Scott Soames

The Substance and Significance

of

the Dispute over Two-Dimensionalism:

Reply to E. J. Lowe, Frank Jackson, and Josh Dever

To Appear In

Philosophical Books

Reply to E. J. Lowe

In answering "No" to his question "Does the descriptivist/antidescriptivist debate have any philosophical significance [beyond semantics]?" Lowe gives what at first sounds like an exciting answer to an interesting question – until one identifies his reason. That reason is the belief – now widely shared -- that a decisive resolution of this semantic debate would *not* allow one, using only secure non-philosophical knowledge, to establish interesting metaphysical principles, beyond philosophical doubt. Though this belief is widespread, the idea that its truth would show the semantics of modality to have no significance for metaphysics (or other areas of philosophy) is preposterous, and, as far as I know, the sole possession of Professor Lowe. Are we to suppose that for any area of philosophy A, and any debated question Q in A, the resolution of Q has no significance for any other part B of philosophy, unless that resolution, absent all other philosophical principles, is sufficient (together with secure non-philosophical knowledge) to establish interesting positions in B, beyond philosophical doubt*?* It is hard to imagine anyone agreeing with that.

Lowe's failure to find the philosophical significance of semantic anti-descriptivism comes from looking in the wrong place. Its importance lies in *expanding* the range of metaphysical hypotheses to be taken seriously, not in limiting debate by proving metaphysical theorems from non-metaphysical premises. The unraveling of Quine's attack on the intelligibility of essentialism is a case in point. That attack, which distorted discussion of the subject from 1943 until well into the sixties, was based on faulty semantic premises about modality, singular terms, and quantification. Particular troublesome were Quine's identification of necessity with analyticity, and his implicitly descriptivist conception of singular terms. It was recognized early on – by Smullyan, Fitch, Marcus, Follesdal, and the young Kripke, among others – that the availability of nondescriptive terms would blunt the attack. However, it wasn't until *Naming and Necessity* that this line of thought came together in a decisive

1

way, when a convincing analysis of ordinary names as non-descriptive rigid designators was paired with an explicitly non-linguistic, metaphysical conception of necessity. The result, though no proof of interesting essentialist theses, refuted Quine's objection to their intelligibility. With this, the dialectical situation reverted to one in which the presumptive truth of much of our counterfactual talk, and the evident plausibility of at least some essentialist intuitions, had to be given their proper weight.[1]

Out of this came Kripke's distinction between necessity and apriority. Using non-descriptive semantics plus a rehabilitated conception of essentialism, he showed how to generate instances of the necessary *aposteriori*. If t is a name or indexical that rigidly designates o, F expresses an essential property of o, and knowledge that o has this property requires empirical evidence, then the proposition expressed by ⌈If t exists, then t is F⌉ is both necessary and knowable only *aposteriori*.[2] Thus, necessity and apriority are different, and the idea that they arise from something beyond the linguistic is vindicated. As I argued in *Philosophical Analysis in the Twentieth Century*, this recognition crucially affects the evaluation of some of the most important arguments in the analytic tradition. This, rather than the fantasy of deriving metaphysical theorems from non-metaphysical premises, is a measure of the philosophical significance of the semantic issues discussed in *Reference and Description*.

The two-dimensionalism of David Chalmers and Frank Jackson challenges the Kripkean story in three ways. First, it treats names and natural kind terms as rigidified descriptions. Second, it identifies metaphysical possibility with conceptual possibility -- which is seen as linguistic in nature.[3]

---

[1] Quine's objection and Kripke's refutation are discussed, and a brief description of the relevant history is given, on pp. 347-354 of Vol. 2 of my *Philosophical Analysis in the Twentieth Century*, (Princeton: Princeton University Press), 2003.

[2] This recipe is stated more precisely, and discussed at greater length in my, "The Philosophical Significance of the Kripkean Necessary Aposteriori," E. Sosa and E Villanueva, eds., *Philosophical Topics* 16, 2006.

[3] Chalmers describes necessity and apriority as different kinds of "truth in virtue of meaning," on p. 62 of *The Conscious Mind*, (New York: Oxford), 1996. On p. 69 of *From Metaphysics to Ethics* (Oxford: Clarendon Press), 1998, Jackson says that instances of the necessary aposteriori involve *analytic, conceptual*, or *logical* necessity.

Third, it offers an ersatz, linguistic account of those examples of the Kripkean necessary aposteriori that are accepted as genuine, while repudiating his metaphysically essentialist explanation. For example, Jackson explicitly rejects the usual Kripkean explanation of

1.   Water = H_2O,

which takes *having as instances only those things that are composed of molecules with two hydrogen atoms and one oxygen atom* to be an essential property of the substance water (that it can be known to have only aposteriori). In its place, he offers an account in which sentence (1) is associated with two propositions -- one given by (2), which we ask about when we ask whether (1) is necessary, and one given by (3), which we ask about when we ask whether (1) is knowable apriori.

