could have *turned out* that ‘water’ refers to XYZ. So perhaps there is a sort of amalgam of epistemic and contextual ideas at work in this phrase.

More clearly, the discussion of how to evaluate a primary intension has a strong epistemic element. I say:

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? For example, if it had turned out that the liquid in the lakes was H₂O and the liquid in the oceans was XYZ, then we would probably have said that both were water; if the stuff in oceans and lakes was a mixture of 95 percent A and 5 percent B, we would probably have said that A but not B were water.

Here, the suggestion seems to be that a term’s primary intension is constituted to a speaker’s or a community’s dispositions to apply the term, depending on what is discovered to be the case. This suggests something at least in the vicinity of an epistemic intension. There is still a metalinguistic element in “what would we say?”; for this reason, it seems hard to extend this heuristic to such cases as evaluating “language exists” in a language-free world, and so on. But the idea of capturing the dependence of judgments about extension on discoveries about the actual world suggests something fundamentally epistemic.

Further evidence for an epistemic interpretation stems from two endnotes (notes 26 and 29, p. 366) in which it is stated that one can evaluate a primary intension in worlds that do not contain the original concept. I give the example of “I am in a coma”, suggesting that the primary intension should be true of centered worlds where the individual at the center is in a coma and not thinking anything. This is more compatible with an epistemic interpretation than with a contextual interpretation.

The strongest evidence is in footnote 21 (p. 364), which responds to an objector who holds that ‘water’ on Twin Earth is a different word:

If one is worried about this . . . one can think of these scenarios as *epistemic* possibilities (in a broad sense) and the conditionals as epistemic conditionals, so that worries about essential properties of words are bypassed.

This response suggests the basis of an epistemic understanding, albeit in quite sketchy terms.³⁵ This interpretation also fits the claim (p. 64) that a sentence is a priori when it has a necessary primary proposition (where here a proposition is an intension for a statement), and the use of primary intensions to make an inference from conceivability to possibility (roughly, from a claim’s a priori coherence to the existence of a world satisfying a claim’s primary intension). These moves will be invalid in general if primary intensions are contextual intensions. But if primary intensions are epistemic intensions, it is possible that they are correct.

Autobiographically: I think that primary intensions as I conceived them (both in Chalmers (1995) and in Chalmers (1996)) were much more like epistemic intensions than like contextual intensions. But the distinction between contextual and epistemic understandings was not sufficiently clear in my mind at the time of writing, and is certainly not clear on the page. (The current paper is in part a mea culpa.) Certainly, one must interpret primary intensions as epistemic intensions to make sense of the applications of the two-dimensional framework in these works. (For example, the main conceivability-possibility argument in Chalmers (1996) turns on a version of the thesis of Metaphysical Plenitude outlined earlier.) If one does so, I think the resulting arguments are sound.

5.6 Jackson’s A-intensions

Jackson (1998a) discusses “a distinction between two fundamentally different senses in which a term can be thought of as applying in various possible situations”. He says:

We can think of the various situations, particulars, events, or whatever 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. In the first case we are considering, for each world W , what the term applies to in W , given or under the supposition that W is the actual world, our world. We can call this the A-extension of term T in world W —‘A’ for actual—and call the function assigning to each world the A-extension of

³⁵ I say a bit more to suggest epistemic evaluation in Chalmers (1995) which develops the framework to yield an account of the narrow content of thought. Here I say that to evaluate a primary intension, one can ask questions such as “If W turns out to be actual, what will it turn out that water is”? This sort of “turns-out” conditional is closely related to an indicative conditional, and like an indicative conditional is most naturally interpreted in epistemic terms (see the discussion in Section 3.3 of this paper). Chalmers (1995) suggests that this sort of conditional is superior in a way to “If W is actual, what would the concept refer to?”, since it makes clear that the concept is not required to be present in the world. Chalmers (1998) suggests that indicative conditionals can be used to evaluate primary intensions. My current view is that indicative conditionals provide a good heuristic for evaluating primary intensions, but should not be taken as definitional.

T in that world, the A-intension of T. In the second case, we are considering, for each world W , what T applies to in W given whatever world is in fact the actual world, and so we are, for all worlds except the actual world, considering the extension of T in a counterfactual world. We can call this the C-extension of T in W —‘C’ for counterfactual—and call the function assigning to each world the C-extension of T in that world, the C-intension of T.

