Soames on Two-Dimensionalism
David Chalmers
Commentary at Arizona State University, January 23, 2004

Scott Soames describes and argues against "pragmatic" two-dimensionalism (attributed to Stalnaker), "strong" two-dimensionalism (attributed to Chalmers and Jackson), and "weak" two-dimensionalism (unattributed). I reject all of these. Before saying why, I'll outline the core claims of two-dimensionalism as I see it. I'll then discuss strong and weak two-dimensionalism, Soames' arguments against them, and whether these arguments provide any reason to reject the core claims of two-dimensionalism.

Core Two-Dimensionalism

(1) Every expression token (of the sort that is a candidate to have an extension) is associated with a primary intension, a secondary intension, and a two-dimensional intension. A primary intension is a function from scenarios to extensions. A secondary intension is a function from possible worlds to extensions. A two-dimensional intension is a function from (scenario, world) pairs to extensions (or equivalently, a function from scenarios to functions from worlds to extensions).

Here the extension of a sentence token is a truth-value, the extension of a name is an individual, and so on. Scenarios are most often understood as centered possible worlds – i.e. ordered triples of worlds, individuals, and times - although they may also be understood in other ways.

Terminology: when the primary intension of S is true at scenario V, we can also say that V verifies S, or that S is true at V considered as actual. When the secondary intension of S is true at world W, we can also say that W satisfies S, or that S is true at W considered as counterfactual.

(2) The extension of an expression token S is identical to the value of the primary intension of S when evaluated at the actual scenario of S (e.g., a world centered on the speaker and the time associated with S). Likewise, it is identical to the value of the secondary intension of S when evaluated at the world in which S is uttered.

(3) A sentence token S is metaphysically necessary iff the secondary intension of S is true at all worlds.

(4) A sentence token S is a priori (epistemically necessary) iff the primary intension of S is true at all scenarios.

Consequences:

(5) A sentence token S is necessary a posteriori iff the secondary intension of S is true at all worlds but the primary intension of S is false at some scenario.

(6) A sentence token S is contingent a priori iff the primary intension of S is true at all scenarios but the secondary intension of S is false at some world.

(7) 'A==B' is metaphysically necessary iff A and B have the same secondary intension.

(8) 'A==B' is a priori (epistemically necessary) iff A and B have the same primary intension.

Here 'A==B' is a statement that is true iff A and B are coextensive. E.g. if A and B are sentences, then this is equivalent to 'A iff B'. If A and B are singular terms, then this is equivalent to 'A = B' (modulo worries about empty extensions).

Note that (8) is reminiscent of the Fregean thesis: 'A==B' is cognitively insignificant iff A and B have the same sense. Both entail that there is a semantic value that is constitutively connected to the epistemic domain. For example, given that identities such as 'Hesperus is Phosphorus' are not a priori (one may want to build something like this in to core two-dimensionalism also), it follows that 'Hesperus' and 'Phosphorus' have different primary intensions.

Soames' "Strong Two-Dimensionalism"

ST1. Each sentence is semantically associated with a pair of semantic values – its primary intension, and its secondary intension. Its primary intension is a proposition which is true with respect to all and only those contexts C to which the Kaplan-style character of S assigns a proposition that is true at C. When contexts are identified with world-states, and propositions are taken to be sets of such states, the primary intension (proposition) associated with S is the set of world-states w which are such that the character of S assigns to w (considered as a context of utterance) a set of world-states (i.e. a proposition) that contains (i.e. is true at) w. The secondary intension of (or proposition expressed by) S at a context C is the proposition assigned by the character of S to C.

I reject this, because I hold that

(i) two-dimensional intensions are not Kaplanian characters
(ii) scenarios are not contexts of utterance
[(iii) intensions are not propositions]
[(iv) the association is with sentence tokens, not sentence types]
[(v) the force of "semantically associated" is not clear]

ST1* [a minimally modified acceptable version of ST1]
Each sentence token is associated with a pair of semantic values – its primary intension, and its secondary intension. Its primary intension is a true with respect to all and only those scenarios V to which the two-dimensional intension of S assigns a secondary intension that is true at the world w corresponding to V. The secondary intension of S is the secondary intension assigned by the two-dimensional intension to the scenario V.

