Introduction

THE ESSAYS IN THIS VOLUME are concerned with four main topics—propositions and attitudes, modality, truth and vagueness, and skepticism about intentionality. The significance of these issues extends well beyond the philosophy of language. In addition to being semantically encoded by sentences, propositions are asserted, believed, and known. Questions about what they are, and how we come to believe or know them—as well as questions about which propositions are expressed by which sentences, and which are asserted by which utterances—are crucial to epistemology and the philosophy of mind, as well as being the touchstone of the systematic study of meaning. The next topic, modality, brings together the study of reference and essence, indexicality and actuality, and the distinction between metaphysical and epistemic possibility. Here, the central issues include the role of evidence in our knowledge of the necessary a posteriori, the metaphysical makeup of possible world-states, and the different ways we acquire knowledge of them. The third topic, truth and vagueness, must be covered by any systematic study of language. But that's not all. In addition to being properties of expressions, truth and vagueness apply to that which these parts of language express, or designate. Sentences are true when the propositions they express are true, and—to the extent that they are vague—it is often not because it is vague which perfectly precise propositions they express, but because the propositions they clearly express are vague. The same is true of predicates, the vagueness of which is often tied to the vagueness of the properties they express. Even singular terms, and the objects they designate, are not exempt. The idea that, apart from language, the world and its objects are pristinely precise is a metaphysical prejudice. The spatial and the temporal boundaries of many things are vague. More generally, clarity about the nature of truth, and an appreciation of how vagueness limits our ability to give precise answers to certain questions, are required in every area of philosophy. Finally, skepticism about intentionality is not just skepticism about meaning, but also skepticism about belief, and mental content. There is no understanding, or rebutting, one without doing the same for the other.

In addition to addressing these topics, the essays that follow will, I hope, illustrate the interpenetration of issues in the philosophy of language with those in other core areas of philosophy. It's not that philosophy of language

is "first philosophy," providing solutions to heretofore intractable problems in other areas. The old view—so prevalent in the mid–twentieth century—that philosophical problems are simply linguistic confusions to be dissolved by a correct understanding of meaning has, mercifully, fallen by the wayside. What exactly has replaced it is, I think, not yet clear. However, whatever it turns out to be, language will, I am confident, remain important. The discipline provided by systematic logical and semantic frameworks, plus explicit pragmatics of assertion and implicature, is too valuable to be neglected by those working in other areas. Of course, influence also flows the other way. Philosophers of language working, as I do, on the necessary a posteriori and the contingent a priori, the semantics of 'actually', and skepticism about intentionality must draw upon discussions in other areas. In some cases, I fear. I have not done enough of this. Still, I hope that these essays, however imperfect, illustrate how the philosophical investigation of language may influence, and be influenced by, investigations in other core areas.

PART 1: REFERENCE, PROPOSITIONS, AND PROPOSITIONAL ATTITUDES

The aim of the essay 1, "Direct Reference, Propositional Attitudes, and Semantic Content" (1987), is to pull apart two conceptions of semantics. According to the first, the central task of semantics is to tell us "what sentences say" (relative to contexts in which they are used).[1] According to the second, the task is to specify the truth conditions of sentences (relative to contexts). On the first conception, the meaning of S is a function from contexts to "what is said by S" in those contexts. On the second, it is a function from contexts to "the truth conditions of S," as used there. When the essay was written, it was common to endorse both conceptions—implicitly embracing the idea that "what is said by S" (in a context C) can be identified with the truth conditions of S (in C). Against this, I argued that there is no reasonable understanding of truth conditions on which this identification is sustainable. Sentences do

---

[1] The phrase "what sentences say" is, of course, a barbarism. Strictly speaking, sentences don't say (assert) anything; speakers do. The idea behind the barbarism is, roughly, that what a speaker typically asserts by uttering S in C is (just) the semantic content of S in C. As readers of volume 1 of this collection know, this (overly simple) conception of the relationship between semantic content and assertion is one I don't accept. At the time the essay appeared, however, it was the dominant, though typically unspoken, conception of that relationship. For present purposes, this is not an issue, since the essay's argument about what is, and what is not, the semantic content of S, in the sense of what its semantics systematically contributes to what is said by normal utterances of S, is independent of the precise relationship between semantic content and assertion.

have semantic contents (relative to contexts), and these contents are systematically related to what speakers standardly assert by uttering them.[2] In addition, the semantic contents of many sentences, as well as what they are used to assert, have truth conditions. But neither these contents, nor the asserted propositions, can be identified with sets of truth conditions. For any such set, there are many semantic and assertive contents which, though true in precisely those conditions, differ sharply from one another.

To establish this, I formulated an abstract conception of truth conditions in which different conceptions of truth-supporting circumstances (at which sentences are evaluated for truth) can be arranged on a continuum from the coarsest-grained (complete, consistent, metaphysically possible world-states) to the finest-grained (partial, and sometimes inconsistent, abstract situations). Next to be abstracted were rules, specifying the truth conditions of complex sentences in terms of the truth conditional contributions of their parts—for example, rules specifying the circumstances supporting the truth of a conjunction (disjunction) as the intersection (union) of those supporting the truth of the conjuncts (disjuncts). Finally, three assumptions were articulated—one treating ⌜A believes/asserts that S⌝ as reporting an agent as believing/asserting the semantic content of S (relative to a context and assignment), one taking these attitude verbs to distribute over conjunction, and one treating names, indexicals, and variables as directly referential—that is, as having semantic contents identical with their referents (relative to contexts and assignments). It is shown that if these assumptions are correct, then the semantic content of a sentence cannot be identified with the set of circumstances supporting its truth—no matter how fine-grained these are taken to be.

The first examples used to establish this were attitude ascriptions containing proper names in their content clauses. However, since the directly referentiality of names was still contentious, analogous examples involving indexicals, or variables bound from outside the clauses, were shown to work equally well. Since it is hard to deny that at least some of these are directly referential, the strategy of blocking the result by denying direct reference was rejected. A similar fate befell another strategy for avoiding the result—namely, denying that attitude ascriptions report attitudes toward the semantic contents of their complement clauses. Against this, it was argued that any relational semantics of such ascriptions compatible with obvious linguistic facts will require one to distinguish the semantic contents of sentences from sets of circumstances in which they are true.

---

[2] My view of this relationship is set out in chapters 3 and 4 of Soames (2002), and modified in essays 9, 10, 11, and 14 of volume 1 of this collection.

With this negative result in place, the essay articulates a positive account of semantic content, and the objects of propositional attitudes. The account divides into two parts—a metaphysical account of the facts in virtue of which attitude ascriptions are true, and a semantic account of the contents of sentences, appropriate to serve as objects of the attitudes. The former sees sentences and mental states as content-bearing vehicles of assertion and belief. The latter identifies semantic contents of sentences with structured, Russellian propositions. The idea is illustrated using a two-stage semantic theory for a simple, first-order language with quantifiers, lambda abstraction, and propositional attitude verbs. Stage 1 is a recursive assignment of structured propositions to formulas relative to contexts and assignments. Stage 2 is a specification of the truth conditions of structured propositions at truth-supporting circumstances. The attitude problems that defeated strict truth-conditional accounts (without structured propositions) are avoided. However, the substitution problems posed by Frege's puzzle for directly referential terms are not. The suggestions at the end of the essay for defusing these problems should be seen as initial ideas, to be supplemented by later work—including *Beyond Rigidity*, as well as essays 9, 10, and 11 of volume 1.

Essay 2, "Why Propositions Can't Be Sets of Truth-Supporting Circumstances" (2008), rebuts Walter Edelberg's (1994) objection to the main negative argument in essay 1. According to that argument, any truth-conditional theory (satisfying certain assumptions) that, correctly, makes (1) and (2) true, must, incorrectly, characterize (3) as also being true.

(1) Hesperus is Phosphorus.
(2) The ancients believed that Hesperus was the brightest body seen in the evening sky and Phosphorus was the brightest body seen in the morning sky.
(3) So, the ancients believed that there was an object that was both the brightest body seen in the evening sky and the brightest body seen in the morning sky.

Since these theories fail to capture the truth conditions of (1)–(3), they must be rejected.

Edelberg objects that the argument rests on a false claim—namely that the truth-conditional theories characterize all instances of (3′) as being *model-theoretically entailed* by the corresponding instances of (1′) and (2′).