Here, the talk of “hypotheses about which world is the actual world”, and “given or under the supposition that W is the actual world”, strongly suggests that we are thinking about these worlds as a sort of epistemic possibility. One might think for a moment that talk of “what the term applies to under various hypotheses” suggests something contextual, but on reflection there is no more reason why that should be the case here than for the corresponding usage about counterfactual worlds.

Jackson does not say much more about evaluating A-intensions than this. He does say one thing that might suggest a contextual element: he says that the A-proposition (A-intension for a sentence) of ‘Some water is H_2O ’ is contingent, because the sentence is “epistemically possible in the following sense: consistent with what is required to understand it, the sentence might have expressed something both false and discoverable to be false”. The claim about what the “sentence might have expressed” strongly suggests something contextual: together with the talk of understanding, it may suggest a sort of cognitive contextual intension. But this locution is not used elsewhere.

One possibility (suggested by conversation with Frank Jackson) is that Jackson’s use of the two-dimensional framework rests on a prior commitment to descriptivism. Jackson has argued elsewhere (1998b) that proper names are equivalent to certain rigidified descriptions. If so, then the rigidified description determines a 1-intension (picking out whatever satisfies the unrigidified description in a world) and a 2-intension (picking out whatever actually satisfies the description in all worlds). These intensions will behave just like the name’s epistemic and subjunctive intension. The main difference is that here the framework is used to analyze an independently established quasi-Fregean aspect of content, rather than to independently ground such an aspect of content. This understanding of the framework might be seen as a semantic understanding akin to Evans’, resting on a prior notion of content, although Jackson’s conception of the relevant sort of content differs from Evans’.

In any case, most of Jackson’s discussion is reasonably consistent with an epistemic understanding of A-intensions. Further, the purposes to which he puts the framework strongly suggest constitutive ties with apriority and the epistemic domains, and epistemic intensions work best for these purposes. If so, then A-intensions are arguably best interpreted as epistemic intensions.

5.7 Kripke’s epistemic duplicates

Although the work of Kripke (1980) provided the impetus for many of the two-dimensional approaches in the literature, Kripke does not embrace a two-dimensional approach himself. There are numerous remarks that suggest a tacit element of two-dimensional thinking: for example, frequent remarks of the form “*Given* that such-and-such is the case (empirically), such-and-such is necessary.” But there is little that formally and explicitly suggests such an approach. While it is possible to analyze many

of Kripke's epistemic claims using possible worlds, Kripke himself generally stays away from this sort of analysis.

There is one exception, however. Kripke notes that in cases where P is the negation of an a posteriori necessary statement, there is some intuition that "it might have turned out that P ", even though P is strictly speaking impossible. For example, when a table is made of wood, there is an intuition that the table "might have turned out to be made of ice", even though that is impossible. Kripke denies the modal claim involving "might have turned out" is false in these cases, but he wants to explain it away. Similarly, Kripke notes that for statements such as 'heat is the motion of molecules', there is a sense of "apparent contingency", even though the statement is strictly necessary. Again, Kripke wants to explain this sense of contingency away.

Kripke suggests the following strategy. In these cases, although a statement is necessary, we can say that *under appropriate qualitatively identical evidential situations, an appropriate qualitatively identical statement might have been false*. And he suggests this explains the sense of apparent contingency.

What, then, does the intuition that the table might have turned out to be made of ice . . . amount to? I think it means simply that there might have been *a table* looking and feeling just like this one and placed in this very position in the room, which was in fact made of ice. (1980, 142)

He applies a similar strategy to 'heat is the motion of molecules', and other cases. The general principle is that when there is an intuition of apparent contingency associated with a necessary truth P , there is a qualitatively identical contingent truth P^* , such that P^* might have been false in an evidential situation qualitatively identical to the original situation. Where P is 'heat is the motion of molecules', P^* might be 'heat sensations are caused by the motion of molecules'.

This apparently innocuous principle packs considerable power: it enables us to reason from epistemic premises to modal conclusions. When P is "apparently contingent", or such that it seems that "it might have turned out that P ", P has a distinctive *epistemic* status: to a first approximation, these claims come to the claim that P is not ruled out a priori. But the conclusion here is a *modal* one: that a certain state of affairs (involving a subject, evidence, and a statement) might really have obtained. In effect, Kripke reasons from a premise about the epistemic status of a statement to a conclusion about the possible truth of a statement token that shares a type with the original statement.