Note: One must distinguish between *contextual* and *epistemic* understandings of two-dimensional semantics (see "The Foundations of Two-Dimensional Semantics").

On the contextual understanding (Kaplan, Stalnaker), the objects at which primary intensions are evaluated are contexts of utterance, and primary intensions encapsulate the *context-dependence* of an expression's extension on the world. The primary intension of S, evaluated at context C, is

true iff a token of S, as used within context C, would be true. E.g. the primary intension of 'water' picks out what 'water' would refer to as used within context C. This approach cannot ground core two-dimensionalism: necessity of primary intension comes apart from apriority.

On the epistemic understanding (Chalmers, maybe Jackson), the objects at which primary intensions are evaluated correspond to (maximal) epistemic possibilities, and primary intensions encapsulate the *epistemic dependence* of an expression's extension on the world. In the first instance, the primary intension of S, evaluated at scenario V, is true iff: under the hypothesis that V is actually the case, the speaker should judge that S is the case.

Example: *'water'*. For all we know a priori, it could be that water is H_2O and it could be that water is XYZ. Both the H_2O-scenario and the XYZ-scenario represent epistemic possibilities for us. We are in a position to know: *if* the H_2O-scenario is actual, then water is (actually) H_2O. And *if* the XYZ-scenario is actual, then water is (actually) XYZ. So the primary intension of 'water is H_2O' is true at the H_2O-scenario and false at the XYZ-scenario.

[These are clearly intuitively acceptable indicative conditionals, satisfying the Ramsey test. One can further argue that the corresponding material conditionals are a priori: justification of the belief they express does not essentially depend on empirical knowledge. See Chalmers and Jackson, "Conceptual Analysis and Reductive Explanation".]

Example: 'Godel'. Let V be a scenario where someone called 'Schmidt' proved incompleteness but someone named 'Godel' stole it and became known to us as the prover. We know that if V is actual, then Godel didn't prove incompleteness. That is: if we were to accept that V is actual, we would rationally conclude that Godel didn't prove incompleteness. So: the primary intension of 'Godel proved incompleteness' is false in V. With respect to V, the primary intension of 'Godel' picks out the stealer, not the prover.

Note that primary intensions (epistemically defined) are quite different from characters (contextually defined). It may be that terms such as 'Godel' and 'water' have their referents essentially: if so, they have constant character. However, this is quite compatible with the claim that they exhibit the nontrivial epistemic behavior exhibited above, and so have nontrivial primary intensions.

#ST2. *Understanding S consists in knowing its character and primary intension. Although this knowledge, plus complete knowledge of a context C, would give one knowledge of the secondary intension of – i.e. the proposition expressed by -- S in C, one does not always have such knowledge of C. Since we never know all there is to know about the context / world-state C, sometimes we don't know precisely which proposition is expressed by S in C. However, this does not prevent us from using S correctly in C.*

I reject this as I don't know what it is to "know" an expression's primary intension. Perhaps there is some weaker relation one can stand in: e.g. one grasps a primary intension P with one's

use of S when for all scenarios V, one's (rationally idealized) judgments about whether S is the case given that V is the case would yield the value of P at V. (Likewise, mutatis mutandis, for secondary and two-dimensional intensions.) I'm also reluctant to make claims about understanding (though Jackson makes such claims).

ST2* [highly tentative]: When one understands S, one grasps its primary intension and its two-dimensional intension. Grasp of a two-dimensional intension, plus complete knowledge of one's actual scenario A, would put one in a position to grasp the secondary intension of S. However, one does not always have such knowledge of A, so one is not always in a position to grasp the secondary intension expressed by S. However, this does not prevent us from using S correctly in C.