(1′) $a = b$
(2′) c believes that (Fa and Gb)
(3′) c believes that $\exists x$ (Fx and Gx)

He is right in maintaining that this entailment claim is incorrect. Theories in which truth-supporting circumstances are allowed to be metaphysically

impossible can make (1′) and (2′) true, and (3′) false, relative to certain *impossible* circumstances in which distinct objects a and b are identified. However, he is wrong in taking my argument to make this false claim. Rather, it claims (i) that in order to be correct, a semantic theory must assign Venus as referent of both 'Hesperus' and 'Phosphorus', while assigning truth conditions to (1) and (2) that make them turn out (actually) to be true, and (ii) that any truth-conditional theory (of the relevant kind) that does this will wrongly characterize the semantic content of (3) as being a *truth-conditional consequence* of the semantic contents of (1) and (2). In short, Edelberg's objection confuses *model-theoretic consequence* with *truth-conditional consequence*.

> Let T be a truth-conditional theory of the relevant sort. A sentence S* is a *model-theoretic consequence* of a set S of sentences, according to T, iff for every model M conforming to T, and every context C and circumstance E of M, if all the sentences in S are true in M with respect to C and E, then so is S*. Let T be a truth-conditional theory of the relevant sort, with intended model M. The semantic content of a sentence (or formula) S*, relative to a context C of M and assignment A of values to variables, is a *truth-conditional consequence* of the semantic contents of a set S of sentences (or formulas), relative to C and A, iff for every circumstance E of M, if all members of S are true in M with respect to C, A, and E, then so is S*.

The upshot is a cautionary tale on the perils of confusing these two, and a discourse on certain respects in which any semantics that aspires to be a theory of meaning must go beyond a theory of the truth of sentences relative to arbitrary models.

Essay 3, "Belief and Mental Representation" (1990), explores the way in which a semantic theory of attitude ascriptions incorporating structured propositions as objects of the attitudes, bears on prominent computational models of belief in the philosophy of mind. My target is a series of books and papers in which Jerry Fodor argues that belief is a relation born by an agent to an internal representation in the agent's "language of thought." According to Fodor, the (truth conditional) content of the representation specifies the way the world is represented to be, while its syntax is the basis for computational operations responsible for the belief's causal powers. Fodor's goal is to vindicate our ordinary belief/desire explanations of behavior, while also providing cognitive psychology with a properly scientific notion of belief. The aim of my essay—which neither endorses nor opposes Fodor's general program—is to show that his best hope of success lies in incorporating the insights of contemporary semantic treatments of indexicals and propositional attitudes.

The early part of the essay explicates Fodor's position, resolves crucial ambiguities in his formulations, and eliminates their most implausible features. This leads to a reconstruction according to which 'believe' expresses a two-place relation that holds between an agent A and a structured proposition p in virtue of A's standing in a certain cognitive relation to an internal syntactic representation R that expresses p. Since the structure of p encodes much of the syntax of R—which, in turn, is responsible for the causal power of the belief—the representational content of the belief and its causal efficacy are brought into line. Although the alignment is not perfect, it is, I argue, close enough to vindicate our ordinary belief/desire explanations.

Misalignment occurs when syntactically identical representations expressing the same proposition contain different directly referential terms. In these cases, what one does may be sensitive not just to the contents of one's beliefs and desires, but also to the particular representations that are in play. As a result, claims of the form (4) cannot expected to be true, exceptionless, universal generalizations.

(4)   If x believes that A is an action that x can perform, and if x believes that performing A will bring it about that Q, and if x wants it to be the case that Q, then x will act in a way intended to be a performance of A.

Nevertheless, there is no cause for alarm. Though principles like (4) are important parts of our "folk psychology," they are so only when supplemented by *ceteris paribus* clauses. Most of the explanations we give of individual actions are perfectly compatible with taking them in this way, and many are counterfactual supporting, despite the fact that (4) may fail in distant possible world-states in which the same beliefs (desires) are held in significantly different ways. As for scientific psychology, the need (sometimes) to include clauses not just about what is believed (desired), but also about how the beliefs (desires) are held, is neither a threat to, nor a radical departure from, our ordinary conception of the mental.

If this is right, then my reconstruction of Fodor's computational model preserves its most plausible features, while avoiding its most implausible ones. Among the latter is his startling claim that "the language of thought" is entirely innate, and that natural language expressions are learned by connecting them with already understood mentalese counterparts. This is no part of my reconstruction, which allows one's internal system of mental representations to include elements of the natural language one speaks—which are, of course, not innate. Because of this, *some* beliefs may depend on the contents of internal representations, *which in turn may depend on the contents of expressions in a public language*. This fits well with linguistic accounts that attribute the contents

of names and natural kind terms to causal-historical connections relating a speaker's use of terms to other speakers, and ultimately to objects in the world. As I see it, social processes like these are crucially involved in determining the contents of some expressions, which in turn play a role in determining the contents of some mental states. Thus, I reject Fodor's reductionist view in which the intentionality of natural language is reduced to the intentionality of propositional attitudes, which is then to be reduced to the intentionality of "the language of thought."

The appendix to essay 3 connects these issues to his views about "narrow psychological content," and its alleged role in scientific psychology. For Fodor, the notion of content needed for the latter must be individuated in terms of the causal powers of mental states, which in turn supervene on brain-states. Since genuine "broad" semantic content is not individuated in this way, and does not so supervene, a notion of "narrow" content is invoked. Roughly put, he takes the narrow contents of mentalese expressions (assigned to brain-states) to be functions from world-states to the "broad" semantic contents these expressions would have if embedded in those world-states. However, there is less here than meets the eye. In order for the parts of Fodor's system to fit together, expressions in the language of thought must be isomorphic to their narrow contents. For each such expression there must be a content, and for each content an expression. Thus appeal to narrow contents doesn't allow one to accomplish anything that can't be accomplished by appealing to expressions. A better way of looking at the two is to see narrow content as that which individuates expressions in the internal representational system that Fodor postulates.

Essay 4, "Attitudes and Anaphora" (1992), uses facts about attitude ascriptions to investigate how sentences containing pronouns anaphoric on singular-term antecedents are understood. The aim is to illustrate the use of *hyperintensional contexts* to distinguish different, but intensionally equivalent, linguistic analyses. The argument rests on semantic assumptions about structured propositions and attitude ascriptions, a standard account of quantification with anaphoric pronouns functioning as variables bound by c-commanding quantifiers, and pragmatic assumptions about what it is to assert and believe bare, Russellian propositions. These assumptions are used to refute analyses according to which pronouns anaphoric on singular-term antecedents inherit the semantic contents of their antecedents (no matter whether these contents are Millian or Fregean). In place of these failed analyses, it is suggested that anaphoric pronouns with c-commanding, singular-term antecedents function as variables bound by a lambda abstraction operator introduced by the anaphoric relation itself.

There is, however, a problem. While this analysis yields satisfying results in cases to which it applies, anaphoric relations in which the pronoun isn't

c-commanded by its singular-term antecedent seem to call for the same treatment. What makes this problematic is that accounts of variable binding in natural language are standardly restricted to cases in which bound occurrences of variables have c-commanding antecedents. Essay 4 suggests that either this restriction must be lifted, or some non-variable-binding analysis must be constructed that yields the desired results in the non-c-commanding cases. Although several such analyses are considered, all are found wanting. Thus, the problem is left unsolved.

Part 2: Modality

The essays in this section illustrate the interpenetration of metaphysical views about modality, epistemological views about knowledge and belief, and linguistic views about reference and semantic content. Essay 5, "The Modal Argument: Wide Scope and Rigidified Descriptions" (1998) rebuts two strategies for avoiding Saul Kripke's celebrated modal argument against descriptivism. According to that argument the meanings (semantic contents) of names cannot be identified with the meanings (contents) of the descriptions speakers associate with them because (i) names are rigid designators, whereas (ii) those descriptions aren't. One strategy for avoiding the argument is to deny (i) by treating names as special descriptions the behavior of which under modal operators simulates rigidity by obligatorily taking wide scope. The other strategy denies (ii) by identifying names with rigidified descriptions. I argue that both strategies fail.