This reasoning can be modeled using the two-dimensional framework, understood contextually. We might say that the *evidential contextual intension* of a given statement is a function that is defined at centered worlds in which there is a subject with qualitatively identical evidence, uttering a qualitatively identical statement, and that returns the truth-value of that statement. Then Kripke is in effect suggesting that when a statement is "apparently contingent", its evidential contextual intension is contingent.

The intension in question is not fully defined, as Kripke does not define what it is for evidential situations or statements to be qualitatively identical. But it is natural to suggest that two evidential situations are identical when they are phenomenologically

equivalent: that is, when what it is like to be in the first situation is the same as what it is like to be in the second. As for qualitatively identical statements: one might first suggest that this occurs when they have similar descriptive or Fregean content, but that suggestion might be inappropriate in the current context. An alternative suggestion is that two statements are qualitatively identical when they (or corresponding thoughts) play similar cognitive roles for the subject. This goes beyond Kripke and is somewhat loose, but it seems at least compatible with his discussion.

Reconstructed this way, Kripke's principle takes on a familiar shape. In effect, the claim is that when a statement is apparently contingent, it has a contingent evidential contextual intension. If one substitutes aposteriority for apparent contingency and rearranges a little, one gets a familiar-looking result: if a statement has a necessary evidential contextual intension, it is a priori.

So Kripke's principle here suggests something in the vicinity of the Core Thesis for an evidential contextual intension. This should raise alarm bells. We have already seen that the Core Thesis appears to be false for any sort of contextual intension. Applying the sort of reasoning from our earlier discussion, one can straightforwardly come up with a counterexample. 'I have such-and-such evidence' is one example. For a more interesting example, let 'Bill' be a name that rigidly designates the phenomenological quality instantiated at the center of my visual field. Let us say that that quality (for me now) is phenomenal blueness. Then 'Bill is phenomenal blueness' is plausibly a posteriori, but it has a necessary evidential contextual intension.

This point is not simply an artifact of our reconstruction; it applies equally to Kripke's original claim. 'Bill is phenomenal blueness' is apparently contingent in the same sort of way as paradigmatic apparently contingent statements. We have an intuition that it might have turned out that Bill was not phenomenal blueness—it might have been that Bill was phenomenal redness, for example. This intuition seems to be on a par with our intuitions about the table, heat, and so on. But there is no qualitatively identical evidential situation in which a qualitatively identical statement would have been false. So the Kripkean reasoning is invalid in this case.

Perhaps Kripke could shrug this off and say that the reasoning was never intended to apply in *all* cases. But this reply would have a significant cost, since the reasoning is crucial to Kripke's argument against mind–body identity theories. We have a strong intuition of apparent contingency associated with 'pain is C-fibers'; and Kripke argues that this cannot be explained away by the claim that in a qualitatively identical evidential situation, a qualitatively identical statement would have been false:

To be in the same epistemic situation that would obtain if one had a pain *is* to have a pain; to be in the same epistemic situation that would obtain in the absence of pain *is* not to have a pain. The apparent contingency of the connection between the physical state and the corresponding brain state thus cannot be explained by some sort of qualitative analog as in the case of heat. (1980, 152)

Kripke uses this point to argue that as the apparent contingency cannot be explained away, the claim that 'pain is C-fibers' is not necessary at all. It follows that the claim is not true, since if it is true, it must be necessary. But it is clear that everything in the quoted passage applies equally to 'Bill'. To be in the same epistemic situation that

would obtain if one had Bill *is* to have Bill; and so on. So the ‘Bill’ case equally cannot be explained by Kripke’s paradigm. But here, one *cannot* reason from failure to satisfy Kripke’s paradigm to contingency and thus to falsity: ‘Bill is phenomenal blueness’ is clearly true and necessary. So by parity, one cannot apply this reasoning in the case of ‘pain’. Insofar as Kripke’s argument against the mind-body identity theory relies on this reason, it appears that the argument is unsound.