#ST3a. Examples of the necessary aposteriori are sentences the secondary intensions of which are necessary, and the characters of which assign false propositions to some contexts. Since the character of such a sentence sometimes assigns propositions to contexts that are false in those contexts, the primary intension of such a sentence is contingent.

ST3b. Examples of the contingent apriori are sentences the secondary intensions of which are contingent, and the characters of which assign true propositions to every context. Since the character of such a sentence always assigns propositions to contexts that are true in those contexts, the primary intension of such a sentence is necessary.

The first sentences of these are OK if one replaces 'characters" by 'two-dimensional intensions', 'propositions' by 'secondary intensions', and 'contexts' by 'scenarios'. I'd reject the "since" claim in the second sentence if this is read to suggest explanatory priority, as I don't think that two-dimensional intensions are explanatorily prior to primary intensions.

ST3a* Examples of the necessary aposteriori are sentences the secondary intensions of which are true at all worlds, and the primary intensions of which are false at some scenario. (Consequently, there are scenarios such that when the two-dimensional intension of such a sentence is evaluated at that scenario, it assigns a secondary intension that is false at the world corresponding to that scenario.)

ST3b* Examples of the contingent apriori are sentences the secondary intensions of which are false at some world, and the primary intensions of which are true at all scenarios. (Consequently, when the two-dimensional intension of such a sentence is evaluated at any scenario, it assigns a secondary intension that is true at the world corresponding to that scenario.)

ST4a. All proper names and natural kind terms have their reference semantically fixed by descriptions not containing any (uneliminable) proper names or natural kind terms.

ST4b. These names and natural kind terms are synonymous with context-sensitive, rigidified descriptions (using dthat *or* actually*).*

I reject these claims (see "On Sense and Intension", 2002, sections 8 and 9) because

(i) The primary intension of an expression need not be encapsulable in a description. That is, there may be no description not containing names or natural kind terms whose primary intension is the primary intension of the original expression. (The two-dimensionalist may take this to be a moral of Kripke's epistemic arguments.) What matters is the intension, not any associated description.

(ii) Different tokens of the same name may have different primary intensions. (E.g., 'Julius' as it passes through the community. Or: 'Bill Smith = William Smith' may be a priori for one speaker but not for another. Cf. Frege's view of senses in natural languages.) So even if the primary intension of name tokens could be encapsulated in descriptions, these descriptions would vary between tokens of a name, and there would be no synonymy between names (qua expression types) and rigidified descriptions.

(iii) It is plausible that the referent of a name is essential to it, whereas the referent of a rigidified description is not. This tends to suggest that the name and the rigidified description are not synonymous (they'll have different Kaplanian characters, for example), though they may be a priori equivalent for a speaker and their primary intensions may be the same.

ST4*: In those cases where a primary intension of a name (as used by a speaker) can be encapsulated in a description D, then the speaker's claim 'N. if it exists, is the actual D' will be a priori and necessary (and such that it is a priori that it is necessary).

ST5a. **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 (metaphysically possible) world-states w* that are possible relative to w.*

This is acceptable.

ST5b. **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 modal and epistemic operators.*

I reject these claims (see "The Components of Content", 1995/2002, section 8): the view that 'x believes that S' is true if the subject stands in a belief relation to the primary intension of S is clearly hopeless. E.g.

> (i) Stew believes that I am Australian, but he does not stand in a belief relation to the primary intension of 'I am Australian. ' (If he did, he'd believe that he is Australian).
>
> (ii) Twin Oscar stands in a belief relation to the primary intension of 'water is wet', but he does not believe that water is wet (he believes that twin water is wet).
>
> (iii) My utterance 'Pierre believes that London is pretty' may be true, even though Pierre's term 'Londres' has a primary intension quite different from that of my term 'London'.

