Midwest Studies in Philosophy
Volume 45 (2021), pp. 541–560
DOI: 10.5840/msp2021111623

Metasemantic Relationism

SETH YALCIN
University of California, Berkeley

Abstract: Consider a language incorporating a mirror-image form of assertion, where the norm is to express what you take to be false rather than what you take to be true. Why aren't ordinary languages like that? Why do we generally assert what we take to be true rather than what we take to be false? If Lewis (1975) and Massey (1978) are right, there is a sense in which the question is based on a mistake, and in which English (etc.) could be described either way. I explore that idea, which centers on the role of duality in language. One of the main questions in the air is whether the symmetry of duality can be used as a guide to 'real structure' in semantics and pragmatics. I try to think through it with an analogy to relationism about space.

Key words: assertion, truth, meaning, semantics, pragmatics, relationism, holism, dynamic semantics, attitude verbs

A: There's something about assertion I'm stuck on. Why do we generally say what we take to be true?

B: You mean, why not lie? You're stuck on this?

A: Sorry, let me set up the question. Here is maybe the most basic thing about assertion: when you assert p in conversation, you're trying to get p added to the common ground—to the stock of things taken to be true by the parties to the conversation. Whatever else comes along with the practice of assertion, that much seems hard to argue with.

B: Sure. Stalnaker 1978, etc.

A: Yet there's another way things could have been set up—another conceivable convention we could have had for our most basic way of transmitting information linguistically. Instead of expressing the propositions we want to take on board in the conversation, we could have had a practice whereby we express the propositions we mean to rule out, to set aside. As it stands, we normally assert what we believe, or what we know, or what we take to be true. But, imaginably, we could have had a sort of mirror-image practice, one whereby the things we say are understood to be the things we doubt, or the

things we take ourselves to know aren't so, or the things we take to be false. That too would work for getting information across. And it seems like it'd work exactly as well.

But looking around at natural languages, one doesn't seem to find this weird alternative—*subtractive* assertion, let's call it. Why not? Why is it that we express the propositions that we want to for granted in conversation, rather than ones that we want to set aside?

B: OK, I think I'm getting it. So the idea is that speakers of your subtractive languages are not pathological liars, though they always say what they take to be false. (Maybe we need air-quotes around "say" there.) They are just employing a different sort of convention for using the meaning of a sentence to convey information.

A: Right.

B: I guess we could put the question using Stalnaker's framework. Suppose the common ground of a conversation fixes a *context set*, the set of worlds compatible with the propositions that are common ground. Think of the assertion of a proposition as having a characteristic update effect on the context set. There is some *update function* for assertion, call it $\cdot[\cdot]$, that maps the proposition asserted and an input context set to an output context set. Now we can ask: what is the update function for assertion? Stalnaker's specific modeling proposal is that assertoric update has an *intersective* character: for any context set c and proposition p,

Intersective update. $c[p] = c \cap p$

The simple idea is that when you assert something, and what you asserted becomes accepted and taken for granted in the conversation, what happens is that possibilities incompatible with the proposition asserted are eliminated from the context set.

Now given that sort of background, you're observing that we could have had this rule:

Subtractive update. $c[p] = c - p$

And you're asking: "Why do we have intersective update rather than subtractive update?" Am I still getting it?

A: I don't know that the problem depends on Stalnaker's particular model, but anyway, yes, that is a way we could sharpen things a bit. I mean, we could follow Stalnaker in thinking of information as the sort of thing that eliminates possibilities; in theorizing about the common ground via context sets; and in taking for granted the idea that assertion involves expressing a proposition. We can even take on board the idea that the proposition expressed is, at least in normal cases of assertion, used to add information—

eliminate possibilities—from the context set. But none of this forces intersective update rather than subtractive update on us. So I'm left wondering why subtractive update isn't a thing.

B: Right. Because sarcasm is definitely not a thing.

A: OK—sarcasm is a thing. But it's the exception that proves the rule, right? Part of the fun of sarcasm is exactly the way it violates expectations, flouting the ordinary pattern of assertoric update. In that way, it only highlights the fact that the ordinary pattern *isn't* subtractive, and is instead closer to the sort of update Stalnaker identified.

B: Agreed. I think the deeper answer to your question is that the choice between describing assertoric update in English (or whatever) as intersective or as subtractive is arbitrary.

A: What? Surely if I say 'Caesar was murdered' and you take me to be proposing to eliminate worlds where Caesar was murdered from our context set, you've just made some kind of blunder understanding how assertion works in ordinary English.