The driving force behind both is the desire to preserve the Frege-Russell explanation of the behavior of names in attitude ascriptions, while also accommodating their behavior in modal constructions. In the former, different descriptions associated with coreferential names are used to explain the apparent possibility of *substitution failure*. In the latter, the wide scope given to these descriptions by the first strategy is invoked to explain the seeming rigidity of names, and the accompanying guarantee of *substitution success*. This strategy treats modal operators and predicates—which combine with an argument A expressing or designating a proposition—as inherently shifty. When A contains no proper names, the modal element is applied to the proposition associated with A. When A does contain a name, the modal is applied to a different proposition.

Although such operators and predicates are perfectly coherent, it is clear that modal operators and predicates in English are not shifty in this way. If they were, the following argument would be understood to have true premises and a false conclusion when n meant ⌜the G⌝, but G didn't express an essential property of the referent of n.

P1. ⌜The proposition that if n exists, then n is G=the proposition that if the G exists, then the G is G⌝

P2. ⌜The proposition that if the G exists, then the G is G is a necessary truth⌝

 C. ⌜Therefore, the proposition that if n exists, then n is G is a necessary truth⌝

Since, in reality, arguments such as this one are recognized as valid, the wide-scope analysis gives the wrong semantics for English.

The second strategy for saving descriptivism treats names as rigidified descriptions. This analysis comes to grief over the interaction of modals with attitude ascriptions. Consider, for example, the claim that 'Aristotle' is synonymous with some rigidified description 'the individual who actually was so-and-so'.[3] If this were so, then it would be *impossible* to believe *that Aristotle was a great logician* without believing *that the individual who actually was so-and-so was a great logician*. But that can't be right. Since to believe *that the individual who actually was so-and-so was a great logician* is to believe, of the actual world-state @, *that the individual who was so-and-so in @ was a great logician*, it is *impossible* to believe *that the individual who actually was so-and-so was a great logician* without believing something about @. By contrast, some agents in merely possible world-states believe that Aristotle was a great logician, without having beliefs about @. Since it is *possible* to believe *that Aristotle was a great logician* without believing *that the individual who was so-and-so in @ was a great logician*, 'Aristotle' isn't synonymous with 'the individual who actually was so-and-so'. The point generalizes to all names and 'actually' rigidified descriptions.[4] Thus, Kripke's modal argument remains intact.

Essay 6, "The Philosophical Significance of the Kripkean Necessary A Posteriori" (2006), identifies and assesses two Kripkean routes to the necessary a posteriori. According to the first, successful, route, these truths attribute properties to objects or kinds that, though essential to them, can be known to be possessed by them only empirically. This leads to a sharp distinction between metaphysical and epistemic possibility. When p is a

---

[3] Rigidifying descriptions with David Kaplan's 'dthat' operator is not an attractive option for the descriptivist. Since the content of ⌜dthat (the D)⌝ is the object denoted by ⌜the D⌝, sentences in which coreferential, 'dthat'-rigidified descriptions are substituted for one another express the same proposition. This undercuts the chief motivation for descriptivism by depriving the analysis of names of the Frege-Russell account of their apparent behavior in attitude ascriptions. For more limitations on 'dthat', and on its use in analyses of names, see Soames (2005, 320–22). Further relevant discussion can be found in Soames (2003, 2:414–16).

[4] A slightly expanded and updated version of this argument is given in Soames (2002, chap. 2).

necessary, a posteriori truth attributing an essential property to something (conditional on its existing), the falsity of p is perfectly conceivable, even though it is metaphysically impossible. Thus, world-states in which p is false are epistemically, but not metaphysically, possible.

World-states can be identified with properties. Metaphysically possible states are maximally complete ways the real concrete universe could have been. Epistemically possible states are maximally complete ways the universe can coherently be conceived to be, which it cannot be known a priori not to be. Just as there are properties that ordinary objects could have had, and others they couldn't have had, so there are maximal properties that the universe could have had, and others it couldn't have had. Just as some properties that objects couldn't have had are properties that one can conceive them as having, and that one cannot know a priori they don't have, so some maximal properties that the universe couldn't have had are properties that one can conceive it as having, and that one cannot know a priori it doesn't have. These world-states are epistemically, but not metaphysically, possible. The reason empirical evidence is needed for knowledge of Kripkean necessary a posteriori truths is to rule out metaphysically impossible, but epistemically possible, world-states in which they are false. On this picture, conceivability is a useful, but fallible, guide to possibility. It is fallible because before we know much about what is actual, there are many epistemically possible states that appear to be genuinely possible, and so remain candidates for being metaphysically possible. The more we learn about the world, the more we whittle down this field of candidates, and the better able we are to identify the scope of genuine possibility. In short, our guide to metaphysical possibility is conceivability plus actuality.[5]

Kripke's revival of essentialism, together with his antidescriptivist semantics, also gave impetus to externalism about linguistic and mental content, while turning analytic methodology away from its previous overreliance on conceptual and linguistic analysis. However, this wasn't the whole of his legacy. Unfortunately, these positive developments were threatened by a confusion embedded in his second route to the necessary a posteriori. Although nearly all familiar Kripkean examples of such truths can be reached by the first, essentialist, route, identity sentences involving *linguistically simple* names or natural kind terms cannot. For these, Kripke appeals to a second route, which he (implicitly) generalizes to all his examples. On this route, the sharp distinction between epistemic and metaphysical possibility is lost, as are the deep lessons for philosophy, including its liberation from the confines of linguistic and conceptual analysis.

[5] For further discussion see my "Kripke on Epistemic and Metaphysical Possibility" (Soames n.d.), as well as Soames (2005, 196–209).

According to the second route, instances of the necessary a posteriori express necessary propositions about individuals or kinds. However, the reason empirical evidence is required to know these propositions is not to rule out possible world-states in which they are false. Rather, evidence is needed to rule out the falsity of certain contingent—descriptive or metalinguistic—propositions the truth of which we rely on in coming to believe the necessary truths. Essay 6 makes the principles connecting these two sets of truths explicit, and uses Kripke's own examples to rebut them—thereby undermining his second route to the necessary a posteriori, and restoring the philosophical significance of his first.

Kripke's discussion of the mind-body identity thesis illustrates the perils of not distinguishing the two routes. His discussion centers on two claims.[6]

(5) Heat = mean molecular kinetic energy
(6) Pain = C-fiber stimulation

Initially, both appear to be contingent. However, Kripke argues, in the case of (5) this impression is illusory. Suppose that 'heat' and 'mean molecular kinetic energy' are rigid designators—that is, that the state which is heat couldn't have existed without being heat, and similarly for the state of having such-and-such mean molecular kinetic energy. Then (5) will be necessary, if true. Since it is true, it is also necessary.

What, then, is responsible for the illusion that it isn't? For Kripke, the answer has to do with how we identify, or fix the referent of, 'heat'. Since our primary means of identifying heat is by the sensations it causes, he imagines us relying on the description 'the cause of such-and-such sensations' to identify it. The illusion that (5) is contingent comes from mistaking this (nonrigid) identifying description for a synonym of 'heat', thereby confusing the necessary truth expressed by (5) with the contingent truth expressed by (5*).

(5*) The cause of such-and-such sensations in us = mean molecular kinetic energy.

One who is confused in this way wrongly takes genuinely possible world-states at which kinetic energy exists without the usual accompanying sensations to be world-states with kinetic energy but no heat. Hence, the illusion of contingency.

Kripke finds the situation with (6) to be different. For the sake of argument, we assume, as before, that the terms flanking '=' are rigid, in which case (6) is necessary, if true. In this case, he argues, there is no way of dismissing the initial impression of contingency. With (5), the illusion

---

[6] For simplicity, we treat these as strict identities, with singular terms designating kinds (rather than instances of the kinds) flanking the identity signs.

of contingency is, he thinks, due to the fact that we rely on our sensations to detect the phenomenon that causes them, which is what we use 'heat' to talk about. With (6), the sensation is the very thing we use 'pain' to talk about. We don't say to ourselves: *What a horrible sensation! Let's use the word 'pain' to talk about whatever causes it.* Instead, we use the word to designate the sensation itself. Thus, whereas one can dispel the illusion that (5) is contingent by distinguishing between heat and the sensation used to detect it, one can't dispel the impression of contingency of (6) that way. Since Kripke didn't see any other way to do so, he concluded that (6) *must really be contingent, if it is true at all.* But, it can't be contingent, since its terms are rigid. Thus, he concludes, isn't true.