The same goes for other attitude ascriptions, e.g. knowledge ascriptions and (ipso facto) a priori knowledge ascriptions. [So Soames' interpretive point (ii), which attributes such a view of a priori knowledge ascriptions to me, is mistaken, as is the corollary (iii).]

I do hold:

ST5a*: A sentence token S is a priori iff the primary intension of S is true at all scenarios.

However, the claim that a token S (produced by speaker x) is a priori is quite different from the claim that the speaker knows a priori that S. On this token-relative conception of apriority: a token S is a priori when it expresses a priori knowledge for the speaker, or when it expresses a thought (a propositional attitude) that is justifiable a priori. (Cf. Kripke's use of 'S is a priori for a speaker'.)

The apriority of a token can come apart from the truth of a corresponding attitude ascription, e.g as follows.

> (i) Stew's token of 'If I exist and am located, I am here now' is arguably a priori. But it is clearly false that Stew knows a priori that if I exist and am located, I am here now.
>
> (ii) A token of 'the actual president is the president', uttered in a counterfactual world where Gore is president, is a priori. But (arguably) the speaker does not know a priori that the actual president is the president [Soames].
>
> (iii) On some views it could be that Pierre knows a priori that London is in England (say his term 'Londres' was introduced to him as a descriptive name for the capital of England), even though his token of 'London is in England' is not a priori.

#ST6a. S is an example of the necessary aposteriori iff the secondary intension of S (with respect to C) is a necessary truth, but the primary intension of S is contingent and, though knowable, not knowable apriori. ...

ST6b. S is an example of a contingent apriori truth iff the secondary intension of S (with respect to C) is true, but not necessarily true, while the primary intension of S is necessary and knowable apriori. ...

These are acceptable with appropriate adjustments along the same lines as before.

Corollaries:

ST7a. There is no proposition that is both necessary and knowable only aposteriori; nor is there any proposition that is contingent yet knowable apriori.

ST7b. The necessary aposteriori and the contingent apriori are, in effect, linguistic illusions, born of a failure to notice the different roles played by primary and secondary intensions in modal and epistemic sentences.

ST8. A proposition is necessary iff it is knowable apriori.

I reject all of these claims (see "On Sense and Intension", section 7). Lewis and Jackson accept versions of them, because they identify propositions with sets of worlds (i.e. with intensions). My view is that if one *stipulates* that a proposition is a set of worlds (as I did in *The Conscious Mind*), then there might be reason to accept the corresponding versions of ST7a and ST8. But if one is not making a stipulation, then I think one should reject the claim that a proposition is a set of worlds. In particular, if one regards a proposition as something like the complete semantic

value of an utterance and as a bearer of epistemic and modal properties, then it is plausible to say that the proposition expressed by 'water is H_2O' is both necessary and a posteriori.

If propositions are to be modeled two-dimensionally, then they will need to involve (at least) complex two-dimensional structure. It could be that they involve more than two-dimensional structure, and that propositions cannot be adequately represented in a possible-worlds structure. On such a view propositions may determine associated intensions but will not be reducible to such intensions. Or it could be that the notion of 'proposition' is somewhat indeterminate, with no single object that can play all the roles that propositions are supposed to play.

A two-dimensionalist need not be committed to any specific claim about the nature of propositions. A two-dimensionalist may well be a *semantic pluralist* (as I am). Utterances can be associated with all sorts of quasi-semantic values via all sorts of quasi-semantic relations, and there may be no deep fact about which of these is "the" semantic value of the utterance.

Soames' "Weak Two Dimensionalism"

#WT1- WT3b, WT5a.

These theses are acceptable with the same sorts of modifications made to ST1, ST2, ST3a, ST3b, ST5a.

WT4a. All proper names and natural kind terms have their reference semantically fixed by descriptions not containing any (uneliminable) proper names or natural kind terms.

WT4b. These names and natural kind terms are synonymous with descriptions rigidified using actually *or* dthat*.*

These theses are to be rejected for the same reasons that apply to ST4a and ST4b.