B: Well, that's right, but that's not quite what I had in mind. Rather, my idea is that the semantics-pragmatics of English, or whatever natural language you want, is invariant under a certain kind of symmetry transformation. Consider a language I'll call *anti-English*. Semantically, anti-English is just English with the truth-conditions reversed.[1] That's to say, where an English sentence is assigned the truth condition (set of worlds) p, the corresponding anti-English interpretation assigns that sentence the complement truth-conditions p'. Pragmatically, anti-English employs subtractive assertion rather than the intersective kind. Now my claim is that these are really just notational variants. The distinction between English and anti-English is a distinction without a difference.

A: Whoa. Let me process this. I agree that it would at least be epistemically hard to tell English and anti-English apart. In anti-English, 'Caesar was murdered' expresses the proposition that Caesar wasn't murdered, but given this proposition is *subtractively* asserted, it has the same dynamic effect on the context set as the sentence does in ordinary English. That does make it look like the two languages amount to the same thing. But are you sure that the truth-condition of a whole sentence in anti-English can be understood

1. For the basic idea, see Geach's notion of Unglish in Geach 1972 and 1982, Lewis's notion of an anti-L in Lewis 1975, and Massey's notion of a dualistic manual in Massey 1978. (Geach credits Wittgenstein [1922, 4.062–.0621], for some reason.) These three discussions appear to be independent. Lewis's discussion is brief in comparison to those of Geach and Massey; and unlike those authors, Lewis does not discuss the question of compositionality of anti-English. Massey uses the possibility of dual interpretation to frame an argument for the indeterminacy of translation. See also Casati and Varzi 2000.

as determined by the meanings of its parts and their syntactic configuration? That is, is it so clear that anti-English can be compositional, in the way a natural language needs to be? If it's not compositional, then that'd seem reason enough to reject it as a possible interpretation for a natural language.

B: Why think there's any problem about compositionality?

A: Well, let's think through a case. Suppose we have two English sentences, ϕ and ψ, and we have their conjunction, $\phi \wedge \psi$. Anti-English is supposed to give every English sentence *s* the same truth-condition that English gives to $\neg s$, so the translation scheme for these three sentences should be:

Anti-English	English translation
ϕ	$\neg\phi$
ψ	$\neg\psi$
$\phi \wedge \psi$	$\neg(\phi \wedge \psi)$

B: Right.

A: But look, if anti-English is also compositional, we ought to translate any ϕ as $\neg\phi$, not only when it is "alone," i.e., unembedded, but also when it is in embedded contexts, such as when it performs as a conjunct. But that predicts the translation scheme will shake out as:

Anti-English	English translation
ϕ	$\neg\phi$
ψ	$\neg\psi$
$\phi \wedge \psi$	$\neg\phi \wedge \neg\psi$

But then, the conjunction $\phi \wedge \psi$ does *not* get mapped to $\neg(\phi \wedge \psi)$ after all. So the requirement that anti-English be compositional seems to conflict with the goal of assigning every sentence to its negation.

B: Hang on—what are you assuming the anti-English translation of 'and' is?

A: Well, I guess I was just assuming it means the same thing that it does in English. But I see from your raised eyebrow that this is a nontrivial assumption. Maybe we should assume instead that 'and' in anti-English means the same as 'It is not the case that both . . . and . . .'? But that wouldn't work either:

Anti-English	English translation
ϕ	$\neg\phi$
ψ	$\neg\psi$
$\phi \wedge \psi$	$\neg(\neg\phi \wedge \neg\psi)$

Plainly $\neg(\neg\phi\wedge\neg\psi)$ is not logically equivalent to $\neg(\phi\wedge\psi)$, but the latter is what we want if anti-English is to map every English sentence to its English falsity conditions.

B: The solution is to say that in anti-English, 'and' means what 'or' does in English (Massey 1978, Geach 1982). So the correct translation scheme is:

Anti-English	English translation
ϕ	$\neg\phi$
ψ	$\neg\psi$
$\phi\wedge\psi$	$\neg\phi\vee\neg\psi$

Thus $\phi\wedge\psi$ gets mapped to $\neg\phi\vee\neg\psi$, which is logically equivalent to $\neg(\phi\wedge\psi)$—and that's just what we wanted. So compositionality is preserved.