2.   Water = Water

3.   The clear, drinkable stuff that falls as rain, fills the lakes and rivers etc. = H_2O

This trivialization of the necessary aposteriori, rejection of essentialism, and substitution of philosophically motivated, but semantically unsupportable, linguistic claims for substantive metaphysical theses was my target in *Reference and Description*.[4]

It's a pity that one with Lowe's zealous regard for the independence of metaphysics didn't see this. The problem may have been his failure to grasp how a Kripkean explanation of the necessary aposteriority of (1) would go -- or at any rate my version of it.[5] The account holds that 'water' is a non-descriptive, directly-referential term designating a substance -- where substances

---

[4] Jackson's rejection of genuine essentialism is a piece with his rejection of Kripkean necessity (pp. 69 and 79 *From Metaphysics to Ethics*). Although 'Water is essentially H_2O' is allowed to come out true, this is only because its truth is reduced to the necessity of (1), which, in turn, is ascribed to the trivial necessity of (2). The result is a linguistic trivialization of what is, in fact, a substantive metaphysical thesis.

[5] For fuller discussions see chapter 10 of *Beyond Rigidity* (New York: Oxford), 2002; "Knowledge of Manifest Natural Kinds," *Facta Philosophica* 6, 2004, 159-181; *Reference and Description*, pp. 58-68, and "Kripke on Epistemic and Metaphysical Possibility," in Alan Berger, ed., *Saul Kripke* (Cambridge: Cambridge University Press) forthcoming.

are taken to be physically constitutive kinds (instances of which share the same basic physical constitution).[6] It is further assumed that a kind of this sort may have different instances in different world-states, and that if a and b are kinds with the same instances in all possible world-states, then a is b. These are clearly metaphysical assumptions, to which we add the natural corollary that for any substance s, if, in some possible world-state, instances of s have a certain molecular structure, then instances of s have that structure in every world-state. In other words, we assume that it is an essential property of a substance that instances of it have the molecular structure they do. From this it follows that (1) is necessary if true, and that 'H_2O' – which I take to be equivalent to 'the substance instances of which have a molecular structure with two hydrogen atoms and one oxygen atom' – is a rigid designator. Being true, (1) is, therefore, necessary. Since knowing the proposition it expresses requires knowing of a certain substance that its instances have a particular chemical structure, (1) is knowable only aposteriori. Note, there is no attempt here to derive a metaphysical truth about essence from non-metaphysical premises. On the contrary, the point of the exercise is to show how compelling, and widely accepted, examples of the necessary aposteriori can be *explained* using plausible essentialist assumptions. The point of *Reference and Description* is that they can't be explained without them.

Reply to Jackson

Frank Jackson focuses on two main issues underlying our dispute over two-dimensionalism.

I₁.   The role of descriptive information in explaining what words mean, and what it is to understand them.

I₂.   The relationship between metaphysical and epistemic, or conceptual, possibility.

---

[6] 'Water' may also be thought of as designating a natural kind of a slightly different sort -- what I call *an explanatory* kind in "Knowledge of Manifest Natural Kinds." The differences between the two accounts is not relevant here.

In discussing I_1, he notes that sentences represent the world, and so, provide information. From this he draws the lesson that "a theory of reference for names is an account of the contribution that names make to how sentences containing them represent things to be" -- a lesson I endorse, provided one carefully distinguishes, as he does not, the semantic content of the sentence from additional information conveyed by uttering it. This distinction is crucial to identifying the meanings of names. The question is, *Are names synonymous with rigidified descriptions that semantically fix their reference?* He says "Yes;" I say "No."

His main argument rests on the very intuitions used against descriptivism. Given a candidate D for a description semantically fixing the referent of n, anti-descriptivists refute the analysis by citing possible scenarios in which the referents of D and n clearly diverge. For us, the combined weight of all such refutations – indicating widespread speaker ignorance and error -- is strong inductive evidence against there being any correct descriptivist analysis. Jackson objects that this can't be right, since our thought experiments about n implicitly rely on intuitions allowing us to identify its referent in relevant scenarios. Surely, he maintains, it is possible, in principle, to extract necessary and sufficient conditions determining n's reference at arbitrary world-states from this intuitive knowledge. This, he insists, is all that descriptivism requires.

There are two main flaws in his argument. First, the ability to refute particular analyses by finding, for each proposed description D, at least one scenario in which the referents of the n and D diverge does *not* presuppose the logically stronger ability (which the argument requires) of correctly determining, *for every possible scenario*, what n refers to there. Second, our ability, such as it is, to identify the referent of n at w depends not just on our knowledge of the meaning of n, but also on pre-semantic and other contextual information -- as well as on how w is described. Thus, descriptive *semantic* conditions on reference cannot simply be read off our referential judgments

about different scenarios. Even when the reference of a term t is *fixed independently of descriptions speakers associate with t,* we may still be able to identify its referents in different scenarios. So, having this ability does *not* presuppose that reference is semantically fixed by description.

This is obvious with indexicals like 'I' and 'this'. Given a scenario, one can identify the referent of 'I' as the user of the expression, and the referent of a use of 'this' as the thing demonstrated by the user. Focusing only on uses of 'I', the unwary theorist might take this to show that its referent is semantically fixed, for a speaker, by the description 'the one producing this token'. Focusing only on uses of 'this', the theorist might take its referent to be semantically fixed for an agent by the description 'the thing I am demonstrating now'. Of course, this can't be. One or both of the referents must be semantically fixed in some other way. Thus, two-dimensionalists exempt 'I' from the thesis that reference is descriptively fixed – which is fine. What isn't fine is their failure to answer Q1 and Q2.

Q1.   Why is taken to be obvious that the attempt to extract semantic, reference-fixing descriptions for names and natural kind terms from our judgments about what they refer to in various possible scenarios won't be defeated by circularities of the sort illustrated by 'I' and 'this'?