WT5b. Standardly, an attitude ascription **x v's that *S***, *taken in a context C, is true of an individual a with respect to a possible world-state w iff there is some meaning (character) M such that (i) in w, a bears R to M, and (ii) M assigns the* ***secondary intension*** *of S relative to C to a related context with a as agent and w as world-state. So propositions are objects of the attitudes, and attitude verbs are two-place predicates of agents and their objects. However, this two-place relation holds between an agent a and a proposition p in virtue of a three-place relation holding between a, a character, and p. To believe p is to accept a character M that expresses p, and to believe that M expresses a truth. To know a true proposition p is to justifiably accept a character M that expresses p, and to know that M expresses a truth.*

I reject this thesis: the view that 'x believes that S' is true iff x stands in a belief relation to the secondary intension of S has obvious and familiar problems. For example, 'Clark Kent' and 'Superman' may have the same secondary intension, but 'Lois believes that Superman can fly' is plausibly true while 'Lois believes that Clark can fly' is plausibly false.

A more plausible thesis ("The Components of Content", 1995/2002, section 8) requires that x be related to the relevant secondary intension under an *appropriate* primary intension, where what counts as an appropriate primary intension may be determined by various contextual factors (intuitively, the concept of Clark/Superman involved in Lois's belief has a primary intension that is appropriate to satisfy 'Superman' ascriptions but not 'Clark' ascriptions). Here one can invoke the approximate structure of "hidden-indexical" theories of belief ascription, with primary intensions playing the role of modes of presentation.

WT5b* (tentative). Standardly, an attitude ascription ***x v's that S*****, taken in a context C, is true of an individual a with respect to a possible world-state w iff there is some two-dimensional intension M such that (i) in w, a bears R to M, and (ii) M assigns the secondary intension of S relative to C to a scenario centered on a in w, (iii) M determines an appropriate primary intension. [N.B. it's best to work with structured intensions here.]**

Summary

The ambitious two-dimensionalist can make the key claims (1)-(8) while denying that

(i) two-dimensional intensions are Kaplanian characters
(ii) names are synonymous with rigidified descriptions
(iii) there are no necessary a posteriori propositions
(iv) belief ascriptions ascribe a relation to primary intensions
(v) belief ascriptions ascribe a relation to secondary intensions.

Soames' Arguments Against Strong Two-Dimensionalism

Argument 1 (simplified): Say W is a counterfactual world where Gore is president, and where Fred believes the president is smart. Strong two-dimensionalism predicts that (1) and (2) are equivalent. But (1) may be true while (2) is false. So strong two-dimensionalism is false.

(1) In W, Fred believes that the president is smart.
(2) In W, Fred believes that the actual president is smart.

Response 1 (on behalf of a strong two-dimensionalist): There are two natural readings of (2) in English: (i) a *de re* reading that requires that in W, Fred believes of Bush that he is smart, and (ii) a *de dicto* reading that in W, Fred believes that the president (in his world) is smart. Strong two-dimensionalism makes no prediction about the former and makes the right prediction about the latter. Any reading of (2) that requires a belief about the actual world is unnatural at best. If it is replied that we are dealing with a stipulative philosophical use of 'actual', the strong two-dimensionalist may stipulate the behavior in epistemic contexts: on the usage relevant for their claims, 'actual' picks out the world of the believer in combined modal/epistemic contexts, reflecting our intuitions in (ii).

Response 2 (on behalf of a weakened strong two-dimensionalist): There is already good reason for a strong two-dimensionalist to give special treatment to indexicals, to handle:

(3) Fred believes that I am Australian.

requiring that Fred believe of *me* that I am Australian, not that Fred believes the primary intension of "I am Australian". If the strong two-dimensionalist were to accept Soames' intuitions about belief ascriptions, they might simply hold that the same sort of special treatment be extended to 'actual'.