A: Whoa again. OK, I think I'm catching on. Clearly it's not just 'and' and 'or' that have to get flipped around when you move from English to anti-English. To imagine anti-English, we have to imagine flipping the intensions of all predicates to their anti-intensions, swapping the tables for 'and' and 'or,' swapping the semantics of the possibility and necessity modals, the semantics of the existential and universal quantifiers . . . Am I following?

B: You've got it. You have to perform a sort of symmetry transformation for the whole language, inter alia flipping all the dual operators. At the end of the day, anti-English will just as compositional as English, but delivering the reverse truth-conditions. Now pair that sort of a semantics with a subtractive assertion rule in your pragmatics. Then what I'm saying is, really what you'd get from that combination just is English, albeit under a nonstandard description.

A: I'm impressed that this looks like it might actually work. I'm not entirely convinced yet that every English expression admits of a compositionality-preserving anti-English translation—we've hardly proven that—but I won't press the case for now (though I want to circle back to this). It seems like there are some other more basic questions to ask. Don't these languages differ rather obviously at a logical level? Like: 'Caesar was murdered' entails 'Someone was murdered.' That entailment holds in English, but fails in anti-English.

B: That's true in a certain sense. But the thing to notice is that anti-English speakers will be happy to (subtractively) assert the sentence:

'Caesar was murdered' entails 'Someone was murdered.'

That's because in anti-English, 'entails' means, roughly, *doesn't preserve falsity*. And I guess more importantly, in the metalanguage for anti-English, entailment would be defined as that sort of notion.

A: I'm getting a brain cramp trying to keep track of the layers of negation involved in the translation scheme from English to anti-English, so I'll just take your word for that. But I guess I'm catching on enough to anticipate what you were going to say to my other objection. I was going to object: 'Caesar was murdered' is true, but what it says is false in anti-English, so anti-English couldn't be English. But I think you'll say that anti-English flips the meaning of 'true' and 'false.' So if you say:

'Caesar was murdered' is true.

in anti-English, it's semantically equivalent to

'Caesar was murdered' is false.

in English. And then if you subtractively assert that anti-English content, it has the same update effect as intersectively asserting the ordinary English content.

B: Yes. Like other predicates, the extension of 'is true' in anti-English corresponds to its English anti-extension.

A: Alright, let me try to sum up what you're suggesting. In semantics-pragmatics, we have an array of dualities—between truth and falsity, extension/intension and anti-extension/anti-intension, $\wedge$ and $\vee$, and $\Diamond$, $\exists$, and $\forall$, intersective and subtractive assertion, etc.—that exhibit a certain sort of interdependence. Is the extension of 'grass,' say, the grass or the non-grass? If we hold fixed our usual way of thinking about assertion, and hold fixed the usual way of stating the semantics of other expressions, we'll want to say that the extension of 'grass' is the grass, at least if the resulting theory is supposed to get English right. But there's *another*, "dual" way to get English right, which has it that the extension of 'grass' the non-grass—this is the anti-English way of giving semantic values, defining consequence, and defining assertion. And your view is that the usual theory and its anti-English counterpart are not just epistemically indiscernible by us—rather, they are not really different theories at all.

B: Just right. English and anti-English seem opposed if we narrow our attention to isolated corners of the language, but if we step back and see them each as the full package deals they are, we see the same deal, just described two ways. The point of bringing up anti-English is to help us to avoid wrapping into our characterization of English artifactual features of our model of it. Maybe the clearest way to put the lesson is like this: take our traditional way of stating the semantics and pragmatics of English, and pair it with the anti-English style of description. The real content of our theory of English is given by this pair, by what's common between these.

A: You seem very calm about this. But I am a little destabilized at the thought that it's not determinate whether 'grass' means *grass* or *non-grass*, or whether 'and' means *and* or *or*. You seem to have described maybe "the most dramatic example yet of indeterminacy of translation" (McGee 2004, 289).[2] But I thought philosophers of language agreed to stop talking about the indeterminacy of translation somewhere in the mid-nineties. Now I'm worried I shouldn't have taken all those classes in natural language semantics.

B: Well, I wasn't planning to take this in a Quinean direction, towards a sort of meaning skepticism. The way I think of it, all we really have here is a special case of the familiar point that precise models of things pretty much always involve artifactual elements, elements that don't correspond to real structure in the thing being modeled.

A: I think I dimly see what you're saying. Maybe it would be good to try to think up an analogy.