The argument contains several points of contention. However, its most glaring weakness is its neglect of other explanations of the impression of contingency—most notably, the distinction between epistemic and metaphysical possibility. Think again about (5). Although Kripke identifies one source of the illusion of its contingency, he neglects another. Imagine a person who does *not* take 'heat' to be synonymous with any phrase about our sensations—on the grounds that the existence of heat doesn't depend on there being sensations to detect it. Such a person might still regard (5) as contingent, because it is *conceivable* for heat to be something other than the motion of molecules. After all, the person might reason, it was an empirical discovery that how hot something is depends on how fast its molecules are moving. Since empirical evidence was needed to rule out possibilities in which heat is something other than molecular motion, it must be possible for it to be something else. Hence, the person may conclude, (5) must be contingent. The mistake here is that of confusing epistemic possibility with metaphysical possibility, which is a second source of the illusion that (5) is contingent. But now the flaw in Kripke's argument against (6) is obvious. Having dismissed only one of two sources of the "illusion" that it is contingent, if true (supposing that the relevant terms really are rigid), he is not entitled to his conclusion, unless he can dismiss the other source as well.

This flawed argument is part of a larger problem. The real danger lies in losing the distinction between epistemic and metaphysical possibility, and in wrongly characterizing the relationship between conceivability and possibility. The error to be avoided is in supposing that whenever something genuinely impossible seems to be conceivable, it is *always* because we are confused about *what we are conceiving*—in the way in which Kripke imagines one to be confused who takes the conceivability of molecular motion without heat sensations to be the conceivability of molecular motion without heat. The Coherent Conceivability Thesis, enshrining this error, is the price of relying on his faux second route to the necessary a posteriori. According to this thesis, if we can coherently conceive—without confusion of

the sort indicated by (5/5*)—of a world-state in which p is true (false), then there are genuine (metaphysically) possible world-states in which p is true (false). Instances of the necessary a posteriori in which I attribute essential properties to myself, using the first-person, singular pronoun, illustrate the falsity of this thesis. Since I identify the referent of my use of 'I' directly—without detour through identifying descriptions—confusion of the sort provided by Kripke's heat example doesn't arise. Yet the necessary truths expressed are easily conceived to be false. Since in these cases there is no confusion about *what we are conceiving*, the metaphysical impossibility of what is conceived refutes the Coherent Conceivability Thesis.

Essay 6 closes with a sketch of how the missteps in Kripke's second route to the necessary a posteriori, and his discussion of mind-body identity, are systematized and made worse in David Chalmers's philosophically ambitious system of two-dimensional modal semantics. I close by noting how the 2D version of Kripke's uncharacteristic error serves a larger historical agenda of attempting to reinstate the linguistic analysis of the modalities that his genuine insights showed us how to replace, and, in so doing, of returning philosophy to the confining orthodoxy of conceptual analysis, from which we thought we had escaped.

Essay 7, "Knowledge of Manifest Kinds" (2004), investigates the linguistic and epistemological underpinnings of examples of the necessary a posteriori involving natural terms governed by reference-fixing intentions of the following sort:

(7) a. 'Green' is to designate the property of surfaces that causes (nearly) all members of a certain class of paradigmatic samples to appear similar to us (and different from certain other samples).

   b. 'Water' is to designate the property shared by (nearly) all members of a certain class of paradigmatic samples (rain drops, puddles, lakes, etc.) that explains their most salient characteristics (their boiling and freezing points, their clarity, potability, etc.).

The properties mentioned here are individuated by their possible instances, so that necessarily equivalent properties are identical. Like other abstract objects, they are capable of having different instances at different world-states. However, it is also natural to think of them as parts of the world that exist, and are known by us, through their instances. These worldly properties are natural kinds. For example, the kind water turns out to be *the property of being made up of molecules containing two hydrogen atoms and one oxygen atom*. However, 'water' is not synonymous with this codesignative phrase. Since simple natural kind terms are directly referential, the property designated by 'water' is also its meaning

(semantic content). By contrast, the meaning of the phrase is a structured complex the constituents of which are the meanings of its syntactically significant parts. Since this complex content determines the same worldly property at all possible world-states, the phrase rigidly designates the kind rigidly designated by 'water'. Thus (8) is a necessary truth, which expresses a proposition different from those expressed by (9a) and (9b).[7]

(8) For all x, x is water iff x is made up of molecules containing two hydrogen atoms and one oxygen atom.

(9) a. For all x, x is water iff x is water.
   b. For all x, x is made up of molecules containing two hydrogen atoms and one oxygen atom iff x is made up of molecules containing two hydrogen atoms and one oxygen atom.

The knowledge expressed by (8) is *de re* knowledge of the kind water. Just as our *de re* knowledge of individuals standardly depends either on our own acquaintance with them, or on the acquaintance of others who pass important parts of their knowledge on to us, so our *de re* knowledge of manifest kinds standardly depends either on our own acquaintance with them, or on the acquaintance of others who transmit their knowledge to us. Since acquaintance with these kinds is achieved by acquaintance with their instances, this means that the knowledge expressed by (8) rests ultimately on acquaintance with instances of water, and knowledge, of those instances, that they have a certain molecular structure. This knowledge can be had only empirically. Thus, the proposition expressed by (8) is knowable only a posteriori. The same is true of innocuous-seeming examples involving descriptions used to introduce, or semantically fix the reference of, natural kind terms. As I explain in the essay, (10a) is knowable only a posteriori (where W is the set of paradigmatic samples used in the stipulation, (7b), fixing the reference of 'water'), for the same sort of reason that (10b) is.

(10) a. All, or nearly all, members of W are water, if they are members of any one natural kind (of the relevant sort).
   b. He [said pointing at the man standing in front of me] is standing in front of me, if any (unique) man is standing in front of me.

Essay 8, "Understanding Assertion" (2006), discusses Robert Stalnaker's (1978) model of how the assertive contents of utterances are determined by

---

[7] For more on the distinction between properties in the worldly sense, in which they can be identified with natural kinds, and properties in the linguistic sense, in which they are often taken to be meanings of complex predicates, see my "What Are Natural Kinds?" (Soames n.d.).

the interaction of the semantics of the sentences uttered with the presuppositions of conversational participants, obvious features of the context, and the pragmatics of assertion. The model's chief positive lesson is that even when metaphorical and other special uses of language are put aside, there is often a substantial gap between what one asserts and the semantic content of the sentence one uses to assert it. Stalnaker's linguistic task is to articulate principles capable of filling this gap. His philosophical task is to use those principles to reconcile Kripkean instances of the necessary a posteriori with his treatment of propositions as functions from possible world-states to truth-values, and his restriction of the epistemically possible to the metaphysically possible. My aim in the essay is to explain how the model is supposed to work, and why, in the end, it doesn't—while distinguishing what can be salvaged from what can't.

For Stalnaker, the aim of discourse is to distinguish the actual world-state from other world-states compatible with everything previously assumed or established. The function of assertion is to narrow down this set of actual-world-state candidates by eliminating those incompatible with what is asserted. When a sentence S semantically expresses a necessary truth, assertion of that truth would fail to eliminate any world-states, and so be pointless. Thus, Stalnaker thinks, uttering S results in the assertion of a different proposition. Typically this is the so-called "diagonal proposition"—the (unique) proposition that is true (false) at a candidate-state w iff the proposition S would express were w to be actual would be true (false). This is the deflationary core of what was to become the 2D account of instances of the necessary a posteriori.

In my essay, I argue that Stalnaker's deflationary 2D explanation is defeated by the inability of his model to accommodate *de re* beliefs and assertions. If these are allowed, the world-states needed for the assignment of diagonal propositions to utterances predicating essential properties of individuals or kinds turn out to be either metaphysically impossible, or incompatible with what has already been assumed or established in the conversation. Both alternatives violate central tenets of the model. In the end, there is no satisfactory solution. Although the range of counterexamples can be reduced by countenancing epistemically possible world-states that aren't metaphysically possible, the basic problem can still be re-created. Nor is it feasible to exclude *de re* beliefs and assertions altogether. Although doing so might render world-states needed for suitable diagonal propositions compatible with the non–*de re* beliefs and assumptions of conversational participants, such an exclusion would mischaracterize the information carried by utterances. Nor can such an exclusion be internally justified, since the *de re* knowledge of world-states presupposed by the model rests on the very *de re* knowledge of individuals and kinds that would have to be excluded.