Q2.   Why, if one kind of non-descriptive reference fixing (the indexical) is compatible with our ability to identify referents in different scenarios, isn't this ability – to whatever extent we have it -- also compatible with other kinds of nondescriptive reference fixing (including those involving baptisms, communicative chains, and the like)?

Q1 questions whether our intuitions about reference are rich enough to determine a description that semantically fixes the referent of every name and natural kind term. Q2 asks whether the intuitions we do have might represent something more, or other, than knowledge of meaning.

The former worry is illustrated by a speaker s who relies on the description 'the capital city of Honduras' to identify the referent of 'Tegucigalpa', while relying on 'the country in which

Tegucigalpa' is located to identify the referent of 'Honduras'. Guided by these descriptions, s can identify the referent of either name in scenarios described using the other. However, s can't semantically fix the referent of each name by appeal to the corresponding description. Nor can the referents of the pair be simultaneously fixed by the descriptive information at s's disposal – which is merely that one term designates the capital city of the referent of the other. Although this example is artificially simple, the problem it poses isn't. To prevent circularity from ever blocking the extraction of semantic, reference-fixing descriptions from intuitions about Kripke-style thought experiments, Jackson must maintain either (a) that for every name, natural kind term, and related expression, we can correctly determine its referent in possible scenarios *described without using any such terms whatsoever*, or (b) that there are some such expressions for which we can do this, which can then be used in describing scenarios from which our intuitions allow us to extract reference-fixing descriptions for additional expressions (until all names, natural kind terms, and related expressions are covered), or (c) that using all our intuitions about what refers to what in different scenarios, however described, we can extract a single statement, *expressible without any names, natural kind terms or related expressions*, that semantically fixes the referents of all such expressions simultaneously. Since it has never been shown that (a), (b), or (c) are feasible, Jackson's argument fails. The difference between the widely acknowledged intuitions about Kripkean thought experiments, to which anti-descriptivists appeal, and the strengthening of those intuitions needed to extract semantic, reference-fixing descriptions is crucial. Since the former do *not* place heavy restrictions on how the scenarios in which we identify referents are described, while the latter do, the existence of the former doesn't guarantee existence of the latter.

Q2 points to a different problem – the possibility that intuitions about reference-change across different scenarios might reflect something other than constant descriptive meanings that

7

generate different extensions in different contexts. Consider, for example, scenarios involving 'and'. It is easy to recognize scenarios in which it is clear that speakers are using 'and', not to stand for the familiar truth function for conjunction, but something else – e.g. the truth function for disjunction. Suppose, for the sake of argument, that we could extract from intuitions about these cases an extension-determining description D that picked out the conjunctive truth function in some possible contexts, and the disjunctive function in others. Would this show that 'and', as we use it, is equivalent to the rigidification of D, that its extension is semantically fixed to be whatever in the context of utterance satisfies D, or that it retains a single (descriptive, indexical) meaning across contexts in which its extension varies? Of course not. No such description is part of the meaning of 'and'. To the extent that such a description is extractable from intuitions about what 'and' expresses in different possible scenarios, it incorporates pre-semantic conditions (of which speakers have some awareness) that are responsible for the fact that 'and' has the non-descriptive meaning it does. By analogy, the anti-descriptivist maintains that the suitability of names and natural kind terms for baptizing individuals and kinds, and the ability to acquire these terms from others, are *not* parts of their descriptive meanings, but rather are pre-semantic facts that determine the simple, non-descriptive meanings they really have. Since nothing in Jackson's argument rules out a pre-semantic, or pragmatic, explanation of our reference-identifying intuitions, he would not be justified in concluding that names and natural kind terms are *semantically* descriptive, even if he could do what he, in fact, cannot – namely, show that fully fledged reference-determining descriptions are extractable from those intuitions.

What do speakers know about the use of names to refer to things – e.g., to people? They know various things – e.g. that children (conventionally) share the last name of a parent, that parents typically choose children's first names, that one can legally change one's name, that nicknames can

be acquired by a more informal process, and so on. Though knowledge of these facts is common, it is *not* a condition of semantic competence. While unusual, a competent speaker *could* use and understand names without it. What competence with a name n requires is simply acquisition of a referential intention determining an object o as referent, and awareness that to assertively utter ⌈n is F⌉ is to say of o that it "is F". Arguably, a third condition can sometimes be added specifying a general property of o – like being a human being – knowledge of which is required for competence with n.[7] On the whole, however, it is striking how austere the conditions for understanding names are. One can acquire the needed reference-determining intention either (a) by speaking a language in which n is already used to name o, and intending to use n as a name in accord with the reference it already has, or (b) by being acquainted with o and introducing n to name it. These conditions are special cases of the general abilities (a_g) to pick up and use any word w with the meaning it already has in one's linguistic community, provided one satisfies minimal conditions for competence with w, and (b_g) to introduce a new word, provided one is capable of passing it on to those with whom one interacts. Although (a) and (b) are crucial to the use of names, they are *not* parts of their meanings. Nor do they provide speakers with precise, non-circular knowledge of how reference is fixed.