Response 3: (on behalf of me): My view of belief ascriptions ("The Components of Content", 1995/2002; see also WT5b*) holds roughly that:

> 'x believes that S' is true iff the subject has a belief with the [structured] secondary intension of S and with a contextually appropriate [structured] primary intension

This handles Soames' intuitions straightforwardly: 'The president is smart' and 'The actual president is smart' have different (structured) secondary intensions, so (1) and (2) are not necessarily equivalent.

Argument 2 (simplified): Say W is a world where Mars is the evening star, where John believes this, and where John has no beliefs about Venus. Then (4) is false and (5) is true. Strong two-dimensionalism predicts that (4) or (5) are equivalent, so it is false.

(4) In W, John believes that Hesperus is the evening star
(5) In W, John believes that the evening star is the evening star.

Response: This is a good reason to reject any view on which belief ascriptions simply ascribe a relation to a primary intension. It is no problem for my view above: 'Hesperus' and 'the evening star' have different secondary intensions and may also determine different constraints on appropriate primary intensions.

Argument 3 (simplified): (6) entails (7) but strong two-dimensionalism entails that (6) can be true while (7) is false. So strong two-dimensionalism is false.

(6) In W, Hesperus exists, and John believes that Hesperus is the evening star.
(7) In W, there is an x such that John believes that x is the evening star.

Response: ditto. Satisfying (1) requires a relation to a (structured) secondary intension, which handles the entailment. Note that the two-dimensionalist is not committed to any particular semantics of de re belief attributions such as (2). But I think something like the following is plausible (see "The Components of Content"):

> 'x believes of y that it is F' is true iff the subject has a belief with the (structured) secondary intension of 'y is F' and with an acquaintance-appropriate primary intension.

Soames' Argument Against Weak Two-Dimensionalism:

Argument 1 (against equivalence of names with actualized descriptions): If 'Kripke' is equivalent to 'the actual SK', then (8) and (9) should be equivalent, but they're not, as (9) requires a belief about the actual world and (8) does not.

(8) In W, Fred believes that Kripke is smart
(9) In W, Fred believes that the actual SK is smart.

Responses: (i) Much as with argument 1 against strong-two dimensionalism, the argument requires an unnatural reading of (9); (ii) If we accept that reading, a view in the vicinity of weak two-dimensionalism can accommodate Soames' intuitions by appealing to structured secondary intensions; (iii) The case poses no problem for my own view of belief ascriptions.

Argument 2 (against equivalence of names with 'dthat'-involving expressions**):** If 'Kripke' is equivalent to 'dthat(the SK)' and Kaplan is equivalent to 'dthat(the DK)', then weak 2D predicts that (10) is a priori, and that this entails that (11) is a priori.

` (10) dthat(the DK) ≠ dthat (the SK who isn't DK)
(11) Kaplan ≠ Kripke.

Response: if the view predicts that (11) is a priori, so much the worse for the view.

Argument 3:

> "Why couldn't there be a sentence S in any possible language which was such that (i) S consists of a proper name for an object o together with a predicate expressing a necessary property of o, (ii) the character S is a **constant** function from contexts to a necessary proposition p, yet (iii) justifiably accepting that character, and knowing it to express a truth, requires empirical evidence, and (iv) there is no other, apriori, route to p?"

Short answer: primary intensions (unlike character) are defined in epistemic terms, with a constitutive connection to the a priori. If S has a necessary primary intension then for all scenarios W, it is a priori that if W is actual, then S is the case. Further substantive claim (defended at length in e.g. "The Foundations of Two-Dimensional Semantics"): if S is not a priori, then there is some scenario W such that it is a priori that if W is actual, then S is not (determinately) the case. It follows that if S has a necessary primary intension then S is a priori.

Soames' Arguments Against Reference-Fixing Descriptions

I don't think that expression tokens have to be associated with reference-fixing descriptions. But I think they have to be associated with primary intensions, which are in certain respects akin to reference-fixing descriptions. Some aspects of Soames' discussion generalize to these.

Soames' Argument 1: 2-D confuses facts about what gives a term its meaning with facts about what's involved in understanding a term.