B: How about space? Think of our geometric representations of space, and consider the question of which parts of the representation reflect what's real about space itself, independent of the representation. Suppose we posit a coordinate structure to represent and reason about spatial reality. Then we could pose the old question of relationism versus absolutism about space. We could ask: are there real spatial locations, corresponding to the coordinates in our geometric representation? Or are those just handy tools for representing spatial relations between things, the latter relations being all there really is spatially speaking? The relationist says the spatial relations are all there is to space, whereas the absolutist goes in for more in reality—usually, an ontology of spatial points or regions. And normally the absolutist will offer to *explain* spatial relations between objects in space in terms of entities from their spatial ontology. She might say: a pair of shoes are two feet apart, for example, *because* they are each at certain real locations in space, and the spatial locations they respectively occupy are in turn two feet apart.

A: Whereas for the relationist, the explanation stops earlier.

B: Right. Another way to see the difference between the views is to observe that the absolutist recognizes more possibilities than the relationist. Is there another possible world where the spatial distance relations between material things are exactly as they actually are, but everything is uniformly moved in space, say, two feet that way? The absolutist can make sense of that, but the relationist can't. The relationist sees at most a notational difference.

2. This is McGee responding to Massey. See also Quine 1992, 51, where Quine characterizes this sort of example (Massey's version) as one of the two best developments of his indeterminacy of translation thesis.

The idea of a real distinction here is a "chimerical supposition," an "impossible fiction" (Leibniz 1716, 23), etc.

A: OK, so what's the analogy? I'm guessing your view of meaning is sort of like the relationist's view of space?

B: Right. The relationist about space also believes in a kind of indeterminacy of translation—"translation" in the spatial-geometric sense. Once you've described a full arrangement of things in space, fixing the whole pattern of distance relations between everything, it *seems* like you can talk about translating everything uniformly in the space in arbitrary distances and directions. But the relationist has it that these different translations reflect equally good descriptions of the same state of affairs—nothing in reality settles one as better. You can of course make sense of translating particular objects in various ways while holding fixed the position of other things—that'd change the spatial relations between things, and so make for a new possibility. But once you start talking about translating *everything at once uniformly*, you're talking distinctions without a difference. Likewise with rotation, by the way. To get at what's spatially real, you have to look at what's invariant under these sorts of global transformation of the geometric representation. In the same way, what we've called English and anti-English are equally good descriptions of the same language. The real content of our semantic-pragmatic theory of the language is best understood as given by what's invariant between those two representations.

A: That was an ambitious pun on "translation."

B: But I do think the analogy works pretty well.

A: It does help me see the view. Actually, I wonder if an even better analogy to your relationist view about semantics-pragmatics would be relationism, not about spatial location or degree of rotation, but about left-right orientation. Say the *orientational relationist* holds that there are no facts about the orientation of a thing that go beyond the pattern of distance relations between it, its parts, and other things. On this view, left hands are (approximately) congruent with other left hands and incongruent with right hands, but there is no further spatial-geometric property of *being left-hand-shaped* per se that separates the left hands from the right hands. To settle which hands are (in)congruent with which others is to settle everything that there is to settle about the orientation of a hand. So a mirror-reflection of this universe is just the universe redescribed, not a genuinely different way things could have been. This sort of relationism goes in for a sort of binary indeterminacy—there are just these two equally good ways of describing the orientations of things. That's rather like the duality you're arguing for in language.

B: Yes. I love it. Not that the two are related, but I'd be happy to sign up for both sorts of relationism.

A: Great. Because these two views seem to me hard to believe in roughly the same way.

B: I'm listening.

A: I find myself attracted to the following refutation of orientational relationism. (*Holds up hands.*) Here is a left hand. And here is a right hand.

B: LOL. You're offering a Moorean proof of the reality of leftness per se? (Or was that the Jedi mind-trick hand gesture?)

A: Take a good strong look at Lefty over here. Can't you just *see* that it's a left hand?

B: Well, sure, I guess.

A: Does it look like you need to examine anything *extrinsic* to the hand to detect its leftness?

B: I concede that it's tempting to say that the leftness of the hand is just part of the shape of the hand. And I grant that no particular extrinsic relation between the hand and something else jumps out as obviously relevant, perceptually speaking.

A: What else you need? The idea that things have orientations, and yet are *not oriented any particular way*, is just about the most disorienting thought there is. I have trouble holding that idea in my head while I look at my hands.

