13

Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics

Robert Stalnaker

Beginning more than twenty-five years ago, two-dimensional modal semantics has been applied to the interpretation of speech and thought in a number of different ways. More recently, the two-dimensional semantic apparatus has been deployed (by Frank Jackson and David Chalmers, among others) in philosophical arguments about the role of conceptual analysis and reductive explanation, and in the clarification of the notions of a priori knowledge and truth. Different philosophers have applied the apparatus to examples in the same way, but have given contrasting interpretations of those applications, and have drawn different philosophical conclusions about their significance. My intention in this paper is to try to clarify the question of interpretation, and the contrasting ways of answering it. I will defend one kind of interpretation, but my main aim is to draw a contrast between two very different ways of thinking about intentionality that I think are implicit in the different ways of understanding the framework. I will begin with a look back at my own early attempts to deploy this framework, at the way I understood the problem to which it was a response, and the way I was interpreting it. Second, I will sketch a contrasting interpretation, which generalizes David Kaplan's semantics for context-dependent expressions. Third, I will look at David Chalmers's different way of contrasting the different interpretations, and at the account that he defends. Finally, I will describe and criticize the internalist approach to intentionality that I think is required by the alternative interpretations of the two-dimensional semantic framework.

The occasion for this paper was the conference on two-dimensional semantics at the ANU in February, 2002, but it was written after the conference. The paper that was the basis of both the talk I gave there and at the Barcelona conference on two-dimensional semantics the previous June was Stalnaker (2004b), which discusses some similar themes. I am grateful to the participants in both of these conference for the high level of discussion that helped me to get clearer about the different ways of understanding and using the two-dimensional semantic framework. Thanks particularly to Philip Pettit for his comments on my paper at the ANU conference.

1. The "Assertion" Story

In my paper "Assertion,"[1] I began with a simple abstract but intuitive picture of what it is to say or think something. According to this picture, a representation is a way of distinguishing between possibilities. As Frank Jackson puts it, "to represent is to make a division into what accords with, and what does not accord with, how things are being represented as being."[2] A *proposition*—the content of a representation—can be modeled, according to this picture, by the set of possible situations that *are* the way the world is being said to be. These possible situations are the *truth conditions* of the representation—the conditions that would have to obtain for the proposition to be true. An *assertion* can be understood as a proposal to exclude from the possible situations compatible with the context those in which the proposition asserted is false.

Two-dimensional semantics came into the picture as a response to a problem that this conception of representation brought into focus. The problem is this: with some statements, there is a tension between global intuitions about the information that the statement conveys, as represented by the possibilities that the statement seems to exclude, and what semantic theories that are otherwise well motivated say about the truth conditions of the statement. The tension is most acute with statements that seem to be informative (and so to exclude possibilities), but also necessarily true (and so to exclude no possibilities). The clearest cases of this kind are the necessary a posteriori statements that Saul Kripke brought to our attention in *Naming and Necessity*. It seems intuitively clear, for example, that identity statements with proper names such as "Hesperus is Phosphorus" and statements about the nature of natural substances such as "water is a compound of hydrogen and oxygen" convey substantive information about the world, but it also seems that such statements say something that could not possibly be false.

The first step to get clear about the problem is to ask what information it is that statements of this kind seem to be conveying—what kinds of possible situations the statement seems to be excluding. If Daniels has reason to tell O'Leary that Hesperus is Phosphorus, it must be that he thinks that O'Leary doesn't know it already—that certain possibilities need to be excluded, and that saying that Hesperus is Phosphorus will succeed in excluding them. If O'Leary thinks that Hesperus might not be Phosphorus, what does he think the world might be like? If we can give a plausible answer to this question, the second step is to ask how it is that the statement "Hesperus is Phosphorus," which our semantic theory tells us is a necessary truth, manages to do the job of excluding those possibilities.

If O'Leary doesn't know that Hesperus is Phosphorus, then it seems reasonable to think that possible worlds that satisfy the following description are compatible with his knowledge: There is a heavenly body that appears in the evening where Venus in fact appears, and a distinct heavenly body that appears in the morning where in fact Venus appears. The first has come to be called "Hesperus" and the second has come

[1] Stalnaker (1978). [2] Jackson (2001: 617).

to be called "Phosphorus," so that what people would be saying in this kind of world if they were to say "Hesperus is Phosphorus" would be false.

The answer to this first question is much the same as the answer that a Fregean might give. The Fregean might describe such a possible world by saying that it is one in which distinct objects are presented by two of the modes of presentation which in the actual world present the same thing—Venus. On the Fregean view, the thought expressed, in the actual world, by the statement "Hesperus is Phosphorus" is a contingent proposition that is false in the possible world described, and it is the very same proposition as the one that would be expressed by someone in that counterfactual possible world who said "Hesperus is Phosphorus" there. But my aim was to reconcile the fact that the statement is informative with a direct reference account of the semantics for names according to which the possible world we have described is not one in which Hesperus is distinct from Phosphorus, since there is no such world. On this account of the semantics of names, the possible world we have described is one that differs from the actual world not only in its astronomical facts, but also in its semantic facts: it is a world in which the expression "Hesperus is Phosphorus" expresses a different proposition. The direct reference theory of names gave an externalist account of the facts that determine reference: statements containing names have the content that they have because of the way speakers using them are causally connected with things in the world. Consequently, in possible worlds where the astronomical facts are different, the semantic values of names referring to astronomical bodies may be different, and so different propositions may be expressed with those names.