Response: The association of primary intensions with a term turns on the term's *epistemic role*, something tied deeply to understanding. The epistemic intuitions above reflect a key aspect of the speaker's understanding of terms such as 'water' and 'Godel'

Soames' Argument 2/3c: Judgments about what the term 'water' would refer to as it occurs in counterfactual scenarios don't yield firm conclusions about meaning, as 'water' may have a different meaning in the counterfactual situation. There are worlds where 'and' means disjunction, but this doesn't entail anything about the meaning of 'and' in our world.

Response: These points are correct, but they miss the important difference between the epistemic and contextual understandings of two-dimensionalism. On the epistemic understanding, claims about the primary intension of 'water' don't turn on metalinguistic counterfactuals about what 'water' would have referred to as it occurred in different situations. Rather, they turn on first-order judgments such as: if the XYZ-scenario is actual, then water is (actually) XYZ. There's nothing metalinguistic about these judgments. To see this, note that a *belief* that the XYZ-scenario is actual and a *belief* that water is XYZ stand in the relevant inferential relation to each other (if one accepts one, one should accept the other). There's no need to *mention* the term 'water' to generate this pattern of judgments.

Soames' Argument 3a: The twins Castor and Pollux satisfy the same descriptions, but Castor's term 'I' refers to Castor and not to Pollux.

Response: No problem for primary intensions over centered worlds.

Soames' Argument 3b: Kripke's arguments don't presuppose that speakers can determine reference of a name with respect to all scenarios, just with respect to some scenarios.

Response: Kripke's arguments don't *presuppose* that names have (full, determinate) primary intensions, but they clearly can't *refute* the claim that names have such intensions.

Kripke-style epistemic arguments argue against the equivalence of names (e.g. 'Godel') with descriptions (e.g. 'the prover of incompleteness') proceed by exhibiting epistemic possibilities V with respect to which N and D come apart: if we discover that V is actual, we'll deny that Godel is the prover. The two-dimensionalist handles this by stipulating (more or less definitionally) that primary intensions follow our rational judgments here: with respect to V, the primary intension of 'Godel' picks out the stealer, not the prover. It is in this sense that Kripke-style epistemic arguments (based on our judgments about cases) cannot possible refute the claim that names are associated with primary intension: the two-dimensionalist simply can appeal to an intension that mirrors the Kripke-style judgments about cases.

Soames' Argument 4: Metalinguistic descriptions don't always work and aren't part of meaning.

Response: Cases where they don't work are handled the same way as Kripke cases: the primary intension goes with our judgment about the case, not with the description. The issue about whether epistemically-defined primary intensions are part of "meaning" is a partly terminological one that depends on what one requires of "meaning".

Summary

Soames' arguments have no force against core two-dimensionalism. Soames' central arguments are not directed at core two-dimensionalism at all, but against certain implausible ancillary views (e.g., about belief ascriptions). All of these arguments are handled straightforwardly by, e.g., the form of two-dimensionalism that I have defended in the literature

Concluding Conciliatory Comments

One upshot of the above diagnosis (the view Soames attacks is not the view I defend) is that the fundamental locus of disagreement, if any, is not entirely clear.

Soames' remarks about epistemic possibility at the end have a strikingly two-dimensional flavor. We have a set of epistemically possible world-states (scenarios?) and a set of metaphysically possible world-states relative to each of these. I have defended a similar picture (in e.g. "The Foundations of Two-Dimensional Semantics" and "The Nature of Epistemic Space"). Presumably one can even associate expression tokens with a "primary intension" (the set of epistemically possible world-states that verify it), such that when a sentence is a priori it has a necessary primary intension.

Residual (potential) disagreements: (i) Soames thinks that 'Hesperus is Phosphorus' is a priori, so there is no epistemically possible world-state that falsifies it. (ii) Soames embraces a correspondingly counterintuitive view of belief ascriptions. But we can work on these things.