B: You're seeing your hands clearly enough, but I'm not sure you're seeing the relationist's thesis. Maybe you would also say it *looks* as if the sun goes round the earth? But then, this *just is* what it looks like for the earth to rotate. If your experience seems incompatible with relationism, we have to ask: how supposedly would things look different, were orientational relationism true? And I think it's harder than you suppose to show that there is a tension between the appearances and this view of spatial reality.[3]

A: I'm not sure *I'm* the one carrying the burden of needing to square appearance with reality—

B: Moreover, saying things "aren't oriented any particular way" according to the relationist is a bit misleading. I mean, your two hands are still incongruent with each other on this view—which is to say, not oriented the same way. There are facts of sameness of orientation, and of difference.[4] It's just that the (in)congruence facts are where the orientational facts stop. You could still as it were divide all the hands in the world into two piles sorted by

3. For further discussion of the experience of orientation, see Lee 2006 and Baker 2012.

4. Though in a nonorientable space (like a Möbius strip) such facts can only be characterized relative to locally orientable regions. I assume an orientable space for simplicity in the discussion, but nothing important hangs on it.

congruence, and then give the piles names—there's the 'left' pile, over there is the 'right' pile.

A: OK, that's fair enough. It's just so tempting to me to want to *explain* the incongruence of my hands in terms of further facts about the handedness of each hand.

B: I had that temptation once. But it receded once I realized that the standard definition of congruence does not require the idea that hands (or other chiral things) have intrinsic orientations.

A: Anyway, this all brings me back to meaning. It seems to me that given the sort of indeterminacy about meaning you described—or maybe we should say, your *relationism* about meaning—sentences have truth values in something like the way hands have orientations, on the relational view of orientation. And relatedly, a certain kind of explanation we might have thought we had for a sentence's having the truth value it has isn't actually available.

B: That all sounds vaguely agreeable to me. Keep talking?

A: Well, take any declarative sentence you please that seems obviously to have a truth value—'Grass is green,' say. Is it true or false?

B: Well, true, obviously.

A: Yeah. But of course, in the context of anti-English—from that metasemantic way of describing things—it's false. And your view just is that English and anti-English are just redescriptions of the same language.

B: True.

A: Before you got us all twisted with your indeterminacy argument, I would have been perfectly happy saying that the truth of 'Grass is green' owes to (i) the sentence's representing the world as being a certain way, and (ii) the world's actually being that way. That is, the sentence has truth-conditions, and the conditions in question obtain. (By the way, I don't have in mind a merely disquotational understanding of truth-conditions here—I'm thinking of truth-conditions in the more robust correspondence sense.[5]) Of course, I would have allowed that our *epistemic* route to knowledge of this fact about the sentence might have a relatively holistic character, coming by way of confirming a broader semantic-pragmatic theory for the language at large, etc. Still, I would have been inclined to say that metaphysically speaking, the truth of the sentence is at most tied up with *that* sentence and its meaningful parts—that the truth-conditions of the sentence don't metaphysically depend on pieces of the language outside that sentence. But if your indeterminacy argument is correct, then it doesn't seem that I can still have this picture.

B: Why not?

5. I discuss this distinction further in Yalcin 2018.

A: Well, if it's right that the English/anti-English distinction is one without a difference, then although 'Grass is green' is still "tied up with" with a certain condition of the world, a certain division in logical space, it is not determinate whether it represents that condition as satisfied or as not satisfied! I would like to have said: *here's the meaning of the sentence*, and presented you with a truth-condition dressed in the usual model-theoretic clothing. But I'd feel deflated doing that now, knowing that with suitable adjustments I could have served you the *opposite* truth-condition and had thereby provided a theory *no less correct*.

B: Now I think I see what you're (I'm?) getting at. Yes, the truth value of a sentence is sort of like the handedness of a hand, if you take a relational view of both domains. Hands can be congruent or incongruent with each other. That's a real thing about hands. But if a pair of hands are (say) incongruent, that's not something further metaphysically grounded in one of the hand's having a robust spatial property of being left (right) per se. There are no such orientational properties. The congruence facts are all there is, orientationally speaking. I guess in the same spirit, I have to say: sentences have different truth values or the same truth value. That's a real thing about sentences. But those differences aren't metaphysically grounded in particular truth values had by the sentences. Truth is not a "robust property" of sentences, any more than leftness is a robust property of hands.[6]

A: It's one thing to say leftness per se isn't a thing. But truth? Truth is basically my favorite thing.