Externalist accounts of names, and more generally of propositional content, focused attention on the fact that it is a matter of contingent fact that the words we use have the meaning and content that they have. This is of course not a fact that is restricted to externalist theories—any theory of speech and thought must give an account of the facts in virtue of which marks, sound patterns, and states of the brains of people and animals have the representational properties that they have. But externalist theories of intentionality make this fact particularly salient. The two-dimensional framework was deployed, in the first instance, as a piece of descriptive apparatus for representing the way that semantic values depend on the facts. We need two dimensions since we start with the fact that the truth value of a proposition (at least a contingent proposition) depends on the facts. But since the identity of the proposition expressed in a given utterance also depends on the facts, the truth value of the utterance will depend on the facts in two different ways: first, the facts determine what is said; second, the facts determine whether what is said is true. We can represent the two different roles of the facts in determining a truth value with what I called a *propositional concept*: a function from possible worlds to propositions, where a proposition is a function from possible worlds to truth values, or equivalently, a function from a pair of possible worlds to a truth value.

This descriptive apparatus is of interest independently of our particular problem since it is apt for representing the interaction of speakers and addressees in a conversation. The information that speakers take for granted when they speak includes a mix of semantic information and information about the subject matter of a conversation.

Speakers make assumptions about what their addressees know or believe about what they are talking about as well as about what their words mean and what the relevant contextual parameters are relative to which their words will be interpreted. These assumptions will influence what they choose to say, and how they choose to say it. Since the speaker will normally presuppose that he is speaking (and that the addressee knows this), the possible worlds compatible with what is presupposed will be possible worlds in which the utterance event in question takes place, and in which it has a meaning and a content that may be different from the meaning and content that the utterance has in the actual world. In a case where the addressee knows that the utterance event has taken place, but mishears or misinterprets it, the content of what is said, and sometimes even the words that are uttered, will be different in possible worlds compatible with what the addressee believes than they are in the actual world. Cases where someone is ignorant or mistaken about the content of an utterance will be cases where the propositional concept for the utterance, relative to the relevant possible worlds, will be a variable one: the function will determine different propositions relative to different possible worlds. An example: The policeman has stopped a driver, and after examining his driver's license, says, "your license says you need corrective lenses, but you're not wearing your glasses." The driver responds, "I've got contacts." The policeman replies, "I don't care who you know, you have to wear your glasses." The proposition expressed in the actual world by the driver's utterance "I've got contacts" is different from the proposition expressed by that utterance in the possible worlds compatible with the policeman's beliefs, and this can be represented with a variable propositional concept, defined on the relevant possible worlds.

The descriptive apparatus is relevant to our problem since the possible worlds that the problematic statement, "Hesperus is Phosphorus," seems to be excluding are worlds that differ from the actual world both in astronomical and in semantic facts, and so it seemed that a two-dimensional representation of this kind might help us to answer the second question about such problematic statements: how is it that a necessarily true statement could be used to convey contingent information? To approach this question, we consider an intuitively natural context for our example in which Daniels tells O'Leary that Hesperus is Phosphorus, and construct a propositional concept for it, on the assumption that the semantics for names is the way Kripke argued that it is: names like "Hesperus" and "Phosphorus" are rigid designators—they denote the same thing in all possible worlds, and so the proposition actually expressed by "Hesperus is Phosphorus" is necessarily true. But if the facts that determine the reference of the names had been different—for example if the astronomical facts had been as O'Leary thinks they might in fact be, then an utterance of "Hesperus is Phosphorus" would instead have expressed the necessarily false proposition, since the two names would have denoted different things. So if *i* is the actual world, and *j* is the world described above that is the way O'Leary thinks the world might be, then the propositional concept for the utterance, relative to these two possible worlds, will be the one pictured in this matrix:

	i	*j*
i	T	T
j	F	F

Daniels's purpose in stating that Hesperus is Phosphorus is clearly to exclude worlds like *j*—to inform O'Leary that the actual world is not like that, while including worlds like *i*. Neither the necessarily true proposition expressed in the actual world nor the necessarily false proposition expressed in world *j* (according to the semantics we are assuming) accomplishes this, but the *diagonal* proposition—the one that for each world *x* is true in world *x* if and only if the proposition expressed in *x* is true in *x*, seems to be the proposition that does the right job.

The proposal made in "Assertion" was that in special cases, where there was a prima facie violation of certain conversational rules, utterances should be reinterpreted to express the diagonal proposition, rather than the proposition expressed according to the standard semantic rules. The proposal followed the pattern of interpretive reasoning that Paul Grice spelled out in his theory of conversation: Certain maxims of conversation are argued to be truisms required by the general purposes of rational discourse, and so to be common ground among the participants in a conversation. The presumption that such rules are being followed constrains the interpretation of what is said. It is presumed that hearers will try to find a way of understanding what is said that conforms to the maxims, and speakers may exploit this presumption by saying things that would be manifest violations of conversational rules if they were interpreted in a standard way, and so that will require reinterpretation. In the case of our problematic statements, the relevant maxim is that speakers presume that their addressees understand what they are saying. In terms of the two-dimensional apparatus, this presumption will be satisfied if and only if the propositional concept for the utterance is constant, relative to the possible worlds that are compatible with the context. Our problematic example, and all cases of necessary truths that would be informative (in the sense that the addressee does not already know that they are true) will be prima facie violations of this maxim, and so will require reinterpretation. Reinterpreting by taking the diagonal proposition to be the one the speaker intends to communicate brings the statement into conformity with the rule, and seems to give the intuitively correct result.