The moral here seems to me that there is nothing special about one’s *evidence* in diagnosing apparent contingency. This suggests a natural response: one might *weaken* Kripke’s principle, holding simply that when a statement is apparently contingent, an appropriate qualitatively identical statement might have been false. This claim would cover the standard Kripkean cases, and would also encompass cases involving contingency of evidence itself, such as that of ‘Bill’. One might think that to do this would be to kill off the argument against the identity theory, which relied crucially on qualitatively identical evidence. But there is another strand in Kripke’s argument that might still apply.

Immediately after noting the considerations above, Kripke says:

The same point can be made in terms of the notion of what picks out the reference of a rigid designator. In the case of the identity of heat with molecular motion the important consideration was that although ‘heat’ is a rigid designator, the reference of that designator was determined by an accidental property of the referent, namely the property of producing in us sensation *S*. It is thus possible that a phenomenon should have been rigidly designated in the same way as a phenomenon of heat, with its reference also picked out by means of the sensation *S*, without that phenomenon being heat and therefore without its being molecular motion. Pain, on the other hand, is not picked out by one of its accidental properties; rather it is picked out by the property of being pain itself, by its immediate phenomenological quality. Thus pain, unlike heat, is not only rigidly designated by ‘pain’ but the reference of the designator is determined by an essential property of the referent.

Kripke says that this is the “same point”, but in fact it is not. A close examination suggests that this point has nothing to do with evidence: it does not rely on the notion of a statement in a qualitatively identical evidential situation, but rather on that of a designator with the same manner of reference-determination. So we can apply the weakened principle suggested above, where a “qualitatively identical statement” is understood, perhaps, as a statement with the same reference-fixing intensions. Under this paradigm, the apparent contingency of ‘Bill is phenomenal blue’ *can* be explained away, as it is possible for a statement associated with the same reference-fixing intentions—to refer to the quality in the center of one’s visual field—to be false, for example, in a case where the quality is phenomenal red. But the case of ‘pain’ *cannot* be explained away in these terms. So the argument against the identity theory would now seem to go through.

Nevertheless, more problems immediately arise. The reasoning here in effect invokes the claim that an apparently contingent statement has a contingent *qualitative contextual intension*, where this sort of intension is defined at worlds centered on a statement qualitatively identical to the original, returning the referent of that token. For familiar reasons, there are counterexamples to this claim. One such is ‘A sentence token exists’. Or let ‘L’ be a descriptive name that rigidly designates

the number of actual (spoken or written) languages. Then ‘ $L > 0$ ’ is apparently contingent: it seems that it might have turned out that L was zero, and so on. But ‘ $L > 0$ ’ has a necessary fixing contextual intension. So the principle is false. Further, it is easy to see that even the weakened Kripkean reasoning suggested above does not apply in the case of ‘ $L > 0$ ’: it is not possible for a qualitatively identical statement to be true, but the statement is necessary all the same. So even the weakened reasoning, from failure to satisfy the weaker paradigm to contingency, is invalid. So although Kripke’s second argument against the identity theory is different from the first argument, it is also invalid.³⁶

I think that the moral in both cases is that Kripke’s diagnosis of intuitions about apparent contingency is incorrect: they do not turn essentially on intuitions about qualitatively identical evidential statements, and they do not turn essentially on intuitions about qualitatively identical statements.³⁷ Rather, they turn on intuitions about the direct evaluation of epistemic possibilities. And we have seen that this sort of epistemic evaluation does not turn essentially on the status of possible tokens. This suggests in turn that to reason about apparent contingency, then instead of the Kripkean principles above, which tacitly involve a contextual understanding, one should appeal to principles involving an *epistemic* understanding: for example, that when a statement is apparently contingent, there is some scenario (considered as actual) in which it is false. Or, if we understand scenarios as centered worlds: that when a statement is apparently contingent, there is some centered world in which the statement’s epistemic intension is true. I think that once we understand things this way, an argument against the mind–body identity theory of a sort analogous to Kripke’s has a chance of succeeding. But I have written about that elsewhere.³⁸

In any case, we can see that while Kripke does not explicitly endorse a two-dimensional approach, these issues are quite close to the surface in his discussion. One can even see here an instance of a familiar two-dimensionalist pattern: trying to capture epistemic phenomena with a contextual approach, coming close, but not