Speakers do, of course, typically realize that their use of a name *somehow* inherits its reference from other uses – without having any precise knowledge of how this occurs.[8] One can say, if one wishes, that such a speaker associates n with the description D, 'the referent of those uses of n from which my present use (somehow) inherits its reference', which, by definition, denotes

---

[7] See chapters 3 and 5 of *Beyond Rigidity.*

[8] For example, speakers need not possess information about the full scope and limits of how names may be introduced, or when, precisely, communicative chains preserve the reference of a term, and when they don't. For discussion, see 366-71 of Volume 2 of *Philosophical Analysis in the Twentieth Century*, and on pp. 299-302 of *Reference and Description.*

what n does. However, D does not semantically fix the referent of n. For that to be so, satisfaction

of D would have to be the means by which reference is established -- which it can't be, since D

determines a referent for a use of n only if that use has inherited its reference independently.[9] When

circularity-generating descriptions of this sort are avoided, the best one can come up with are often

descriptive approximations that pick out the right referent for n in many, but not all, cases.

However, the incompleteness of this knowledge doesn't prevent uses names and related terms from

carrying information. All the information carried by utterances cited by Jackson can be grasped by

speakers who combine their knowledge of the non-descriptive meanings of these terms with pre-

semantic and other pragmatic information they associate with them .[10]

---

[9] Jackson misses a version of this point when he comments on *my* remark that "an otherwise competent English speaker is
counted as a competent user of the word ['water'], if he or she knows that 'water' is a term that stands for some natural kind
that determines its extension at different world-states – even if one doesn't have any reliable way of describing that kind,
other than the kind that the word stands for in English …" (my emphasis, *Reference and Description,* 183-184) About this,
Jackson says, "under what conditions would it be correct for someone to use the word 'water' to describe how they take
their world to be in a sentence like 'There is water nearby'? Surely just when they believe that … 'some natural kind ….
that is the kind the word ['water'] stands for in English …' is nearby. *This is exactly the kind of account I* [FJ] *like. Soames
has given us his version of 'watery'* [Jackson's supposedly semantic, reference-fixing description for 'water']. *One might
or might not like the detail of his account but the principle is the same.*" (my emphasis) Here, Jackson confuses meta-
linguistic beliefs that speakers have when they accept certain sentences with the semantic (representational) content of those
sentences in the language. The proposition about the expression 'water' that Jackson describes the speaker as believing is
not the semantic content of the sentence containing 'water' that the speaker accepts. Similarly, satisfaction of 'the natural
kind that 'water' stands for in English' cannot be the semantic means that determines the reference of 'water' in English.

[10] One of Jackson's examples is 'The conference is in London', about which he says: "The way this sentence enables those
who hear it … get to London relies on two facts: the knowledge of the hearers … that i) they should go to somewhere
called 'London', and ii) that various other sentences they come across like 'London is served by the following air lines …'
are a reliable source of information about London, something resting, as English speakers know, on information-preserving
causal chains running from London to the sentence tokens they come across. The way the folk use sentences containing
proper names … as a source of knowledge about our world, shows that they know about naming practices (baptisms) and
about how our response to them generates information-generating chains." This is overstated. All that is required to extract
the needed information is (a) knowledge of the semantic fact that 'London" refers to London in the language of the

I now turn to Jackson's discussion of I_2, and its role in his broader 2D account of modality. His argument for this account is essentially as follows:

(i)   The information carried by an utterance of a sentence S (its representational content) is information about how the speaker represents the world to be.

(ii)   This representational content must be understood as the set of world-states at which S is true.

(iii)   However, for many sentences S (ii) is acceptable only if either (a) the world-states include some that are epistemically possible, but metaphysically impossible, or (b) there are two different dimensions of evaluation, one (needed to understand what is asserted by ordinary utterances of S) yielding the set of metaphysically possible states taken as contexts of utterance at which S expresses a truth, and the other (needed for counterfactual evaluations of S) yielding the set of metaphysically possible states at which S is true, relative to a fixed context of utterance. (b) is the 2D option.

(iv)   The claim that there are epistemically possible, but metaphysically impossible, world-states is incoherent. So, (iiia) is not an option.

(v)   Therefore, the 2D treatment of modality in (iiib) is correct.

Jackson's commitment to (i) and (ii) is signaled by a remark in which he implicitly identifies the proposition expressed by a sentence with the set of (metaphysically) possible world-states in which it is true – a view notorious for its stunningly counterintuitive results. [11]

community and (b) the non-semantic ability to recognize occurrences of 'London' in different sentences. If it is objected that 'London' is ambiguous, counting as a name for several different cities, we might add (c) the ability to spot when 'London' is being used with one of its meanings rather than another. But knowledge of when an ambiguous word is being used with one, rather than another, meaning is not constitutive of any of these meanings. Rather, it presupposes that the word has its meanings already. Finally, Jackson's vague generalizations about baptisms and causal chains are instances of the platitudes that all words one encounters have somehow been introduced (baptisms broadly construed), and acquired their meanings by being used in the same way by members of the linguistic community. To think of all of this as knowledge of what goes into the meaning of a name like 'London' is not to know what meaning is.

[11] Scott Soames, "Direct Reference, Semantic Content, and Propositional Attitudes," *Philosophical Topics*, 15, 1987, 47-87.

"If you say 'There is something round nearby', I know how you are representing things to be and I could arrange matters so that things were as you were representing them to be … [so as] to make the sentence you produced true … So how you represent things to be and the possibilities at which your sentence is true are one and the same."