On this kind of account, diagonal propositions (corresponding to what Frank Jackson calls A-intensions) are derivative from the standard semantics, as it is in the actual world, and in the relevant alternative possible worlds. My project began from the assumption that the standard semantics (which determines what Jackson calls C-intensions) was essentially right—the project was to reconcile it with the fact that statements that were necessary, according to that semantics, could be used to communicate contingent information. Diagonalization was *re*interpretation—interpretation that was parasitic on the standard interpretation, which was assumed to give the right result for what was expressed and communicated in the normal case. If the standard semantics for names did not give the right result in the normal case—that is, if it were not right to say that normally the beliefs one expresses and the information one conveys when one uses proper names are singular propositions about the individual named—then that semantic account would not be defensible. The plausibility of this assumption turns on one's account of intentionality—of what makes it the case that our mental states have the representational properties that they have. It is disagreements about how the problem

of intentionality is to be solved that lie behind disputes about the semantics for names, and I will suggest that this is also the central issue that divides different interpreters of the two-dimensional semantic apparatus.

In several of the early papers in which I applied the diagonalization strategy,[3] I contrasted the use I was making of the two-dimensional apparatus with the use that David Kaplan made of it in his theory of indexicals and demonstratives.[4] Kaplan's theory is a descriptive semantics for a formal language containing context-dependent expressions such as personal and deictic pronouns, tenses, temporal and locative adverbs. In Kaplan's semantics, the meaning of a sentence type (which he called its *character*) is a function from context to content, where content (what is said) is the proposition expressed. So, for example, the meaning of the sentence "I am flying to Canberra tomorrow" is a function that takes a context of utterance in which **a** is the speaker and **t** is the time of utterance into the proposition that is true in possible worlds in which **a** flies to Canberra on the day following time **t**. The thought expressed in a use of that sentence is the proposition determined; the role of the context is to contribute to the means used to express it.

Kaplan's characters are abstract objects that are similar to propositional concepts, but they play a very different role in the explanation of speech, and were not designed to solve the problem to which the diagonalization strategy was applied. In Kaplan's semantics, the paradigm examples of sentences that express necessary a posteriori truths such as "Hesperus is Phosphorus" are not context-dependent, and so have constant character. If an analogue of the diagonal proposition were defined in terms of the character of such a sentence, it would be the same as the content expressed, a necessary truth, and so would not help to explain how such statements can convey contingent information. Kaplan's semantics will play a role in some applications of the descriptive apparatus, since features of context on which content depends are sometimes among the features that hearers may be ignorant or mistaken about, and the reinterpretation strategy may be required in such cases. The two theories are not competing theories for explaining the same phenomena, or competing interpretations of the abstract framework, but complementary theories that use formally similar tools to answer different questions.

One important difference between the two theories is the contrasting roles of the two-dimensional intensions (character, in Kaplan's semantics, propositional concepts in the assertion theory) in the explanation for the fact that an utterance has the content that it has. Suppose we ask why a certain utterance of the sentence "I am flying to Canberra tomorrow" expresses the proposition that is true if and only if RS flies to Canberra on February 19, 2002. The answer is, because the sentence has the character stated above, and the utterance in question was produced by RS on February 18, 2002. One might go on to ask the further question, what made that utterance an utterance of a sentence with that character, but even if there are further questions, the answer we gave is correct. Character precedes content in the order of explanation of the fact that the utterance has the content that it has. But the order is the reverse

[3] Stalnaker (1981b and 1987). [4] Kaplan (1989a).

in the case of the explanation of why an utterance conveys the information that a diagonal proposition represents. Why does "Hesperus is Phosphorus" (uttered in a particular context) convey the contingent information that the heavenly body that appears in the evening and is called "Hesperus" is distinct from the one that appears in the morning and is called "Phosphorus"? The answer begins with the fact that in a world of this kind that is compatible with the context, the semantics, as it is in that world, implies that the sentence expresses a necessary truth, whereas the semantic accounts that hold in other worlds compatible with the context imply that it expresses a necessary falsehood. We explain why the utterance determines the propositional concept that it determines in terms of the content that it has, or would normally have, according to the semantics of the relevant alternative possible worlds. Content (in the various alternative worlds) precedes propositional concept in the order of explanation. The second part of the explanation invokes reinterpretation by diagonalization, but since the diagonal proposition is determined by the propositional concept, the main work of explaining why the utterance conveys the particular content that it conveys is done when we have explained why the utterance determines the propositional concept that it determines. Again, we can ask a further question: what facts about these possible worlds make it the case that the semantics of those worlds determine that the utterance in question expresses the necessarily true, or necessarily false proposition, but even if there are further questions, the answer we gave gives a correct explanation, assuming we are right about the semantics for the sentence in the different possible worlds.