Thus, the basic structure of the model must be given up. Instead of taking world-states to be basic and propositions to be sets of such states, we should take propositions and their constituents to be basic, and think of world-states as deriving from them. Instead of restricting epistemic possibility to metaphysical possibility, we should recognize the former as outstripping the latter. Instead of aspiring to a deflationary account of the necessary a posteriori, we should embrace the metaphysically robust Kripkean account. Finally, we should recast the rules of Stalnaker's discourse model in a way that reflects all this, while preserving his insight that pragmatics has an important role to play in filling the gap between semantic content and assertion. The essay closes with some steps in this direction.

In essay 9, "Ambitious Two-Dimensionalism" (2007), I offer a broad overview of attempts by opponents of the antidescriptivist revolution of the 1970s to use the technical apparatus of two-dimensional modal logic to reinstate descriptivism in semantics, conceptualism in our understanding of the modalities, and linguistic analysis as the core of our general philosophical methodology. To that end, I articulate and criticize four versions of this ambitious philosophical and linguistic program— Robert Stalnaker's pragmatic version from the late 1970s, a strong semantic version suggested in the mid-1990s by Frank Jackson and David Chalmers, a weak semantic version that is a natural retreat from the strong version, and a hybrid version suggested by Chalmers in 2002.[8] I argue that all these systems fail, and that the central ideas motivating them are incorrect.

That said, an interpretive caveat must be noted. The main texts used in arriving at what I call "strong" and "weak" two-dimensionalism— Chalmers's *The Conscious Mind*, and Jackson's *From Ethics to Metaphysics*—do not explicitly state or unequivocally endorse either system. However, that is only because they don't explicitly state or unequivocally endorse any system. The main area of inexplicitness concerns the semantic analysis of knowledge, and other attitude ascriptions. Although Chalmers and Jackson offer no precise analyses, the contentious conclusions they draw about the necessary a posteriori require 2D analyses of these constructions—about which their texts contain many suggestive hints.[9] My strong and weak 2D systems make these analyses explicit. However, since the analyses are refuted, the 2D conclusions about the necessary a posteriori are left unsupported. Since Chalmers and Jackson don't explicitly endorse the analyses, strictly speaking, their texts have

---

[8] See Chalmers (1996); Jackson (1998); and Chalmers (2002). The ambitious philosophical motivations of Chalmers and Jackson were shared by Stalnaker only in part.

[9] See Soames (2005) for textual documentation.

not been refuted—but only because their key arguments were insufficiently explicit to support any far-reaching conclusions.

Essay 10, "Actually" (2007), presents a theory of the metaphysics and epistemology of actuality and possibility, and the language we use to talk about them. World-states—which are consistent, maximally informative properties attributed to the universe—include the actual world-state, which is instantiated, metaphysically possible states, which could have been instantiated, and epistemically possible world-states, which cannot be known a priori not to be instantiated. The contents of these properties are knowable a priori. However, it is argued, empirical knowledge of the actual world-state also arises when it is presented to us indexically. This duality is the key to understanding the actuality operator, and to solving important puzzles about the necessary a posteriori and the contingent a priori.

Whenever S expresses a contingent truth p, ⌜Actually S⌝ expresses the necessary truth *that p is true at @*.[10] However, since ⌜Actually S⌝ is trivially inferable from S, and since the proposition it expresses often doesn't seem knowable in any other way, it has seemed to be knowable only a posteriori, whenever p is. But this is problematic. If p is true at @, then the proposition *that p is true at @* is true, not just at every metaphysically possible world-state, but at every epistemically possible state as well. Why, then, if there are *no* possible world-states at which this proposition is false, is empirical evidence required to know it?

I argue that it isn't. World-states are properties of making certain world-describing sets of propositions true. Imagine, then, a tiny universe consisting of two blocks side by side, with a third on top. This world-state, *Tiny*, is *the property of containing blocks 1 and 2 side by side, with block 3 on top*. We can know, just by thinking about this property, that if it were instantiated, then block 3 would be sitting on blocks 1 and 2. So, when p is the proposition that block 3 is sitting on those blocks, it is knowable a priori *that p is true at Tiny*. This point generalizes to real-world inquiries in which the relevant world-states are finitely specifiable. For every such state w, and every proposition p the truth of which is calculable from the basic propositions defining w, the proposition *that p is true at w* is knowable a priori. This result applies to the actual world-state (relative to an inquiry), as much to any other. Thus, the propositions expressed by uses of ⌜Actually S⌝ are often *knowable* a priori, even when they are not, in fact, *known* a priori. Since the actual world-state, @, relative to many inquiries, will be much more complex than Tiny, we may not be able grasp it in the nondemonstrative way we grasp Tiny—in which case we won't be able to calculate the truth-values of propositions

[10] '@' designates the actual world-state—that is, the world-state of our present context.

from a complete specification of @. In such cases, our only practical way of learning *that p is true at* @ is by inferring it from p. So, when p is a posteriori, our knowledge of the proposition expressed by ⌜Actually S⌝ may be a posteriori, even though *what we know* can, in principle, also be known in another way.

The proposition *that p is true at* @ is entertainable in two radically different ways. One way—which, as a practical matter, may exceed our cognitive abilities—involves grasping the propositional content of @. One who entertains the proposition in this way can know it a priori—by deriving p from the propositions that define @. But when @ is presented in this way, there is no way of knowing that it is instantiated. Hence, when one entertains the proposition *that p is true at* @ in a way that allows one to know it a priori, there is nothing in one's knowledge that allows one to infer p from it. The second, indexical, way of entertaining the proposition *that p is true at* @—which is how it is presented using the actuality operator—doesn't involve grasping the full propositional content of @. When presented with the proposition in this way, we can't determine it to be true a priori, though we can move a priori from it to p, and vice versa. Since on neither way of knowing *that p is true at* @ is there a way of establishing p a priori, p is knowable only a posteriori.

The contingent a priori gives rise to a related puzzle. I have argued that the function of empirical evidence needed to know *Kripkean* examples of the necessary a posteriori is to rule out epistemically possible, but metaphysically impossible, world-states in which they are false. This may seem to suggest (11).

(11)   If p is false at some epistemically possible world-state, then p isn't a priori. So, if p is a priori, then p is true at every epistemically possible world-state.

But then, if p is contingent a priori, it will follow that p is true at all epistemically possible world-states, while being false at some metaphysically possible state. So, if (11) is true, some metaphysically possible world-states are epistemically impossible. This, I argue, is incorrect.

The puzzle can't be solved by denying the contingent a priori—since, as I show, when S is a contingent truth anyone who, at @, knows the a priori truth expressed by ⌜S iff S⌝ is in a position to derive the contingent ⌜S iff actually S⌝ by steps that can be known a priori to be truth preserving. If this is right, then the only way to solve the puzzle is by denying (11). Although the details of the argument are intricate, the basic point is simple. The proposition *that p iff p is true at* @ is knowable a priori by agents evaluating it at @—when it is presented to them by ⌜S iff actually S⌝—because they know that the world-state @ it is

about is the world-state at which they are evaluating it. The fact that it fails to be true when evaluated at world-states *distinct from* @ is irrelevant. No empirical evidence is required to rule out such states because the agents know in advance that they are not evaluating it there. Thus, (11) is false.

So is the closely related principle (11+)

(11+)   If it can be known a priori both (i) that p is false, and (ii) that p is true at w, then it can be known a priori that (iii) w is not instantiated (and so is epistemically impossible).

The falsity of this principle is due to the fact that a priori consequences of propositions that are knowable a priori are sometimes not themselves knowable a priori. In this case, the proposition *that it is false that (~p and p is true at @)*—playing the role of (i) in (11+)—can be known by us a priori, when @ is presented indexically—as it is in ⌜It is false that (~S & actually S)⌝. By contrast, the proposition *that the proposition that (~p and p is true at @) is true at w*—playing the role of (ii) of (11+)—can be known a priori only when it is known a priori that p is true at @. This requires @ to be presented nonindexically, in terms of its propositional content. Since there is no way of merging the a priori routes to (i) and (ii) into a single a priori route to (iii), an agent can't derive (iii) from a priori knowledge of (i) and (ii), and (11+) fails.