³⁶ Bealer (1996) develops Kripke’s proposal using his notion of a “semantic stable” expression (one such that necessarily, in any language group in an epistemic situation qualitatively identical to ours, the expression would mean the same thing). Bealer suggests that semantically stable expressions are invulnerable to “scientific essentialism”. It is not clear exactly what is required for a “qualitatively identical epistemic situation” for a language group, but it is clear that however this notion is understood, terms such as ‘ L ’ above will be semantically stable, as will rigid designators for aspects of a group’s epistemic situation. So there will be a posteriori necessities involving semantically stable expressions. As in Kripke’s discussion and in other cases, a broadly contextual notion (semantic stability) serves as an imperfect substitute for a broadly epistemic notion (semantic neutrality).

³⁷ There is arguably a strand in the passage above that does not turn directly on qualitatively identical statements. This is the claim that while heat is picked out by a property it has accidentally, pain is picked out by a property it has essentially. This suggests that Kripke’s “same point” may come down to three different points, with potentially three different arguments in the background. I think that the third argument has the most chance of success. But I think that to make the argument work in the general case, one has to adopt something like the (highly non-Kripkean) framework of epistemic intensions.

³⁸ See, e.g., Chalmers (2002a) (2003).

quite succeeding. I think that once again, the moral is that the epistemic framework is most fundamental for these purposes.

5.8 Other approaches

There are a number of other approaches that either fit within a two-dimensional framework or have something in common with these ideas. I will discuss some of these briefly.

Tichy (1984) suggests that there are two propositions corresponding to any given statement: the proposition expressed and the proposition associated. The proposition expressed by ‘Phosphorus is hot’ is the proposition that Venus is hot. This proposition behaves like a 2-intension. The proposition associated with ‘Phosphorus is hot’ is the proposition that the sentence ‘Phosphorus is hot’ says in English that Phosphorus is hot. This proposition behaves like a (modified) linguistic contextual intension: it is true at those worlds corresponding to a centered world where the sentence’s linguistic contextual intension is true, and false at all other worlds.

Tichy uses this distinction to argue that there are no necessary a posteriori propositions. He suggests that for truths such as ‘Hesperus is Phosphorus’, the (trivial) proposition expressed will be necessary and a priori, while the proposition associated will be contingent and a posteriori. Whether the latter is so depends on how sentences and languages are individuated: if ‘Hesperus’ picks out Venus essentially, then the proposition associated will be necessary. Even if this possibility is set aside (as Tichy does), there will clearly be cases (e.g. ‘A sentence token exists’) of an intuitively necessary a posteriori statement with a necessary associated proposition.³⁹ Tichy nowhere says that *all* intuitively a posteriori statements have an a posteriori associated proposition, and his claim that there are no necessary a posteriori propositions is arguably independent of this claim.⁴⁰ But this suggests at least that Tichy’s dual-proposition account of the Kripkean phenomena does not quite get to the roots of the phenomena.

Also relevant are some proposals for assigning content to *thought* rather than to language.⁴¹ One such proposal is Lewis’s (1979; 1986; 1994) suggestion that a subject’s system of belief can be modeled by the self-attribution of a property, or equivalently,

³⁹ Tichy sets this possibility aside on the (highly arguable) grounds that Kripke is committed to denying that it is necessary that ‘Hesperus’ in English picks out Venus, since on Kripke’s view all that is essential to ‘Hesperus’ in English are linguistic conventions involving reference-fixing.

⁴⁰ The claim that there are no necessary a posteriori propositions is sometimes regarded (e.g. Byrne 1999; Soames 2004) as a central aspect of the two-dimensional framework. I think it should be regarded as a further inessential claim. If one stipulates that propositions are sets of possible worlds, such a claim may be reasonable. On my own view, it is better to regard propositions as the contents of sentences (leaving open their nature) and allow that the apriority and the necessity of a statement correspond respectively to the apriority and the necessity of the proposition it expresses. If so, some propositions will be both necessary and a posteriori, and propositions will presumably themselves have a sort of two-dimensional modal structure (although they may not be reducible to this structure).