Just to see how the two theories can interact, involving both kinds of explanation, consider an example that involves both diagonalization and demonstratives: Pierre, in London, says "Londres est jolie, mais cette ville-ci n'est pas jolie." (London is pretty, but *this* city is not pretty.) Jacques responds, "Mais cette ville-ci *est* Londres." (But this city *is* London). Jacques's statement communicates to Pierre the contingent information that the world is not the way he thinks it is, not a world in which the city he calls "Londres" is distinct from the city he is currently in. Why does the utterance of this sentence convey this information? Because the semantics implies that this utterance token expresses a necessary truth in worlds in which the place of utterance is London and a necessary falsehood in worlds in which the place of utterance is a city different from London. Since worlds of both kinds are compatible with the context required to interpret Jacques's utterance, the utterance is reinterpreted to express the diagonal proposition, which is true in worlds of the first kind, and false in worlds of the second kind. But why does the semantics imply that this utterance (on the standard interpretation) expresses a necessary truth in worlds of the first kind and a necessary falsehood in worlds of the second kind? Because the semantics (which is common knowledge in the context, and so applies in all the relevant possible worlds) says that "cette ville-ci" is a rigid designator for the city which is the place of utterance, and that "Londres" is a rigid designator for London. So we explain the propositional concept determined by an utterance in terms of the content expressed by that utterance in different possible worlds and the content expressed by the utterance in the different possible worlds in terms of the character that the

sentence used to make the utterance has, and the context in which it is uttered, in those different possible worlds.

2. The Generalized Kaplan Interpretation

Although I have been arguing that the Kaplan semantics and the assertion theory are complementary theories—formally similar in certain respects, but doing quite different jobs—the two theories have often been taken to be slightly different variations on the same theme. Some have proposed that even though the Kaplan semantics as it is does not apply to the phenomena of informative necessary truth, it can be modified and extended so that it does.[5] This kind of project suggests an alternative interpretation of the two-dimensional framework, as applied to our problem.

The idea is to take a Kaplanian character to be a kind of narrow content, for thought as well as for speech.[6] A semantics that fits what I have called the *generalized Kaplan paradigm* treats a much wider range of expressions as context-dependent: almost all descriptive expressions of the language will have a variable character. While in the original Kaplan theory, it was the content determined that was the thought expressed in the use of an expression, in the generalized theory, it is the character (or the A-intension, or diagonal, that it determines) that is the cognitive value of what is expressed. When a person thinks or says that Socrates lived in Athens, or that there is water on Mars, the thought that he has or expresses is a descriptive proposition about whatever the person and city, or substance and planet, are that fit certain descriptions, or that present themselves to the thinker in certain ways. The C-intension determined will be a singular proposition about Socrates and Athens, or a proposition about the actual substance water and the planet Mars, but these are propositions to which the speaker or thinker has only indirect access. The rigidity of the proper names and natural kind terms is the result of a kind of generalized scope device. The character, or two-dimensional intension, for a thought or utterance corresponds to a non-rigid description of a proposition (the C-intension). The A-intension is the proposition that the C-intension that fits this description is true. The content of the thinker's

[5] It is mainly David Chalmers and Frank Jackson that I have in mind as proponents of the generalized Kaplan interpretation, though my rough sketch of the view may not exactly match the way either of them would state it. See Chalmers (1996) and (2002) and Jackson (1998).

[6] Frank Jackson and David Chalmers give different answers to the question whether the two-dimensional apparatus applies to thought as well as to language. Chalmers assumes that mental states as well as utterances are associated with the two different kinds of intension, while Jackson makes the A-intension/C-intension distinction only for linguistic expressions. But as I understand him, Jackson would say that when one makes an assertion, or when one attributes a belief, it is in general the A-intension of the sentence used to make the assertion, or of the sentential clause used to attribute the belief, that is the proposition that the speaker expresses, or that the subject of the belief attribution is said to believe. This is the sense in which, for Jackson, the A-intension represents the cognitive value of an expression. Since for Chalmers, thoughts (mental analogs of sentences) themselves have the two kinds of intension, it is slightly less straightforward to say that it is the A-intension that is the cognitive value of an expression, but on his view, it is only the A-intension to which the thinker has access.

thought is not the proposition described, but the proposition that this proposition, whatever it is, is true.

In contrasting the different interpretations of the two-dimensional framework, I have used the labels "metasemantic" (for the interpretation I want to defend) and "semantic" (for the generalized Kaplan interpretation). The terminology marks a distinction between questions about what the semantic values of expressions are and questions about what the facts are that determine those semantic values. Kaplan introduced it to contrast two different ways of understanding a causal theory of reference.[7] The direct reference account says that the semantic value of a name is simply its referent. (That is the whole *semantic* story.) The causal mechanisms explain what the facts are that make it the case that names have the semantic values that they have. (They are part of the *metasemantic* story.) A contrasting theory, "causal descriptivism," holds that the causal mechanisms belong in the semantic story: one should take the semantic value of the name to be a description something like this: "*the individual who lies at the other end of the historical chain that brought this token to me*."[8] I labeled my interpretation "metasemantic" because the second dimension represents the facts in virtue of which the utterance in question has the semantic content that it has. The generalized Kaplan interpretation was called "semantic" because the two-dimensional intension—the analogue of Kaplanian character—and the A-intension that it determines, are semantic values of the expressions. The generalized Kaplan interpretation is a kind of generalization of causal descriptivism.