Having disposed of puzzles about the necessary a posteriori and the contingent a priori, I turn to a puzzle about language. On the one hand, prefixing the actuality operator to a contingent, a posteriori truth results in a dramatic change in semantic content (from contingent to necessary and from a posteriori to a priori). On the other hand, adding it in ordinary conversation often seems to be a purely rhetorical move, with no effect on what is asserted. For example, the difference between uttering, "Actually, I live in California," and "I live in California" lies not in what is asserted (which is the same), but only in the suggestion, carried by the first utterance, that my assertion may be unexpected. My explanation of this follows naturally from the indexical semantics of the actuality operator that gives it its modal, epistemic, and semantic punch. The essay ends with a discussion of broader linguistic, epistemic, and metaphysical challenges.

Part 3: Truth and Vagueness

Essay 11, "What Is a Theory of Truth?" (1984), explains, and gives a limited defense of, the philosophical significance of Tarski's theory/definition

of truth. The defense depends on taking the notion he defined as a replacement for our ordinary truth predicate that neither captures the meaning of 'true', nor provides a notion suitable for explicating meaning, or advancing contentious philosophical theses. The virtues of Tarski-truth are (i) that it avoids the liar paradox, (ii) that it provides a precisely defined basis for mathematically serviceable notions such as *arithmetical truth*, *definability*, *logical truth*, and *logical consequence*, (iii) that it uses semantic ascent to enrich the expressive power of metatheoretical investigations, and (iv) that it is transparently deflationary—thereby removing the dangerous temptation to read contentious epistemological and metaphysical views into what should be a philosophically neutral notion. The price to be paid for this progress is the frank recognition of the irrelevance of Tarski-truth to reductionist programs like physicalism, to the project of giving theories of meaning for interpreted languages, and to the practice of providing interpretations for uninterpreted languages by stipulating the truth-conditions of their sentences.

The key point is that statements of the Tarski-truth conditions of sentences *provide no information* about their meanings. For example, when 'T-in-L' is a Tarskian truth predicate, (12a) and (12b) are interderivable using elementary truths of logic, set theory, and the syntax of L.

(12) a. 'Ws' is T in L iff snow is white.
      b. Snow is white iff snow is white.

Since knowing the latter provides no information about meaning of 'Ws', knowing the former doesn't either. Although this makes Tarski-truth useless for Davidson's project using truth to explain meaning, for Tarski it is the inevitable by-product of something good. Since the only commitments carried by ⌜S is T-in-L⌝ not carried by S itself are those involving syntax and elementary set theory, we are guaranteed that if L is unproblematic, and syntax and set theory are too, then introducing 'T-in-L' can't cause problems. The defense of Tarski on this point is that Davidsonian semantic theories are not what we want anyway. Thus, the incompatibility of Tarski-truth with these theories of meaning is not, in itself, a black mark against Tarski.

A related defense is given against Hartry Field's criticism that Tarski's theory doesn't do enough to reduce semantic facts to physical facts. The defense is that such a reduction is irrelevant to Tarski's project, which *doesn't do anything* to advance it. If one distinguishes genuine semantic facts (having to do with meaning) from facts stateable using Tarski-truth, and related notions, then—except for his embrace of abstract objects (sets, expression types, etc.)—the latter pose no problem

for physicalism.[11] What about genuine facts about the meanings of the expressions uttered by speakers? In the essay, I play with the idea that if we take meanings to be essential properties of expressions, then we might be able to give a purely formal semantic account—linking expressions with entities they express or designate—while confining philosophically and empirically significant issues to questions about which abstractly characterized expressions speakers should be regarded as using. I am not convinced that this is the best way of thinking of the matter, only that it is a defensible one for formalists about meaning to adopt.

The essay closes with an understated admission of the limitations of my defense of Tarski. I don't think that Tarski-truth is sufficient for all theoretical purposes for which we need a truth predicate. Although I don't accept Davidsonian-style theories of truth and meaning, I believe we do need non-Tarskian notions of truth in semantics that are much closer to our ordinary notion than any he provided. I also believe that more work on truth has yielded, and will continue to yield, better accounts of the liar. However, I appreciate Tarski's deflationary perspective, and remain convinced that deflationism is crucial to any adequate understanding of truth.[12]

Essay 12, "Understanding Deflationism" (2003), is devoted to spelling out what deflationism is. Roughly put, a predicate that applies to all and only the true sentences of a language L is a deflationary truth predicate iff the claim that it applies to a sentence of L is a trivial, necessary, and a priori consequence of the claim made by the sentence itself, and vice versa. In this sense, a Tarskian truth predicate for a restricted fragment E of English for which it can be defined is deflationary. However, the one-place predicate 'is a true sentence of E', formed by filling the second argument place of the relational predicate 'is a true sentence of' that occurs in ordinary English is not. For some purposes, the deflationary truth predicate for E is sufficient; for others, the nondeflationary truth predicate—which can be defined in terms of propositional truth—is required. This raises the

---

[11] Tarski was not as clear as one might have hoped about the distinction between genuine semantics and semantics based on his notion of truth. In the essay, I note in passing that he recognized that his notion of truth-in-L can't be used to give the meanings of the truth-functional connectives in L. However, I fail to note that other comments he makes (Tarski 1944, 1969) indicate an insufficient appreciation of the chasm separating his truth predicate from the ordinary English predicate, and of the related distinction between genuine facts about meaning and facts stateable using his truth predicate. For discussion see Soames (1999b, 238–44); also Soames (1995).

[12] See Soames (1999b, chaps. 4, 6, and 8); also essays 5, 7, and 8 of volume 1 of this collection.

issue of truth for propositions. To say that propositional truth is deflationary is, I suggest, to say (i) that p and the proposition that p is true are trivial, necessary, and a priori consequences of one another, and (ii) that any warrant for believing, assuming, denying, or doubting one is warrant for taking the same attitude toward the other. In this sense, I argue, propositional truth *is* deflationary.

Although this conclusion is, I think, both true and informative, it doesn't provide an explanation of what it is to understand the predicate 'true', when applied to propositions. Since I appeal to necessary and a priori consequence—which are themselves defined in terms of truth—my claim that propositional truth is deflationary doesn't explain the truth predicate in a way that can be grasped by someone who doesn't already understand it. Thus, it's not an analysis of what truth is, or what 'true' means. It is simply a philosophically important truth about truth.

In giving my favorable assessment of deflationism, I identify its chief philosophical opponents as nonfactualists. It is these philosophers—who *deny* that warrant for p is warrant for the claim that p is true (rather than those who merely add contentious doctrines to deflationism connecting truth to what we can know, or what it is useful for us to believe)—who are the real inflationists about truth. Nonfactualists accept (believe, assert) certain propositions p (about one or another philosophically contentious domain) while rejecting (disbelieving, denying) corresponding claims to the effect that p is true. Interestingly, these philosophers don't deny that *the claim that p is true* is a necessary consequence of p. Instead, they maintain, this only shows that they are required to accept *the truth of the claim that p is true* if they accept *the truth of p*—which they don't. In so doing, they reject the idea that one who accepts p, while recognizing that q is a consequence of p, is thereby committed to q. I argue that this rejection is self-defeating.

It also misses the explanatory power of deflationism. We are interested in the relation of logical consequence among sentences, and necessary consequence among propositions, because we take them to be connected to the argumentative commitments we take up when we accept a set of premises. The reason why a definition of consequence in terms of guaranteed truth-preservation serves this interest is that we take the central deflationist point for granted: namely that acceptance of S (and the proposition it expresses) carries with it a commitment to the truth of S (and the proposition it expresses), and vice versa. If we didn't take this for granted, then recognizing an argument to be truth preserving would not give us guidance about what acceptance of its premises committed us to—which, of course, it does. Thus, deflationism about truth is central to our practice of using logical and necessary consequence to track our argumentative commitments.

Essay 13, "Higher-Order Vagueness for Partially Defined Predicates" (2003), explains what the higher-order predicate 'is determinately red' has in common with the partially defined predicate 'is red'.[13] To say that the latter is partially defined is to say that the rules governing its application provide sufficient conditions for it to apply to an object, and sufficient conditions for it not to apply, but no conditions that are both individually sufficient and disjunctively necessary for it to apply, or not to apply. This results in a three-way classification of objects. Those to which the rules of the language, plus the underlying nonlinguistic facts, determine that 'is red' applies are in the determinate-extension of the predicate. Those to which the rules plus facts determine that it doesn't apply are in its determinate antiextension. Those for which the predicate is undefined, because the rules plus facts determine neither result, are excluded from both classes.