B: Well, to be clear, I'm not jettisoning *truths*, any more than my left hand! And it's not like either of us really wants truth per se as a thing in our ontology. I am saying that if we think of truth as a property of sentences and propositions, there's a sense in which it's not a robust property. But maybe my view sounds farther out than it really is. Let's keep in mind that there's no empirically detectable difference between English and anti-English. Further, on both theories, the informational upshot of any ϕ you please is exactly the same—to use Stalnaker's framework, the ultimate dynamic change to the context set is the same. People still count as transferring information and all that. How far out could I really be?

6. Compare: "If someone asks, 'But what kind of entities are these truth-values supposed to be?,' we may reply that there is no more difficulty in seeing what the truth-value of a sentence may be than there is in seeing what the direction of a line may be; we have been told when two sentences have the same truth-value—when they are materially equivalent—just as we know when two lines have the same direction—when they are parallel" (Dummett 1959, 141).

Of course, once all the sentences are grouped into equivalence classes by the sameness-of-truth-value relation, A will complain that there is an important further question we could raise: which is the class of true sentences? And B will say that there is something factually defective about that question.

A: Let's keep poking at this view. One thing I wonder is whether your view counts as a *holistic* view of meaning. And related, I (still) wonder whether it is compatible with thinking compositionality is an important, robust constraint on meaning. I am used to thinking of the truth-condition of a sentence as determined as a function of the meanings of its parts. But on your view, it turns out that the particular truth-condition assigned to a sentence by our theory is in a certain respect an artifact of the theory; and it seems the ultimate import of that assignment turns out to hinge on things remote from that sentence and its parts, such as how assertion updates the context. This gives me holism vibes. Maybe you'd even say "To understand a sentence is to understand a language" (Wittgenstein 1953, §199), or "Only in the context of the language does a sentence (and therefore a word) have meaning" (Davidson 1967, 22).

B: Let's not go crazy here. We should make some distinctions. First, as you noted earlier, an *epistemological* holism about how a semantic-pragmatic theory gets confirmed seems not only fine but hard to argue with. Surely the predictive import of any given assignment of truth-conditions to sentences *always* depended on other elements of the theory, like how assertion works, or how consistency is defined. So let's agree that that's right, and that it has nothing especially to do with the point raised by anti-English.

Another kind of holism in the air here is one about the meanings of theoretical terms. It's a familiar idea that theoretical terms often pick out functional roles that may call on each other in a complicated web, so that it might be plausible to say, at least in some cases, that they gain their sense holistically, as a network, in the context of a particular theory. Maybe such a view is right for the technical jargon of formal semantics and pragmatics. Whether that's right or not, this too would seem like a point about theories in general.

So I think you are asking, not about these holisms, but something about the metaphysics of the semantic properties that semantic-pragmatic theory traffics in. Something like: are these best conceived as holistically-had properties, such that the meaning of a whole sentence might not necessarily be fixed just by the meanings of its parts and their organization?

A: Right.

B: To that I'd say that compositionality isn't lost. If we think of the substantive content of our semantic-pragmatic theory as revealed by what's common between the English and anti-English styles of description, then since you must have a compositional theory either way you slice it, that's reassurance that compositionality hasn't gone anywhere. If there's a certain lack of fixity or determinacy to the truth-condition of a sentence on this picture, that's a feature inherited from its parts.

A: OK—though I'm not sure I still have a grip on what's composing with what anymore. Anyway, let me ask a different sort of question. What about the response that a theory stated in the anti-English style, while possibly descriptively adequate, couldn't possibly be *foundationally* adequate, because it couldn't cohere with the best account of the intentionality of language—the best story about in virtue of what expressions have the meanings they have?

B: Not sure I'm following you.

A: Well, take for instance the Lewisian idea that meaning is fixed by use plus eligibility—specifically, his idea that a correct interpretation must assign relatively natural properties and relations to predicates to the extent possible (Lewis 1983 building on Merrill 1980). That's after all his way out of the more familiar Putnamian indeterminacy arguments. Couldn't it work here too? You've been arguing that it's in a certain sense indeterminate whether 'green' means *green* or *non-green*. But surely the former is a more natural property than the latter, and hence (if Lewis is right) more eligible to be the meaning of 'green.'

B: I'm surprised you think so. I would have thought that the natural properties are 'closed under negation' in the relevant way. That'd work better with the typical "joints in nature" metaphor: what's natural (or not) is the *distinction* the property (relation) carves out, not one particular side of the joint.

A: But surely properties like being metallic or being a kumquat are going to figure more prominently in illuminating metaphysical explanations, lawlike descriptions, etc., than properties like nonmetallicness or nonkumquathood. As for Lewis, he seemed okay with thinking that the extensions of the perfectly natural properties were united by a shared *having* of something, not by a shared lacking of something.