It is not important what gets called "meaning," or labeled "semantic." The significance of the contrast between the two kinds of interpretation is in the order of explanation of the fact that an utterance is associated with the particular two-dimensional intension that it is associated with, and in the kind of account of intentionality that the contrasting stories require. In the metasemantic story, the problem of intentionality is addressed at the level of C-intensions, which are the contents of thought, and the cognitive values of expressions, in the normal case. Propositional concepts are defined, for an utterance token, only relative to possible worlds in which the utterance event takes place,[9] and the diagonal propositions determined by propositional concepts are local and context-dependent. One can define a propositional concept for any context, but in the normal case, where speakers know what they are saying (according to the standard semantic rules) and hearers are presupposed to understand what is said, the propositional concept will be constant, relative to the context, and so the diagonal proposition, or A-proposition, will be

[7] See Kaplan (1989b: 574). I use this terminology to distinguish the two kinds of interpretation of the two-dimensional apparatus in Stalnaker (2001) and Stalnaker (forthcoming).

[8] Kaplan (1989: 574).

[9] As discussed in Stalnaker (1987), one extends propositional concepts to possible worlds not containing an utterance token in applications of the diagonalization strategy to belief attribution by considering what a token that-clause *would* have said if uttered in a certain possible world. But as I emphasized, the counterfactual is vague, and this is an ad hoc, case-by-case procedure that requires charitable interpretation.

the same (relative to the context) as the horizontal, or C-proposition.[10] In contrast, the generalized Kaplan interpretation addresses the problem of intentionality on the level of two-dimensional intensions, or A-intensions. Since these intensions are not defined in terms of what an utterance expresses or would express in the relevant possible world, they can be assumed to be defined for a broader range of possible worlds. What matters is not what the content of an utterance would have been if uttered in some alternative possible world, but what value the actual two-dimensional meaning takes where the argument of the function is the alternative possible world. A-intensions are, in the general case, the cognitive values of expressions, and for all cases where the C-intension depends on the external environment (including all cases involving proper names, natural kind terms, color words, or any terms to which "twin-earth" thought experiments might be constructed), the A-intension will differ from the C-intension.

The two interpretations make different assumptions about what we have cognitive access to because they have different accounts of what cognitive access is. A thoroughly externalist account of intentionality, since it gives an externalist account of thought as well as speech, gives an externalist account of cognitive access. Knowing who Socrates is, and so having cognitive access to singular propositions about Socrates, is a matter of being appropriately causally related to Socrates. Knowing who someone is, and so knowing what singular proposition is expressed by some singular statement, is of course highly context-dependent, and cognitive access will be context-dependent in the same ways. Cognitive access, on an externalist theory, is not simply a matter of the strength of an acquaintance relation. It is a matter of whether a person's state of mind is aptly described in terms of an individual (in terms of a distinction between possible worlds in which that individual has a certain property and worlds in which the individual does not).

The account of intentionality implicit in the generalized Kaplan interpretation is internalist. Two-dimensional and A-intensions are assumed to be determined, in general, by the internal properties of the speaker or thinker, and to be accessible a priori.[11] A prioricity is identified with the necessity of the A-intension—an identification that does not have any plausibility on the metasemantic interpretation. Even paradigm cases of truths knowable a priori (for example simple mathematical truths) will have contingent diagonals in some contexts, on the metasemantic account. Consider a context in which a person is uncertain about whether the intended meaning of a certain token of "$7 + 5 = 12$" is the usual one, or one that uses a base 8 notation, with the same numerals for one through seven. In some possible worlds compatible with the beliefs of this person, the token expresses the falsehood that seven plus five is

[10] To say that two propositions are the same, relative to a context, is to say that the two functions from possible worlds to truth values take the same values for all possible worlds compatible with the context. So, for example, the proposition that the current President of the United States is a Republican is the same as the proposition that G. W. Bush is a Republican, relative to a context in which it is presupposed that G. W. Bush is the President.

[11] The thesis that a sentence is a priori if and only if it has a necessary A-intension is described by Chalmers as "the core thesis" of his interpretation of the two-dimensional framework. It is made true by definition in that interpretation. See Chalmers (2002).

ten, and so the diagonal will be contingent. More generally, any utterance, no matter how trivial the proposition that it in fact is used to express, might have been used to say something false, and a person might have misunderstood it to say something false. So the metasemantic interpretation yields no account or representation of a priori truth or knowledge, and does not depend on any notion of the a priori.[12] This may be regarded as a strength or a weakness of the metalinguistic interpretation, depending on one's attitude toward the notion of a priori knowledge and truth, but it is a clear difference between the two interpretations.