⁴¹ Apart from those mentioned in the text, some other relevant proposals concerning the narrow content of thought include the “immediate object of belief” of Brown (1986) (roughly: those propositions believed by all intrinsic duplicates), the “notional worlds” of Dennett (1982) (roughly:

by a class of possible individuals (the subject's "doxastic alternatives"), or by a class of centered worlds. Lewis (1979) assigns this sort of content to specific beliefs, while Lewis (1986; 1994) suggests a content of this sort can be used to represent a subject's total belief system, and then in turn to show how the subject satisfies various belief ascriptions. It is clear from Lewis's discussion that this set of centered worlds behaves at least something like a 1-intension: a subject who believes *I am hungry* will have only worlds centered on hungry subjects in the set; a subject who believes *water is XYZ* will have Twin Earth centered worlds in the set, and so on.

Lewis's proposal is hard to classify directly in the present system, since he does not say much about how the relevant set of worlds (or the relevant property, or the relevant class of alternatives) is defined, apart from saying that it is determined by the subject's behavior and functional organization by the principles of belief-desire psychology (Lewis 1986, 36–40).⁴² Lewis says nothing to suggest that a token of specific mental or linguistic states is required at the center of the relevant worlds, however, so there is reason to think he is not invoking a contextual understanding. And Lewis's centered worlds can naturally be seen as representing a sort of epistemic possibility for the subject. So Lewis's discussion seems at least consistent with an epistemic understanding, on which the relevant set of centered worlds is a sort of epistemic intension of a subject's total belief state, consisting to a first approximation of those scenarios that verify all of a subject's beliefs.

White (1982) sets out a multi-dimensional proposal for understanding the narrow (internally determined) content of thought and language. To simplify a little, White defines the *partial* character of a word as a function from "contexts of acquisition" (worlds centered on a functional duplicate of the original subject) to the Kaplanian character of the corresponding word in that context. This is actually a three-dimensional function (a function from centered worlds to contexts to worlds to extensions), but one can diagonalize it twice to yield a one-dimensional function: in effect, a functional contextual intension, defined at worlds centered on a functional duplicate of the original subject, returning the extension of the relevant token.

The resulting functional contextual intension will be internally determined by definition, and will be a reasonably good approximation to a Fregean semantic value. It will give anomalous results in a few cases: for example, cognitively significant claims concerning language ('A sentence token exists') or a subject's functional organization ('I have computational structure C') may have a necessary functional contextual intension, and so will have a relatively trivial partial character. But if we assume that the epistemic intensions of a subject's tokens are determined by the

the set of environments in which the organism as currently constituted will flourish), the "realization conditions" of Loar (1988) (roughly: the set of worlds in which a given thought would be true if it were not a misconception), and the "notional content" of White (1991) (roughly: the class of worlds for which a subject's actions are optimal). All of these have some similarity to 1-intensions, characterized in a broadly contextual way.

⁴² For example, the principle that a subject's actions are such that if the subject's beliefs were true, they would tend to fulfill the subject's desires. One can see this principle as mutually constraining a subject's classes of doxastic alternatives and desire alternatives.

subject's functional state, then the diagonalized partial character will give a reasonable approximation of an epistemic intension.

Fodor (1987) gives a related proposal for understanding narrow content. He suggests that the narrow content of a thought is a function from contexts to truth-conditions, where contexts appear to behave like worlds centered on the thought, and truth-conditions behave like 2-intensions. Like White's partial character, Fodor's function is not directly truth-evaluable (this led Fodor to eventually reject the proposal on the grounds that it is not really a sort of content), but as usual one can diagonalize it to yield a truth-evaluable content. The result is something like a conceptual contextual intension, mapping worlds centered on a token of the relevant concept or thought to that token's extension.

As with conceptual (and linguistic) contextual intensions in general, the behavior of Fodor's function will depend on how concepts and thoughts are individuated, in order to know which centered worlds are relevant. If they are syntactic mental types, then one has a sort of orthographic intension, which has uninteresting content. If they are semantic types, then it is unclear how one can specify the relevant semantic value in a noncircular way. These points (and many others) are developed by Block (1991), in a thorough critique of proposals of this sort in accounting for narrow content.

Block makes a point worth noting here: he suggests that proposals involving a mapping from contexts "often seem to engender a *cognitive illusion* to the effect that we know *what the proposed mapping is*". I diagnose things differently. Our judgments about the mapping are grounded in perfectly reasonable intuitions about the evaluation of our terms in epistemic possibilities. The illusion on the part of these theorists is not that they grasp the mapping. Rather the illusion is that the mapping is grounded in context-dependence. Once we recognize the epistemic roots of the mapping, the problems go away.