B: Well, suppose for the sake of argument we take your view of naturalness, so that that the naturalness of the property of being *F* isn't necessarily identical to the naturalness of the property of being a non-*F*—that such pairs of properties can be asymmetrical in respect of their naturalness. Could that asymmetry break the tie between English and anti-English in the favor of English, by way of the reference-fixing role of naturalness? But I don't see why we wouldn't just say that in the context of a theory stated in the anti-English way, natural properties exert their magnetism on the anti-intensions of predicates. That'd give us an equal and opposite Lewisian metasemantics, so that a theory stated in the anti-English way remains on par with the more usual statement.

A: Hmm. Can you make the same kind of move in response to other theories of intentionality? Another familiar idea is that the properties our predicates answer to is, at least sometimes, a matter of our causal or infor-

mation-theoretic connections to the world around us. 'Water' means water rather than XYZ partly because, whether or not we could tell the difference, our actual interactions have been with water. *Water*—not non-water! It's hard to see how you could have a Kripke-like causal chain theory that links our water talk back to non-water.

B: Yes, that is very hard. But again, a predicate meaning always has those two complementary parts—intension and anti-intension. If you theorize in the anti-English way and you hope to equip your semantics with a causal or information-theoretic story about the intentionality of language, your causal or information-theoretic connections will tie the relevant properties of your environment to the relevant anti-intensions of your predicates. That results again in a theory at parity with a theory stated in the more usual way.

A: That does make it feel like the anti-English theory is just sort of moving labels around—which, I suppose, is not far from what you're getting at when you say that the difference between it and the regular theory is a distinction without a difference.

Well, still looking for ways to break the apparent tie between English and anti-English, let me circle back to the apparently empirical question of whether every English expression admits of a compositionality-preserving anti-English translation. I just thought of another case. What about attitude verbs? Take a sentence like:

1. Gerald believes that it's raining. (Bp)

Under what condition is this true in in anti-English? Well, obviously what we want is:

2. It is not the case that Gerald believes it's raining. ($\neg Bp$)

Moreover, we want that result compositionally. How do we get it? I think there's a problem here. If the intension of 'believes' in anti-English is identical to its English anti-intension, then 'x believes that p' is true in anti-English just in case x isn't belief-related to p. Further, 'It's raining' is true in anti-English just in case it's not raining. But then it seems that the truth-condition of (1) shakes out to be

3. It is not the case that Gerald believes it's not raining. ($\neg B\neg p$)

—which is not (2), the thing we wanted.[7]

B: I think once again that the problem is that you have the wrong translation of the relevant anti-English word. Like modal operators, attitude verbs

7. My discussion here is following Geach (1982), who raised this kind of problem for what he calls the 'Frege-Dummett' view of attitudes. Geach's preferred view—"a theory along the lines of Prior's"—is the sort of view B is about to describe.

have their duals. Just as we translate the □'s of anti-English as the ◇'s of English, we translate the belief operator *B* of anti-English as equivalent to ¬*B*¬ in English. If we do that, what we get is:

4. It is not the case that Gerald believes it is not the case that it's not raining. (¬*B*¬¬*p*)

I take it ¬*B*¬¬*p* simplifies to ¬*Bp*, and we have the result we wanted: the truth-condition of (1) is indeed stated by (2).

A: Intriguing. That does looks like it saves compositional translation. But I notice that in reasoning this way, you seem required to deny that attitude verbs classify with ordinary predicates, since their mode of translation into anti-English is obviously different. In this way, you seem to be using the assumption that there must be a compositional translation manual into anti-English as a guide to thinking about the structure of English.[8] But what justifies this?

B: That's a good question. I concede that this is an unusual approach to defend a view about which expressions in English are or aren't semantically alike. I guess I was thinking that if we can preserve the symmetry transformation of English into anti-English, then other things being equal we'll end up trafficking in less structure—ignoring as we can structure that's not invariant across the translations—and less structure is better because simplicity is better.