Next consider the predicate 'is determinately red', which applies to objects in the determinate-extension of 'is red', and does not apply to objects not in the determinate-extension. If every object is either in that determinate-extension or not, then 'is determinately red' is totally defined, and so, on my account, is not vague. That seems wrong. Not only is there no sharp and precise line dividing the objects to which 'is red' applies from those to which it doesn't, there also seems to be no sharp and precise line dividing the objects for which it is *determined*, by the rules of the language and the underlying facts, that 'is red' applies from those for which this is not determined. Thus, there is reason to resist the claim that 'is determinately red' is totally defined.

For some putative rules it is indeterminate whether or not they are rules of the language governing 'is red'. Let R be the class of such putative rules. For certain objects o, the question

Q. Is the claim *that 'is red' applies to o* a necessary consequence of the rules of the language governing the predicate, plus the underlying facts?

---

[13] 'Is red' is also context sensitive. Its *default determinate-extension* is the set of things to which the rules of the language, together with underlying nonlinguistic facts, determine that it applies. Its *default determinate antiextension* is the set of things to which the rules of the language plus underlying facts determine that it doesn't apply. For all objects o, 'is red' is undefined for o just in case o is in neither of these classes. Since these classes don't exhaust all cases, speakers have the discretion of adjusting the extension and antiextension so as to include initially undefined cases. Often they do this by explicitly predicating 'is red' of an object o, or by explicitly denying such a predication. When this happens, and other conversational participants go along, the extension (or antiextension) of the predicate in the context is adjusted so as to include o, plus all objects that bear a certain relation of similarity to o.

can be answered only by assuming that certain members of R *are* rules of
the language governing 'is red', or by assuming that they *aren't*. Since
neither of these assumptions can be established, there is no possible justi-
fication for accepting them. Thus, we should reject both the claim that
these objects are determinately red, and the claim that they are not. Like
the application of 'is red', the application of 'is determinately red' is sub-
ject to a range of vagueness and indeterminacy.

Of course, this range is smaller in the latter case than in the former.
Can it be further reduced by additional iterations of the determinately
operator? The range of indeterminacy for 'is determinately red' is smaller
than the range of indeterminacy for 'is red'. Is the range of indeterminacy
for 'is determinately, determinately red' still smaller? The answer turns
out to be 'no'. Iterating 'determinately' has no effect. The predicate 'is
determinately red' and 'is determinately, determinately red' are equiva-
lent. They are *weakly partial*, in a sense analogous to the sense in which
'is red' is partial.

It is important not to be misled by this result. Nothing in the semantics
of partial, or weakly partial, predicates shows that the indeterminacy to
which they are prone can't be eliminated. Once we understand them, we
can always introduce artificially defined substitutes that are totally de-
fined. By the same token, nothing in the semantics of partial, or weakly
partial, predicates shows that they don't impose sharp and precise lines
distinguishing different categories of objects. They do. However, the re-
sulting fine-grained classifications are less worrisome than they have
sometimes been thought to be.

They do not, for example, pose the threat to our notion of linguistic
competence posed by a sharp and precise division between the objects
that an ordinary vague predicate is determinately true of and those it is
determinately not true of. The distinction between truth and untruth is
important to us; and the norms of language use presuppose that we are
able (given suitable facts) to track it. The same cannot be said of the dis-
tinction between (i) statements that are true because they are determined
to be so by the rules of one's language, together with nonlinguistic facts,
and (ii) statements for which there is no saying whether they are true for
that reason, or true because speakers have made them true by exercising a
minimal amount of discretion in adjusting the boundaries of the partial,
context-sensitive, predicates employed. If my model of vague predicates is
correct, there may be a sharp distinction between (i) and (ii). However
this distinction, unlike the one between truth and untruth, is highly theo-
retical, and need not be reliably tracked by competent speakers. Thus, the
fact that they are typically oblivious to it is not paradoxical.

Essay 14, "The Possibility of Partial Definition," defends this analysis
of vague predicates against an objection due to Michael Dummett (1978)

and Michael Glanzberg (2003). The objection rests on the view that an assertion of a proposition p is correct (in the sense of satisfying the intrinsic norm of assertion) just in case p is true, and incorrect just in case p is not true. If this conception of assertion is correct, then the impossibility of partially defined predicates follows from (i) and (ii), each of which is integral to the theory of partial definition (in my sense).

(i) Assertions of propositions expressed by sentences attributing partially defined predicates to objects for which they are undefined violate the norm of assertion, and so are incorrect.

(ii) Truth is undefined for the propositions in (i)—in which case, the proper response is to reject both the claim that they are true, and the claim that they are not true.

The reply to this objection is that although (i) and (ii) are correct, the Dummett-Glanzberg conception of assertion is not. Instead, it should be replaced by Timothy Williamson's (1996) conception, according to which an assertion of a proposition p is correct (satisfies the norm of assertion) just in case the agent knows p, and incorrect just in case the agent doesn't know p. After giving reasons for preferring this account, I show how the incorrectness of asserting undefined propositions follows from they fact that such propositions are, in principle, unknowable. Thus, the Dummett-Glanzberg argument against the possibility of partially defined predicates is defeated.

Intuitively, this is the right result. No matter how attractive, or unattractive, partially defined predicates may be for various purposes, surely it is at least *possible* to introduce and use them in certain situations. I flesh out this possibility by elaborating an imaginary case in which the color terms of an isolated community living in a restricted environment are introduced by meaning-giving stipulations that seem, transparently, to be partial. These predicates are meaningfully used for a time, after which the discovery of previously unencountered hues gives rise to a practice in which speakers allow themselves a range of conversational discretion in applying the predicates to items for which they were initially undefined. At this point the predicates become context-sensitive, as well as partially defined.

After describing the semantic properties of these predicates, I bring the example closer to home by allowing the initially clear and precise specification of the rules of the language governing them to become muddied—with a consequent blurring of the boundary between what is in their default determinate-extensions, and what isn't. This results in the kind of indeterminacy explored in essay 13. As with vague predicates of English, so with the context-sensitive, partially defined predicates in my imagined language, there is a limited range of cases in which it is indeterminate

whether the claim that a predicate applies to an object is true *because the rules of the language, plus facts about the object, determine it to be so*, or whether it is true *because speakers have exercised minimal discretion in adjusting the boundaries of the predicate to make it apply to the object*.

What precisely is the range of this indeterminacy? I am now inclined to think the Williamsonian view—that, somehow, the question has a definite answer, even though we can't find it—is correct. If so, then context sensitivity and partial definition don't tell the whole story about vague language. I do, however, continue to believe that they are part of the story. The essay therefore concludes with two arguments—one identifying a theoretical advantage in recognizing vague predicates as partially defined, given that they are context sensitive, and the other showing how, given partial definition, context sensitivity can be used to explain our differing reactions to violations of different "laws" of classical, two-valued logic.

Part 4: Kripke, Wittgenstein, and Following a Rule

Essays 15 and 16—"Skepticism about Meaning: Indeterminacy, Normativity, and the Rule-Following Paradox" (1998) and "Facts, Truth Conditions, and the Skeptical Solution to the Rule-Following Paradox" (1998)—are commentaries on Kripke's (1982) work on Wittgenstein's rule-following paradox, read, not as an interpretation of Wittgenstein, but as the development of a novel skeptical position about intentionality. The position has two parts. The first is a skeptical argument for the paradoxical conclusion that there are no facts about meaning, or representational content. The second is a "skeptical defense" that attempts to save our practice of ascribing meanings, and contents, by stripping it of the presumption that such ascriptions purport to state facts.

In essay 15, I argue that Kripke's skeptical argument is guilty of an equivocation similar to one in Quine's argument for the indeterminacy of translation.[14] In Kripke's case, the equivocation centers on the normativity requirement he places on attempts to specify the putative fact that determines addition as the meaning of '+'. I argue that two different formulations of this requirement—N_E and N_M—are extractable from his discussion.