Whether or not this diagnosis is correct, it seems fair to say that in many of the cases we have discussed, various contextual two-dimensional notions are of interest largely to the extent that they approximate epistemic two-dimensional notions. One might regard contextual notions positively as an inexpensive substitute for the epistemic notions, yielding many of the benefits without as many of the costs. Or one might regard them negatively, as a "distractor" from the more important epistemic notions that can lead to confusion because of their surface similarity. I suspect that there is something to each of these attitudes. But in any case, it seems clear that the epistemic understanding yields semantic notions with the deepest connections to the cognitive, rational, and epistemic domains.

6. Conclusion

One might see the project of this paper as an attempt to vindicate Carnap's vindication of Frege, although a number of aspects of the approach diverge from both. I have followed Carnap's lead in using a modal analysis to construct a semantic value that is constitutively tied to the epistemic domain. If the project is successful, it yields an aspect of meaning that can serve as the third vertex in the golden triangle, by virtue of its constitutive connections to reason and modality.

I have taken epistemic notions as primitive in this paper, and have constructed semantic notions from there. But it should be stressed that proceeding in this way does not entail that one of the vertices of the golden triangle is more fundamental than the others. I have appealed to this order of explication because we have a relatively pretheoretical grasp on the relevant epistemic notions, whereas we do not have the same pretheoretical grasp on the corresponding semantic notions. But this no more entails that reason is prior to meaning than our pretheoretical grasp of the macroscopic world entails that it is prior to the microscopic world.

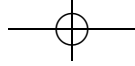
The framework in this paper is compatible with a variety of views about the underlying relations between epistemic and semantic notions. It could be that epistemic properties are grounded in semantic properties (so that thoughts, for example, stand in epistemic relations by virtue of their epistemic content), or it could be that semantic properties are grounded in epistemic properties (so that thoughts have their semantic content in virtue of their epistemic role). Or it could be that, as I am inclined to suspect, neither is more basic than the other. In any case, we can expect that two-dimensional semantics will be helpful in analyzing the complex connections between meaning, reason, and modality.

References

- Bealer, G. (1996). A priori knowledge and the scope of philosophy. *Philosophical Studies* 81: 121–42.
- Block, N. (1991). What narrow content is not. In B. Loewer and G. Rey, eds., *Meaning in Mind: Fodor and his Critics*. Blackwell.
- Block, N. and Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108: 1–46.
- Braun, D. (1996). Demonstratives and their linguistic meanings. *Noûs* 30: 145–73.
- Brown, C. (1986). What is a belief state? *Midwest Studies in Philosophy* 10: 357–78.
- Byrne, A. (1999). Cosmic hermeneutics. *Philosophical Perspectives* 13: 347–83.
- and Pryor, J. (2006). Bad intensions. In M. Garcia-Carpintero and J. Macia, eds, *Two-Dimensional Semantics: Foundations and Applications*. Oxford University Press.
- Carnap, R. (1947). *Meaning and Necessity*. University of Chicago Press.
- (1963). Replies and systematic expositions. In P. Schilpp, ed., *The Philosophy of Rudolf Carnap*. Open Court.
- Chalmers, D. J. (1995). The components of content (1995 version). Manuscript (revised as Chalmers 2002c). [consc.net/papers/content95.html]
- (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- (1998). The tyranny of the subjunctive. [consc.net/papers/tyranny.html]
- (1999). Materialism and the metaphysics of modality. *Philosophy and Phenomenological Research*. [consc.net/papers/modality.html]
- (2002a). Does conceivability entail possibility? In T. Gendler and J. Hawthorne, eds., *Conceivability and Possibility*. Oxford University Press. [consc.net/papers/conceivability.html]
- (2002b). On sense and intension. [consc.net/papers/intension.html]
- (2002c). The components of content (revised version). In *Philosophy of Mind: Classical and Contemporary Readings*. [consc.net/papers/content.html]