3. Contextual and Epistemic Interpretations

David Chalmers is one of those I have in mind as a proponent of the generalized Kaplan interpretation, but he has a different way of contrasting his own interpretation with alternatives. Chalmers distinguishes *contextual* from *epistemic* understandings of the two-dimensional framework. The one "uses the first dimension to capture *context-dependence*," while the other uses it "to capture *epistemic dependence*."[13] This classification cuts across the semantic/metasemantic distinction that I am making, since both Kaplan's semantics for demonstratives and the metasemantic interpretation count as contextual interpretations. In Chalmers's classification, the kind of contextual interpretation that comes closest to the metasemantic account is one that identifies a two-dimensional intension with what he calls a "token reflexive contextual intension." To evaluate such an intension for a given actual utterance token at a possible world, we consider what proposition is expressed by that particular utterance token in the possible world in question. Chalmers argues that the token reflexive account is problematic since it seems to depend on questionable metaphysical assumptions about the essential properties of linguistic and mental tokens—about the way they are identified across possible worlds. To borrow and adapt an old example of Donald Davidson's, suppose Daniels says "Empedocles leaped," but O'Leary took him to be speaking German, saying "Empedocles liebt." Is the token utterance of the German sentence that Daniels utters in the possible world that O'Leary thinks we are in really the very same token as the token of the English sentence that he utters in the actual world? Do we have to assume that it is in order to use the two-dimensional framework to represent the misunderstanding? This is a good question, but I think one can bypass metaphysical questions about the essential properties of tokens. It will suffice for the metasemantic propositional concepts that the tokens in the alternative possible

[12] As David Chalmers keeps reminding me, in the face of my increasingly strident expressions of skepticism about a priori truth and knowledge, I did say, in "Assertion," that a certain two-dimensional modal operator, which says that the diagonal proposition is necessary, could be understood as the a priori truth operator (p. 85). I now think that this was an ill-considered remark. The notion of a priori truth that this identification yields is at best a very local and context-dependent one.

[13] Chalmers (2002).

worlds be epistemic counterparts of the actual token.[14] In a context to which the two-dimensional apparatus can be straightforwardly applied (either in a case where reinterpretation by diagonalization is required, or in a case of ignorance or misunderstanding of what is said), the relevant people will believe, or presuppose, that a particular utterance event takes place, so there will be a uniquely salient utterance token in each of the relevant possible worlds. So long as it is clear which utterance token it is, it does not matter whether it is literally the same one. (Though Chalmers's point does underscore the extent to which the application of the apparatus, on the metasemantic interpretation, is context-dependent. A propositional concept, and the diagonal proposition it determines, will be well-defined only for a limited range of possible worlds.)

What is the epistemic interpretation that Chalmers contrasts with all versions of the contextual theory? Here is the way I understand it: We start with an "epistemic space," a set of possibilities, or scenarios, "ways things might turn out to be, for all we know *a priori*." These scenarios can be described in a canonical language which is "semantically neutral," which means roughly that the terms in it are not "twin-earthable": the two kinds of intensions (A-intensions and C-intensions, to stay with Jackson's notation) will coincide for the terms of the canonical language. It is assumed that this special language is rich enough to give a complete description of the scenarios, or points of the epistemic space. A description is complete if knowing it would suffice to put one in a position to know any truth by reasoning alone. The two kinds of intensions are then defined as functions with the subsets of this space of possibilities as its range. Thought and speech in general are then interpreted by assigning these intensions to thought and utterance types. The project is, in effect, a project of reduction to the canonical language, for which all content is narrow, and knowable a priori.

The first thing to note about Chalmers's account of epistemic space is that since it defines the space of possibilities in epistemic terms, and takes epistemic notions such as a priori knowledge as unexplained primitives, it does not directly address what I regard as the central question of interpretation: what are the facts in virtue of which expressions are associated with the intensions (one or two-dimensional, A or C) that they are associated with? But it does put constraints on the way that question can be answered, and I am skeptical that they can be met. Since the contents of the sentences of the rich but neutral canonical language are narrow contents—determined by the intrinsic properties of the speakers of that language—it, like any version of the generalized Kaplan interpretation, needs an internalist solution to the problem of intentionality. I doubt that any such solution can be made to work, but I think one can say something about the general shape that a successful account of this kind would have to have. All attempts to address the problem of intentionality consist mainly of the waving of hands, but the different kinds of hand-waving suggest very different pictures of our intentional relations to the world. The clearest and best developed internalist account that I know of is a theory David Lewis calls "global descriptivism,"

[14] This problem is briefly noted in Stalnaker (1981: 138, n. 14).

and I think this is the kind of account that any proponents of the generalized Kaplan interpretation should find congenial.[15]

4. Global Descriptivism

Global descriptivism begins by analyzing names and some general expressions in terms of definite descriptions. The descriptions may involve causal notions, and a reference to the speaker (as in "the individual who lies at the other end of the historical chain that brought this token to me"), and the descriptions may be rigidified ("the *actual* man who corrupted Hadleyburg"). One may define several names together ("Cicero and Cataline are the men such that the first denounced the second and . . . "). But as Lewis emphasizes, an analysis of a name or a predicate in terms of a definite description simply passes the semantic buck from one part of the vocabulary of the language to another. The idea of the global theory is to interpret all the non-logical terms of a language at once. The method follows and generalizes Frank Ramsey's proposal for interpreting theoretical terms in which predicates are replaced with variables, bound by quantifiers. One does not refer directly either to individuals or to empirical properties and relations, but instead quantifies over them. The content of one's theory is that there exist properties and relations that are related to each other in the way that the laws and generalizations of one's theory say that they are. The theory is true in the set of those possible worlds that provide an appropriate model for the theory.