More important, however, is the argument's treatment of examples like

4.   That paperweight is made of wood (if it exists),

which predicate an essential property to an object. (Suppose, for the sake of argument, that the material constitution of a thing is essential to it.) An utterance U of (4) in which a paperweight p that is in fact made of wood is demonstrated expresses a proposition that is true at all metaphysically possible world-states. How then can U be informative? The possible answers are, roughly: (a) that U is informative because accepting what is asserted allows one to rule out, as not obtaining or "being actual," metaphysically impossible, but *epistemically* possible, world-states at which p isn't made of wood, vs. (b) that U is informative because accepting U as expressing a truth allows one to rule out *metaphysically* possible world-states in which (4) is uttered demonstrating a different paperweight (≠ p) that isn't made of wood. Jackson argues that (a) is incoherent.

"He [SS] takes the case of a paperweight that is in fact made of wood but, for all you know, might be made of metal or plastic. He describes this as a case where it is metaphysically impossible but epistemically possible that the paperweight be made of plastic … But what precisely is the world-state in question that is both metaphysically impossible and epistemically possible? Let's ask, first, what is supposed to be metaphysically impossible. Presumably, the possibility of that very paperweight being made of plastic. *But, as being made of wood is an essential property of that paperweight, that world-state is that of something made of wood being made of plastic, and that's epistemically impossible ....*" [my emphasis]

12

Incredibly, this argument assumes that an object o that has an essential property P at *the actual world-state* must have P at every *epistemically* possible world-state at which o exists. This presupposes that an essential property of o is one that o has at every *epistemically* possible world-state at which o exists. However, as *Reference and Description* makes abundantly clear, that is *not* what an essential property is. Rather, it is a property that o has at every *metaphysically* possible world-state at which o exists. More simply, it is one that o *couldn't have lacked*. Since it doesn't follow from the fact that o couldn't have lacked P that it is knowable apriori that o has P, world-states in which o lacks P are epistemically possible.

I suspect that Jackson's failure to recognize this elementary point is due to a fundamental confusion about what world-states really are. As I argue on pp. 200-202, and 206-209, they are *not* the alternate concrete universes of Lewisian fantasy, but maximal properties that the one and only universe can be conceived as having. Ignoring this, and thinking of world-states in the former way, Jackson may find it incoherent to suppose that there is a Lewisian universe at which an object o really exists without having property P – when P is a property that is, in fact, necessary for o's existence. But this just reflects his inability to transcend the flawed Lewisian conception. There is nothing incoherent about supposing that there is a certain maximal property M, attribution of which to the universe involves taking it to contain a state of affairs in which o lacks P. Although M is a way the universe couldn't really have been (a metaphysically impossible world-state), it is a way that the universe can coherently be conceived to be, and that the universe cannot be known apriori not to be. Hence it is an epistemically possible world-state. The reason that Jackson's argument against such states fails so miserably -- the reason it fails even to address the relevant issues -- is that he simply doesn't understand the conception of modality of which it is a part.

Because of this, there is, for him, no alternative to the problematic 2D account of representational content given in (iiib). On that account, a use of S in a context C is associated with a pair of propositions (identified with sets of metaphysically possible world-states). These are the proposition expressed by S at C (S's so-called C-intension at C), and a (centered) proposition true at all and only those (centered) world-states at which S expresses a truth (S's so-called A-intension). When the C-intension predicates an essential property of an object, and so is necessarily true, this intension is trivial, and the representational content of the use of S in C is identified with it's A-intension, instead. According to Jackson, this proposition both represents the way the speaker believes the world to be, and is asserted and conveyed by the speaker's remark.

This is the idea behind the following remark.

"He urges that it is very implausible to hold that he, Soames, does not know the proposition expressed by sentences containing his name, 'Scott Soames', and those of his two sons, 'Greg Soames' and 'Brian Soames'. But what Soames plausibly does know is what he is representing about how things are when he uses such sentences [their A-intensions]. But the view of mine he is attacking is that he does not know the proposition expressed *in the sense of the worlds at which the various sentences are true* [the C-intensions]. And that view of mine *is* plausible: fathers know their children and themselves well, but do not know their essences, and that's what's required to know the worlds at which … 'The Soames's are all composed in part of H_2O molecules' is true [when evaluated according to its C-intension]. The example, far from being a problem one for me, makes the point nicely that we must distinguish how things are being represented to be, a sentence's representational content [its A-intension], from the worlds at which the sentence is true in cases like these [its C-intension]."

According to this story, what I first wonder about, later learn, and finally use

5.   The Soames's (Scott, Greg, and Brian) are all composed in part of H_2O molecules

to assert is *not* the C-intension of (5) -- a proposition knowledge of which is knowledge of me, Greg, and Brian that we are composed in part of H_2O molecules – but rather the A-intension of (5) – a proposition also expressed by (6) (where 'D_{GS}', 'D_{BS}', and 'D_{SS}' stand in for descriptions I allegedly use to determine the referents of the names of my sons and me).

6. D_{GS}, D_{BS}, and D_{SS} are all composed in part of H_2O molecules.

For this account to work, the descriptions must designate Greg, Brian, and me in the actual world-state, while designating different people at some other possible states – individuals who perhaps look and sound a lot like us, and share our other descriptive characteristics, but nevertheless are not composed in part of H_2O molecules. Are there such *metaphysically possible* individuals and world-states? Although it is not obvious there are – a point on which I E.J. Lowe and I seem to agree – such individuals and states are crucial for Jackson. Without them, his restriction of epistemically possible world-states to the metaphysically possible would have the unacceptable consequence that both the C and A intensions of (5) are necessary – and so, (for him) uninformative. Surely, however, the informativeness of my remark shouldn't depend on the answer to a deep and contentious metaphysical question.