- (2003). Consciousness and its place in nature. In S. Stich and F. Warfield, eds., *Blackwell Guide to Philosophy of Mind*. Blackwell. [consc.net/papers/nature.html]
- (forthcoming). The nature of epistemic space. [consc.net/papers/espace.html]
- and Jackson, F. (2001). Conceptual analysis and reductive explanation. *Philosophical Review* 110: 315–61. [consc.net/papers/analysis.html]
- Crossley, J. N. and Humberstone, I. L. (1977). The logic of ‘actually’. *Reports on Mathematical Logic* 8: 11–29.
- Davies, M. and Humberstone, I. L. (1981). Two notions of necessity. *Philosophical Studies* 58: 1–30.
- (2004). Reference, contingency, and the two-dimensional framework. *Philosophical Studies* 118: 83–131.
- Dennett, D. C. (1982). Beyond belief. In A. Woodfield, ed., *Thought and Object*. Oxford University Press.
- Evans, G. (1979). Reference and contingency. *The Monist* 62: 161–89.
- (1982). *The Varieties of Reference*. Oxford University Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Frege, G. (1892). Über Sinn und Bedeutung. Translated in P. Geach and M. Black, eds., *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 1952.
- Garcia-Carpintero, M. (1998). Indexicals as token-reflexives. *Mind* 107: 529–63.
- Jackson F. (1998a). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford University Press.
- Jackson F. (1998b). Reference and description revisited. *Philosophical Perspectives* 12: 201–18.
- Kaplan, D. (1978). Dthat. In P. Cole, ed., *Syntax and Semantics*. New York: Academic Press.
- (1989). Demonstratives. In J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*. Oxford: Oxford University Press.
- Kripke, S. A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, C. I. (1944). The modes of meaning. *Philosophy and Phenomenological Research* 4: 236–50.
- Lewis, D. (1979). Attitudes *de dicto* and *de se*. *Philosophical Review* 88: 513–43.
- (1986). *On the Plurality of Worlds*. Blackwell.
- (1994). Reduction of mind. In S. Guttenplan, ed., *Companion to the Philosophy of Mind*. Oxford: Blackwell.
- Loar, B. (1988). Social content and psychological content. In R. H. Grimm and D. D. Merrill, eds., *Contents of Thought*. University of Arizona Press.
- Putnam, H. (1975). The meaning of ‘meaning’. In K. Gunderson, ed., *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press.
- Salmon, N. (1986). *Frege’s Puzzle*. MIT Press.
- Schiffer, S. (1990). The mode-of-presentation problem. In C.A. Anderson and J. Owens, eds., *Propositional Attitudes: The Role of Content in Logic, Language, and Mind*. Stanford: CSLI Press.
- Schroeter, L. (2004). The rationalist foundations of Chalmers’ two-dimensional semantics. *Philosophical Studies*, 118: 227–55.
- Soames, S. (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press.
- (2004). *Reference and Description*. Princeton University Press.
- Stalnaker, R. (1978). Assertion. In P. Cole, ed., *Syntax and Semantics: Pragmatics, Vol. 9*. New York: Academic Press.

- Stalnaker, R. (1990). Narrow content. In C. A. Anderson and J. Owens, eds., *Propositional Attitudes*. Stanford: Center for the Study of Language and Information.
- (1991). How to do semantics for the language of thought. In B. Loewer and G. Rey, eds., *Meaning in Mind: Fodor and his Critics*. Oxford: Blackwell.
- (1999). *Content and Context* (foreword). Oxford University Press.
- (2001). On considering a possible world as actual. *Aristotelian Society Supplementary Volume 75*: 141–56.
- Tichy, P. (1984). Kripke on necessity a posteriori. *Philosophical Studies* 43: 225–41.
- Weatherson, B. (2001). Indicatives and subjunctives. *Philosophical Quarterly* 51: 200–16.
- White, S. (1982). Partial character and the language of thought. *Pacific Philosophical Quarterly* 63: 347–65.
- (1991). Narrow content and narrow interpretation. In *The Unity of the Self*. MIT Press.
- Yablo, S. (1999). Concepts and consciousness. *Philosophy and Phenomenological Research* 59: 455–63.
- (2002). Coulda, woulda, shoulda. In T. Gendler and J. Hawthorne, eds., *Conceivability and Possibility*. Oxford University Press.



Queries in Chapter 4

Q1. Please clarify whether it is 'determines' or 'determined'