I said that according to global descriptivism, a speaker's theory is true in possible worlds that provide an *appropriate* model for the theory since as Lewis emphasizes, an unconstrained global descriptivism would be untenable. Properties and relations, unconstrained, are plentiful enough to provide models for any theory in any possible world, so if one required only that there be some model for the theory, then all theories would be true in all possible worlds (or at least in all possible worlds of the right size). That is Putnam's paradox, and Lewis takes it to refute an unqualified global descriptivism. The main burden of this account of intentionality, as Lewis argues, is to explain the constraints on the properties and relations that the quantifiers range over, and so that define the restricted class of models that are appropriate. Lewis's idea is that the properties and relations must be more or less natural, and the hope is that this kind of constraint will suffice to give a version of global descriptivism that gives empirical claims the kind of substantive content that they seem to have.

It will be agreed by all that this kind of theory will not be even remotely plausible unless the theory being interpreted is a rich and detailed one. Suppose one's theory said only that all swans are black. Then the global descriptivist analysis would say that the content of the theory is that there exist two properties such that everything that satisfies the first also satisfies the second. Even if the class of properties one is quantifying over is restricted to simple, natural properties, this claim will obviously not be a plausible paraphrase of the generalization about swans. To have a chance of

[15] See Lewis (1984). I discuss Lewis's global descriptivism in more detail in Stalnaker (2004c).

plausibility, the global descriptivist account must be applied to a theory with many predicates, and a large number of generalizations about the interrelations between the properties and relations that instantiate them. The individual claims that a theory makes cannot be understood independently of the whole theory.

A proponent of the generalized Kaplan interpretation of the two-dimensional framework, and of the internalist conception of content that it requires, need not subscribe to all of the details of Lewis's account of intentionality, but I think that any internalist account of intentionality will share with his account certain features that seem to me problematic. I will conclude by sketching two objections to global descriptivism, which I will call the holism problem and the indirectness problem. More neutrally, perhaps I should call them two distinctive features of that kind of theory, since others may find them less problematic than I do. In both cases, the presence of these features helps to explain the role of the second dimension in trying to reconcile this kind of account of intentionality with the phenomena of speech and thought.

First, the holism problem: Meanings and contents, on this kind of account, will be extremely unstable and idiosyncratic. Since interpretation goes by way of a total theory, any change in the total theory, however minor, will bring about a change in the contents of everything expressed in the theory, and any difference between your total theory and mine will mean that the contents of all my claims will differ from any of yours. This is a familiar objection to internalist accounts of meaning.[16] Jackson and Chalmers acknowledge that the kind of descriptivism they are defending is holistic in this way, but apparently do not take it to be a problem. It may be, they say, that "Leverrier uses 'Neptune' as a name for whatever planet perturbs the orbit of Uranus," while his wife uses the same name as a name for the "astronomical object for which her husband is searching . . . . 'Neptune (if it exists) perturbs the orbit of Uranus' is a priori for Leverrier but not for his wife." The same kind of variability might affect natural kind terms: they suggest that it might be a priori for a city dweller, but not a beach dweller, that water comes out of faucets, and a priori for the latter but not the former that water is the liquid in the ocean.[17]

The second dimension of meaning is supposed to soften the effect of the fact that we never mean the same thing as others with whom we communicate, or mean the same thing ourselves from day to day. If the descriptions that give the meanings of our names and predicates are rigidified descriptions, then "that will avoid confusion between people who have attached the same term to the same referent by means of different descriptions."[18] It is rigidification that gives rise to the general A-intension/C-intension distinction, and the derivative C-intensions will tend to be much more stable across time and person than the A-intensions from which they are derived. The C-intensions will play an essential mediating and stabilizing role. But the holism feature does yield a peculiar account of communication. It is the A-intensions that are the cognitive value of our thoughts and utterances—the

[16] Jerry Fodor, for example, takes this to be a devastating objection to conceptual role theories of meaning. See Fodor (1987).

[17] Chalmers and Jackson (2002). [18] Lewis (1984: 59).

propositions to which we have access—but they are not what must be the same for successful communication. The thought I express—what I believe when I am sincere and say what I believe—is rarely if ever the same as what you come to believe when you accept what I say.

Second, the indirectness problem: The kind of content to which we have access, according to the global descriptivist theory, is extremely abstract. It is not just that we do not refer directly to particular individuals, and entertain singular propositions. We also do not describe things in terms of ordinary empirical properties and relations, but only in terms of whatever properties and relations are the ones that best fit the abstract structure given by our uninterpreted theory. Again, it is the second dimension that is supposed to mitigate the indirectness and give us a kind of access to the individuals that inhabit our world and to the empirical properties and relations that they instantiate. Our utterances, and perhaps our thoughts, have singular propositions and propositions involving empirical properties and relations as their C-intensions. But on this interpretation, it is only the two-dimensional intensions and A-intensions to which we have cognitive access, and according to the global descriptivist account, we *never* have cognitive access to any propositions except the very abstract ones that existentially generalize over empirical properties and relations. The rigidification operations that give us the second dimension are, in effect, devices for describing propositions that we cannot grasp. We can, for example, describe the proposition that there is water (water itself, not whatever it is that plays the water role) on the floor, but that is not the proposition we believe when we believe that there is water on the floor, or the one we entertain when we consider the possibility that there is. But cannot we, who know that water is H_2O, grasp such propositions? No, because our access to Hydrogen and Oxygen is equally indirect. For the global descriptivist, it is indirect description all the way down.