N_E If the fact that F *determined* that (in the past) one meant addition by '+', then knowing that F should provide one with a sufficient basis for concluding that '125' is the *correct* answer to the

---

[14] I discuss Quine's arguments, and equivocation, in further detail in Soames (1999a; 2003, vol. 2, chap. 10; 2007).

question "What is $68+57$?", and so is the answer one ought to give, provided one intends to use '+' with the same meaning it had in the past. (Similarly for other calculations)

N_M If the fact that F *determined* that (in the past) one meant addition by '+', then one means addition by '+' at any possible world-state at which it is true that F; hence '125' is the *correct* answer, and so the answer one ought to give, to the question "What is $68+57$?", in such a world-state. (Similarly for other calculations)

N_E is plausible, provided that the determination relation in question is given by Det_E.

Det_E P *determines* Q only if, given P, one can demonstrate Q without appealing to any other empirical facts—i.e., only if Q is an *a priori consequence* of P.

Read in this way, the effect of N_E is to restrict potential *meaning-of-'+'-* determining facts *that F* to those knowledge of which would, in principle, allow one to demonstrate that (in the past) one meant *addition* by '+'. Kripke argues that whatever one's past dispositions to calculate using '+' may have been, they do not satisfy this restriction. Since (i) it is plausible to suppose that a similar conclusion can be reached for any nonintentional fact *that F* one might cite, and (ii) intentional facts have previously been dismissed as possible meaning-determining facts (on the grounds that they are subject to the same skeptical argument), Kripke's skeptic reaches the conclusion that no facts determine the meaning of '+'.

Unfortunately for the skeptic, what the argument establishes is not particularly skeptical, or paradoxical. The basic point—that truths about meaning are not *a priori consequences* of nonintentional truths—is philosophically interesting, but not paradoxical. In contrast to the claim that statements about meaning are not *necessary consequences* of nonintentional truths, it does not threaten the very existence of truths about meaning. This is where N_M comes in. When the normativity requirement is understood in this way, it is plausible, provided that the determination relation in question is given by Det_M.

Det_M P *determines* Q only if Q is a *necessary consequence* of P.

Read in this way, N_M restricts potential *meaning-of-'+'-determining* facts to those which have the claim that (in the past) one meant addition by '+' as a *necessary consequence*. If the skeptic can show that no nonintentional facts satisfy this restriction, then—since, surely, facts about what one means can't simply float free of underlying nonintentional facts about one's history, environment, linguistic behavior, and brain-states—the skeptic may plausibly maintain that *there are no facts whatsoever about*

what one means by '+'. However, now he faces a different problem. Although many candidates for the nonintentional fact determining addition as the meaning of '+'—including some simple dispositional facts—fail to satisfy N_M, it is far from obvious that this is true for all such candidates. On the contrary, it is highly plausible to suppose that some appropriately global candidates do satisfy the requirement. Since the Kripkean skeptic does nothing to undermine this supposition, his argument fails.

In essay 16, I turn my attention to Kripke's so-called skeptical solution. What makes the position skeptical is that it purports to accept the conclusion that there are no facts about meaning; what makes it a solution is that it nevertheless defends the correctness of ascriptions like 'Soames means addition by '+''. I argue that there are two interpretations of the position, corresponding to two interpretations of what is meant by 'facts'. According to the first, minimal, interpretation, all instances of the schema, (13), are embraced.

(13) It is fact that S iff it is true that S iff the proposition that S is true iff S

On this interpretation, the skeptical solution claims that meaning ascriptions like (14) don't express propositions, or purport to state facts, but rather have some other, nondescriptive, kind of meaning.

(14) In the past, I meant addition by '+'.

I argue that, when understood in this way, the skeptical solution is indefensible, and ultimately self-defeating—in addition to being unsuitable for certain key argumentative purposes of Kripke's Wittgenstein.

According to the second interpretation, what is denied is not that there are meaning facts in the minimal sense of (13), but that there are facts about meaning, truth, and reference the grasp of which by speakers plays a central role in *explaining* how words come, initially, to be endowed with meaning, and how, later, they are understood. The target here is the *classical truth-conditional conception of meaning*. According to it, all words (and sentences) are endowed with meaning by *conceptually prior* intentions to use them with certain reference (and truth) conditions. Once they have acquired their meanings, one who understands them does so *in virtue of* recognizing these conditions. I argue that there is an important element of truth in the second interpretation of the skeptical solution, which disputes this explanatory picture. For some words and sentences, one does not understand them *because* one knows their reference and truth conditions. Instead, one knows those conditions, in part, *because* one understands them—where understanding crucially involves not just one's private semantic beliefs, but also one's position in a broader linguistic community.

There is much in the argument I give for this that fits Kripke's text. However, there are also important differences. Whereas the critique of the classical truth-conditional conception he offers is global—purporting to show that *no* expressions conform to it—mine is local, establishing only that *some* don't. In fact, my critique doesn't apply to Kripke's paradigmatic example of '+'. This points to a larger interpretive problem. To argue effectively against the classical conception one must show that one class of intentional facts—which includes semantic beliefs and intentions—is not always *conceptually prior to*, and *part of the explanation of*, a different class of intentional facts—which includes understanding expressions and using them to mean certain things. Kripke's Wittgenstein doesn't argue in this way. Instead, he looks for a nonintentional basis of all intentional facts, and suggests that there is none. Because of this, his skeptical argument can't be seen as a reductio ad absurdum of the classical truth-conditional conception of meaning.

Thus, in the end, Kripke's text leaves us in a quandary. The most natural interpretation of the skeptical argument makes no room for a reasonable version of the skeptical solution, while the most reasonable version of the skeptical solution can't be seen as a response to the most straightforward reconstruction of the skeptical argument. Though hybrid interpretations are possible, none is fully satisfying. The reason for this, I suggest, is that there is no single, coherent line of argument running through the entire text. My recommendation is to give up the illusion of a single unifying interpretation, to distinguish different illuminating partial interpretations, and to identify what there is to learn from each. That, at any rate, is what I offer.

REFERENCES

Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.
———. 2002. "Components of Content." In *The Philosophy of Mind: Classical and Contemporary Readings*, ed. David Chalmers, 608–33. New York: Oxford University Press.
Dummett, Michael. 1978. "Truth." In *Truth and Other Enigmas*, 1–24. Cambridge: Harvard University Press. Originally published in *Proceedings of the Aristotelian Society* 59 (1959): 141–62.
Edelberg, Walter. 1994. "Propositions, Circumstances, and Objects." *Journal of Philosophical Logic* 23:1–34.
Glanzberg, Michael. 2003. "Against Truth-Value Gaps." In *Liars and Heaps: New Essays on Paradox*, ed. JC Beall, 151–94. Oxford: Clarendon Press; New York: Oxford University Press.
Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Clarendon Press.

Kripke, Saul A. 1982. *Wittgenstein: Rules and Private Language*. Cambridge: Harvard University Press.

Soames, Scott. 1995. "T-Sentences." In *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*, ed. Walter Sinnot-Armstrong in collaboration with Diana Raffman and Nicholas Asher, 250–70. Cambridge: Cambridge University Press.

———. 1999a. "The Indeterminacy of Translation and the Inscrutability of Reference." *Canadian Journal of Philosophy* 29:321–70.

———. 1999b. *Understanding Truth*. New York: Oxford University Press.

———. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of "Naming and Necessity."* New York: Oxford University Press.

———. 2003. *Philosophical Analysis in the Twentieth Century*. 2 vols. Princeton: Princeton University Press.

———. 2005. *Reference and Description: The Case against Two-Dimensionalism*. Princeton: Princeton University Press.

———. 2007. "What We Know Now That We Didn't Know Then." *Philosophical Studies* 135:461–78.

———. n.d. "Kripke on Epistemic and Metaphysical Possibility: Two Routes to the Necessary Aposteriori." In *Saul Kripke*, ed. Alan Berger. Cambridge: Cambridge University Press, forthcoming.

———. n.d. "What Are Natural Kinds?" Forthcoming in *Philosophical Topics*.

Stalnaker, Robert. 1978. "Assertion." *Syntax and Semantics* 9:315–32.

Tarski, Alfred. 1944. "The Semantic Conception of Truth and the Foundations of Semantics." *Philosophy and Phenomenological Research* 4:341–75.

———. 1969. "Truth and Proof." *Scientific American*, June, 64–77.

Williamson, Timothy. 1996. "Knowing and Asserting." *Philosophical Review* 105:489–523.