There is also another problem. Even if the deep and contentious metaphysical question is ultimately answered in Jackson's favor, his account of what I first wonder about, then learn, and finally assert by uttering (5) is patently false. My progress through this process requires me to acquire evidence ruling out possibilities falsifying what I eventually assert. For Jackson, these can only be possibilities in which *individuals other than Greg, Brian, and me* contain no H_2O. But surely, there are *some* cases in which what I come to be in a position to assert by uttering sentences containing these names requires evidence that rules out possibilities involving Greg, Brian, and me, rather than any artful dodgers. Equally surely, if there are such cases, then (5) is among them. Thus, Jackson's 2D account of representational content fails.

Reply to Dever

Josh Dever targets a class of arguments given in chapter 10 against a particular type of 2D theory called *Strong Two-Dimensionalism,* designed to provide deflationary accounts of the contingent apriori and the necessary aposteriori. According to Strong 2D, an indexical sentence S is assigned both a primary intension – which is a proposition that is true at a world-state w iff the proposition expressed by S at w is true at w – and a secondary intension – which is the proposition delivered by applying the character of S to w. When the primary intension is necessary and the secondary intension is contingent, S is contingent-apriori. When the modalities are reversed, S is necessary-aposteriori. Instead of distinguishing two kinds of possibility – metaphysical and epistemic -- with respect to which a single proposition can differ, Strong Two-Dimensionalism identifies the two, while associating distinct propositions (that come out true at different world-states) with individual indexical sentences. Since the account is supposed to apply to all instances of the contingent apriori and the necessary aposteriori, names and natural kind terms are analyzed as *actually*-rigidified descriptions (and hence as indexical). Crucially, modal operators and predicates are defined as applying to the secondary intensions of their sentential arguments, while epistemic operators and predicates are defined as applying to the primary intensions of their sentential arguments – as indicated by the following assumptions from chapter 10.

T5a. ⌈It is a necessary truth that S⌉ is true with respect to a context C and world-state w iff the secondary intension of S in C is true with respect to all world-states w* that are possible relative to w. Similarly for other modal operators.

T5b. ⌈It is knowable apriori that S⌉ is true with respect to C and w iff the primary intension of S in C is knowable apriori in w; ⌈x knows / believes that S⌉ is true of an individual i with respect to C and w iff in w, i knows / believes the primary intension of S with respect to C. Similarly for other epistemic operators.

The arguments that Dever focuses on -- which he aptly dubs *Nesting Arguments* – involve the interaction of modal and epistemic operators in sentences containing both.   These arguments assume, for simplicity, (i) that the only indexical in the language is the actuality operator, (ii) that both contexts and circumstances of evaluation are simply metaphysically possible world-states, so the semantic theory evaluates sentences relative to pairs of such states, and (iii) that propositions are sets of world-states. The aim of the arguments is to show that Strong Two-Dimensionalism gives a false account of English, and that this falsity includes the falsity of the conjunction of T5a and T5b. As I will demonstrate, nothing in Dever's critique challenges this conclusion. Instead, he questions how far the Nesting Arguments can be generalized. His main conclusion, with which I agree, is that these arguments do not refute all possible 2D systems obtained by abstracting away from the specific dimensions of meaning attended to, the nature of the indices over which they are defined, and the type and complexity of the operators targeting these different dimensions. Nevertheless, the Nesting Arguments against Strong Two-Dimensionalism do refute their intended target.

The arguments have roughly the following form. Let M be a modal operator – e.g. *it is a necessary truth that ___*, or *if it had been the case that ___, then it would have been the case that ___*. Let E be an epistemic operator – e.g. *it is knowable apriori that ___*, or *Jones knows / believes that ___*. Let S_1 and S_2 be sentences with the same primary intensions (according to Strong 2D), but different secondary intensions (with respect to a given context C). For example, let S_1 contain a non-rigid description $\lceil$the x: Dx$\rceil$ that denotes a certain individual (in the context C), and let S_2 arise from S_1 by substituting either the rigidified description $\lceil$the x: actually Dx$\rceil$, or a name n that is (according to Strong 2D) synonymous with $\lceil$the x: actually Dx$\rceil$, for one of more occurrences of $\lceil$the x: Dx$\rceil$ in S_1. (The standard Kaplan-semantics for *actually*, given on pp. 27-30 of *Reference and Description*, is assumed in Strong 2D.) The arguments then proceed as follows:

Step 1   Since S_1 and S_2 have the same primary intensions, and since, according to T5b, epistemic operators operate on the primary intensions of their arguments, $\lceil ES_1 \rceil$ and $\lceil ES_2 \rceil$ have the same primary intensions. This follows from the fact that for any world-state w, $\lceil ES_1 \rceil$ is true with respect to the pair <w,w> (in which w functions both as context and circumstance of evaluation) iff $\lceil ES_2 \rceil$ is true with respect to <w,w> iff the application of the operation associated with E in w to the primary intension of S_1 / S_2 (which does not change from world-state to world-state) produces a truth in w.