Julius provides a paradigm here. In one of the early discussions of the problem we have been concerned with, Gareth Evans stipulated that the name "Julius," as he proposed to use it, should be a rigid designator for the person, whoever he or she is, who invented the zip. It seems intuitively clear—it is part of the point of the example—that some competent users of the name "Julius" (presumably including Evans himself) do not know who Julius is: they do not know to whom the name "Julius" refers. It is of course not at all clear what it takes, in general, to know who someone is, or to know what a name refers to, but this seems to be a clear case. And if someone does not know who it is that "Julius" refers to, then he does not know what singular proposition is expressed by sentences using that name, such as "Julius invented the zip," and "Julius was born in Minsk." In such a case, the person can believe that the singular proposition, whatever it is, is true, but that will not be the same as believing the singular proposition itself. So much is pretheoretical intuition, common ground whatever one's interpretation of the two-dimensional apparatus used to describe the case. It seems clear that the propositions believed by a person who is prepared to affirm the truth of these statements are descriptive propositions: the necessary truth that the inventor of the zip invented the zip, and the contingent but general proposition that the inventor of the zip was born in Minsk. These propositions are the diagonal or A-intensions of the statements.

According to global descriptivism, our access to all properties, relations, and individuals is like our access to Julius, and our relation to the C-intensions of all sentences we understand is like our relation to the contingent singular propositions that Julius invented the zip, and that Julius was born in Minsk. While we are in a sense talking about Julius when we say things using this name, we do not know what we are talking about. On a global descriptivist theory, this is the general case: we never know what we are talking about.

The metasemantic interpretation and the externalist, causal/information-theoretic account of intentionality that motivates it, agree with the generalized Kaplan interpretation with its internalist account of intentionality about the case of Julius. That is, it is common ground that it is the A-intensions that are expressed and communicated with sentences using the name Julius in the kind of context that Evans intended. In any context in which it is presupposed that "Julius" names the inventor of the zip, but in which there is no individual who is presupposed to be the inventor of the zip, diagonalization will be required. The disagreement between the different interpretations that I am contrasting is about whether one should think of this as a model for the general case.

It might be nice if we had a neutral language with an internally grounded semantics, a language that required no factual assumptions for its interpretation and that could provide a complete description of the world, and all possible worlds. It might be nice if there were a pure epistemic space to which we had a priori access and in terms of which we could locate our disagreements about what the actual world is like. But I do not think these things are possible. The only way we can describe the world is to use the materials that the actual world offers us—the things, properties and relations that we find there. Where we disagree about the nature of what is to be found in the actual world, we may as a result disagree about what is possible—about the character of the space of possibilities in terms of which our language and thought are interpreted. Semantic and factual issues become intertwined, and that is a problem. Where the disagreements are about the fundamental natures of things, or about identities, the problem is particularly acute. But our resources for describing the world are rich and diverse, and even if there is no absolutely neutral language, we can usually find ways of describing the possibilities that are neutral on the issues in contention in a particular context. The two-dimensional apparatus, on the metasemantic interpretation, is apt for describing the problems that arise from the mix of semantic and factual information, and some of our resources for solving them.

References

Chalmers, David (1996). *The Conscious Mind*. Oxford: Oxford University Press.

____(2002). "The Foundations of Two-dimensional Semantics." (on line paper, www.u.arizona.edu/~chalmers/papers/foundations.html)

____and Frank Jackson (2001). "Conceptual Analysis and Reductive Explanation," *The Philosophical Review* 110, 315–61.

Fodor, Jerry (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: Bradford Books, the MIT Press.

Jackson, Frank (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

——(2001). "Précis of *From Metaphysics to Ethics*," *Philosophy and Phenomenological Research*, 62, 617–24.

Kaplan, David (1989a). "Demonstratives," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford and New York: Oxford University Press.

——(1989b). "Afterthoughts," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford and New York: Oxford University Press.

Lewis, David (1984). "Putnam's Paradox," *Australasian Journal of Philosophy*, 62, 221–36. (Reprinted in Lewis, *Papers in Metaphysics and Epistemology* (Cambridge Studies in Philosophy, Cambridge: Cambridge University Press, 1999). Page numbers refer to the latter.)

Stalnaker, R. (1978). "Assertion," *Syntax and Semantics*, 9, 315–32. (Reprinted in Stalnaker (1999); page numbers refer to the latter.)

——(1981a). "Indexical Belief." *Synthese*, 49, 129–51. (Reprinted in Stalnaker (1999); page numbers refer to the latter.)

——(1981b). "Logical Semiotic." E. Agazzio (ed.), *Modern Logic—A Survey*, Dordrecht: D. Reidel, 439–56.

——(1987). "Semantics for Belief." *Philosophical Topics*, 15, 177–90. (Reprinted in Stalnaker (1999); page numbers refer to the latter.)

——(1999). *Context and Content*. Oxford: Oxford University Press.

——(2001). "On Considering a Possible World as Actual," *Proceedings of the Aristotelian Society* (supplementary volume). Reprinted in Stalnaker (2004a), 188–200.

——(2004a). *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays*. Oxford: Oxford University Press.

——(2004b). "Conceptual Truth and Metaphysical Necessity." In Stalnaker (2004a), 201–15.

——(2004c). "Lewis on Intentionality," *Australasian Journal of Philosophy* 82, 199–212. Reprinted in F. Jackson and G. Priest, *Lewisian Themes: The Philosophy of David Lewis*. Oxford, Oxford University Press, 2004, 231–44.