Step 2   By the analogous reasoning, $\lceil ES_1 \rceil$ and $\lceil ES_2 \rceil$ have the same secondary intensions. This follows from the fact that for any pair of world-states w_1 and w_2, $\lceil ES_1 \rceil$ is true with respect to <w_1,w_2> (in which w_1 functions as context and w_2 as circumstance of evaluation) iff $\lceil ES_2 \rceil$ is true with respect to <w_1,w_2> iff the application of the operation associated with E in w_2 to the primary intension of S_1 / S_2 (which does not change from world-state to world-state) produces a truth in w_2.

Step 3   From Step 2, together with the semantics of modal operators given by T5a, it follows that $\lceil ES_1 \rceil$ and $\lceil ES_2 \rceil$ are substitutable under M without change in truth value.

Step 4   Since it is a datum of English that such substitution sometimes changes truth value, Strong Two-Dimensionalism is false. (Further examination of particular Nesting Arguments reveals that this falsity includes at least the conjunction of T5a and T5b.)

Although these arguments are sound, Dever raises two critical points about them. The first is that they rely crucially on compositionality to justify the substitutivity of equivalent sentences involving one type of operator (the E-operators in the above argument) under the scope of operators of the other type (the M-operators). Thus, he sees denial of compositionality as a possible escape route for the two-dimensionalist, and cites an idea by Chalmers as a potential way of implementing

18

it. The idea is to treat occurrences of 'actually' that are within the scope of an epistemic operator like 'believes' -- which is itself within the scope of a modal operator -- as designating "the world [state] of the believer," rather than the world-state of the context. About this, Dever says:

> "Since whether an expression is in a combined modal/epistemic context [i.e. whether it is embedded under both modal and epistemic operators] is a non-local feature of that expression, no semantic theory which computes meanings locally, and hence adheres to compositionality, can make the behavior of "actual" sensitive to the presence of such a [linguistic] context. Absent compositionality, ϕ and ψ can carry exactly the same semantic value – and even be fully interchangeable in simple 0_1 contexts [i.e. when embedded under an operator sensitive only to the first dimension of meaning] – but produce radically different results when further embedded in $\square_2$ contexts [i.e. when embedded under an operator sensitive only to the second dimension of meaning], allowing the two-dimensionalist to avoid the incorrect or inconsistent truth value assignments in … the Nesting Argument. The strength of the Nesting Argument can thus be no greater than the justification for implementing a compositional semantic theory."

This is, I think, misleading. Although there is a valid point here, it is highly abstract, and irrelevant to the arguments against Strong Two-Dimensionalism. The valid point is that it is formally possible to introduce a non-compositional operator A* such that $\lceil ES \rceil$ and $\lceil EA*S \rceil$ have the same primary, and even secondary, intensions -- even though the two are not always substitutable under a higher operator M. However, this point is irrelevant to my use of the Nesting Arguments, because one can't do this in a way that saves Strong Two-Dimensionalism as an adequate semantic theory of English.

There are several reasons for this. First, the compositionality needed by the Nesting Arguments against Strong Two-Dimensionalism are built into the Strong-2D semantics for modal, epistemic, and rigidifying operators like 'actually'. Thus, any relaxation of compositionality is thereby a revision of the Strong-2D system. Second, no revision of 'actually' (or any other rigidifier) along the lines suggested by Chalmers can preserve the aims of Strong Two-Dimensionalism.[12] The point is illustrated by (7), where 'actually*' is understood as incorporating Chalmers's suggestion about 'actually', and hence (7b) is equivalent to (7c).

7a.   There exists an x (Aristotle) such that necessarily one believes that *Aristotle* was a philosopher only if one believes something the truth of which requires x to be a philosopher.

7b.   There exists an x (Aristotle) such that necessarily one believes that *the x: actually* Fx* was a philosopher only if one believes something the truth of which requires x to be a philosopher.

7c.   There exists an x (Aristotle) such that for all worlds w, in w the following is true: one believes of w that the unique thing that was F in it was a philosopher only if one believes something the truth of which requires x to be a philosopher.

The problem is that since (7a) is true, whereas (7b) and (7c) are false, the Strong-2D view that names are synonymous with rigidified descriptions can't be maintained -- if the descriptions are rigidified using 'actually*', or any comparable non-compositional operator.[13] Since rigidification using standard, compositional rigidifiers simply reinstates the Nesting Arguments, there is no way of saving the Strong-2D accounts of the contingent apriori and the necessary aposteriori. Finally, names aside, a true premise corresponding to (7a) can be constructed in which 'the individual who actually founded formal logic' is substituted for 'Aristotle', thereby indicating that 'actually*' gives

---

[12] For Strong 2D it doesn't matter whether the rigidifying operator is 'dthat' or 'actually'.

[13] See pp. 46-49 of *Beyond Rigidity* for a discussion of the semantic properties of 'actually*'.

the wrong account of 'actually'.[14] In sum, although non-compositional operators could, in principle, be employed to meet the challenge posed by certain highly abstract versions of the Nesting Argument for certain types of 2D-systems, this strategy is of no use proponents of Strong Two-Dimensionalism, or, more generally, to two-dimensionalists with the philosophical motivations of Chalmers and Jackson.

Dever's second criticism of the Nesting Arguments grants compositionality, while attempting to block the move from the sameness of primary intensions of S_1 and S_2 to the sameness of both primary and secondary intensions of $\lceil ES_1 \rceil$ and $\lceil ES_2 \rceil$, and their intersubstitutivity under modal operators. Abstracting from both the details of Strong Two-Dimensionalism, and my arguments against it, he puts the point as follows:

> "The 1-equivalence [equivalence of primary intension] of ϕ and ψ certainly grounds *some* important semantic relation between $0_1\phi$ and $0_1\psi$ [where 0_1 is an operator on the primary intension of its argument], but I see no argument that the relation must extend so far as both 1-equivalence and 2-equivalence [sameness of primary and secondary intensions].   In particular, I see no reason to think that $0_1\phi$ and $0_1\psi$ must be 2-equivalent [have the same secondary intensions]. If they are not 2-equivalent, then there is no reason to think that they are (1- or 2-) substitutable in $\Box_2$ contexts [i.e. when embedded an operator sensitive only to the second dimension of meaning], and the Nesting Argument fails."

Again, Dever's point is misdirected. The reason he "sees no argument" that $\lceil ES_1 \rceil$ and $\lceil ES_2 \rceil$ must have the same secondary intensions, and so be intersubstitutable under modal operators like 'it is a necessary truth that', is that his abstract versions of the Nesting Arguments ignore central

---

[14] For defense and elaboration of the thesis that the standard philosophical semantics of the rigidifier 'actually' is the correct semantics for the English word 'actually', see my "Actually," available at rcf.usc.edu/~soames.

defining features of the theory – Strong Two-Dimensionalism – that they are arguments against. The

proof of the sameness of secondary intensions given in Step 2 above makes use of the Strong-2D

treatment, T5b, of epistemic operators as operating on primary intensions, thought of as sets of

metaphysically possible world-states. The proof of substitutivity under modal operators given in

Step 3 relies on the Strong-2D account of those operators expressed by T5a. Given these

assumptions, substitutivity follows. It is only by implicitly relaxing or ignoring them that Dever can

raise his objection.

This point is nicely illustrated by his example involving the interaction of 'it is a necessary

truth that' with Stalnaker's dagger operator -- taken as mapping the character of its sentential

argument, S, onto the primary intension of S (i.e. the set of world-states w such that the proposition

that S expresses at w is true at w). It is quite true that two sentences, ϕ and ψ *with different*

characters but identical primary and secondary intensions in the sense defined in Strong Two-

Dimensionalism, can be such that (i) $\lceil$it is a necessary truth that $\phi\rceil$ and $\lceil$it is a necessary truth that $\psi\rceil$

have the same secondary intensions (relative to a certain context), even though (ii) they do not have

the same primary intensions, and (iii) they are not intersubstitutable under the dagger operator. But

this is irrelevant to my use of the Nesting Arguments – and not just because my arguments first

embed the two sentences under an epistemic operator, and then further embed them under a modal

operator (rather than, as Dever does, reversing the order of the embeddings). The example is

irrelevant to my use of the Nesting Arguments because the dagger operator *does not operate on*

primary intension (or secondary intension), as defined in Strong Two-Dimensionalism.

What makes this point easy to miss is that two-dimensionalists -- including Chalmers,

Jackson and others -- sometimes speak of primary intensions as propositions defined in terms of

characters, while at other times they speak of them as simply characters (typically non-constant

functions from contexts to propositions) These two conceptions are strikingly different, and, as I show in *Reference and Description*, sliding from one to the other often reflects an uncritical vacillation between two quite different semantic theories -- which I call *Strong* and *Weak Two-Dimensionalism*. Dever's objection applies to neither. In order for it to work, (i) ϕ and ψ must have the same primary intensions, despite having different characters, and (ii) the dagger operator must operate on primary intensions. In order for (i) to hold, characters can't be primary intensions, in which case (ii) fails. In order for (ii) to hold, characters must be primary intensions, in which case (i) fails. Either way, there is no valid objection.

Nevertheless, there is a positive lesson in Dever's observations, to which I wholly subscribe. Thought of, not as a substantive philosophical theory, but as an abstract semantic framework encompassing different dimensions of semantic evaluation, while allowing the definition of both single and multi-dimensional operators, two-dimensionalism is to be embraced and explored, rather than discouraged. I agree with Dever that the Nesting Arguments pose an interesting challenge to theory-construction in this abstract framework, not a refutation of the entire class of such theories. They were never intended as such. They do, however, refute their target -- the philosophically loaded theory of Strong Two-Dimensionalism.

As for Weak Two-Dimensionalism – according to which epistemic operators apply to the *secondary* intensions of their sentential arguments (which are believed, known, etc. in virtue of the attitudes agents bear to characters that express them) – the arguments I offer to refute it are disjoint from the Nesting Arguments directed against Strong Two-Dimensionalism. To this can be added a third, hybrid, version of ambitious two-dimensionalism, according to which epistemic operators apply to pairs consisting of the primary and secondary intensions of their sentential arguments. I give further arguments to show that this version also fails. At this point, one has a choice. One

can search still further, for more convoluted, versions of two-dimensionalism to account for the contingent apriori, the necessary aposteriori, and the like, or one can look for a different, non-2D, explanation. *Reference and Description* offers just such an explanation. It is the conjunction of all these negative arguments with this positive explanatory alternative that makes me think that the philosophically-laden theories of two-dimensionalists like Chalmers and Jackson can't be salvaged. This doesn't mean that other, productive, uses won't be found for abstract two-dimensionalist frameworks. As we have learned, however, one must be rather more careful in spelling out what they are.