A: It is interesting to consider what one has to say to preserve duality in the face of various examples. I wonder too about things like:

5. A man walked in. He sat down.

If that's in anti-English, I imagine you'd want the English translation of this discourse to come, not from negating each sentence individually, which sounds like nonsense:

6. Nobody walked in. ??He didn't sit down.

Rather, it should be tantamount to:

7. Nobody walked in and sat down.

—at least on one available reading of (5). But that would seem to suggest that discourse-level duality isn't always grounded in discrete sentential dualities. It seems to mean that the compositional semantics of anti-English ought to output something tantamount to (7) for the truth-conditions of the discourse (5). How could it do that? After all, (5) is two sentences, whereas (7) is just

8. Geach (1982) thinks a Priorian analysis of attitudes is "manifestly better" than one that doesn't preserve duality, but he doesn't expand.

one; and in (7) the quantificational subject is plausibly binding variables in each conjunct, making it hard to see how the conjuncts could ever be broken into two separate sentences. So how could (5) ever mean what (7) does? Is this some kind of abstract argument for the sort of dynamic semantics that would allow binding relations to cross sentential boundaries (Heim 1982, Groenendijk and Stokhof 1991, etc.)?

B (*mystified*): Hmm. I guess? I wasn't expecting this view to make such a substantive prediction, but perhaps that's right.

A: You know, all this makes me wonder if there's a lazier way to get to anti-English, one that doesn't require messing with the ordinary semantics of English at all. We've been assuming that the output of the semantics is a proposition, and that proposition is the input to the conversational update rule. But that's to implicitly assume a (simple but substantive) bridge principle connecting semantic value to assertoric content.[9] And one thing we haven't talked about is the possibility of messing with this bridge principle. For instance, we could consider leaving the semantics alone, but postulating bridge principle that has the effect of inverting the usual truth-condition. That would give us the same propositional inputs to the update rule that anti-English delivers, but using the ordinary English semantics.[10]

B: Yes, that's a good point. It seems very obvious that if you pair the ordinary English semantics with this flipped bridge principle and with subtractive update, you get just what you would get with an ordinary bridge principle and intersective update. So that's maybe a simpler way of seeing that there's a certain artificiality to the question of whether a language uses intersective versus subtractive update. But I think the case of anti-English raises deeper questions, because of the way it ropes in meaning and truth: Anti-English superficially seems to have a very different semantics, so if it's true that parity with English is really restored by equipping it with a dual form of conversational update, we have to confront the idea that this superficially very different language is only notationally different.

A: I think I need a few days to mull over your variety of relationism. Half the time I think it's a shocking view, and half of the time it seems sort of trivial—though examples like the last few make it appear pretty substantive in empirical import. By the way, do you have a name for this kind of relationism? Maybe *semantic* relationism?

9. On the distinction between semantic value and assertoric content, see Lewis 1980, Ninan 2010, Rabern 2012, and Yalcin 2014. For more on bridge principles, see Yalcin 2021.

10. Thanks to Brian Rabern for prompting me to consider this case.

B: Well, that label is already trademarked (Fine 2007). Also, the view I've been describing is not *just* a view about *semantic* features—remember, it crucially it involves the update rule for assertion.

A: Is there any relation between this view and Fine's?

B: There are certainly some overlapping themes. Fine too analogizes his view to relationism about space, and he doesn't want the particular semantic values assigned to expressions to be the ground of (all) their semantic features. Further, he has a theory of variable binding which allows for binding across the sentential boundary—something which, if your last point is right, my sort of relationism perhaps recommends, too. However, Fine's stress isn't on duality. Subtractive assertion isn't on his radar. He is mainly out to develop a rival to more typical sorts of semantic theory, for addressing Frege's puzzle among others. By contrast, my points here have been more meta—they've been about how the notion of a dual language might used to separate out the 'surplus structure' in our theory, even if that theory is of a very conventional sort.

A: Maybe *metasemantic* relationism then?

B: I guess? Kind of wordy.

A: Well, we need to wrap up here. Unfortunately though, I don't think we quite answered the question I posed at the beginning. I asked: why isn't subtractive update a thing? Your move was to try to dissolve the question, using the idea of duality to reveal at least one way that the question isn't substantive.

B: Yeah. That's in a way the biggest motivation for the relationist perspective: it makes certain questions ill-posed, and so leaves us with less explaining to do.

A: Sure. But I still find myself wanting something explained. Consider, not anti-English, but *subtractive* English, a language as much like English semantically as possible (theorizing about truth, intensions, consequence, etc., in the usual rather than anti-English way) but which involves subtractive assertion.

B: Oh, that'd definitely be very unlike speaking English! Since you're 'flipping' assertion but not the semantic values, the conversational update of sentences of subtractive English will be equivalent to the corresponding negated sentences in English.

A: Right. Now there's no contradiction in the idea of such a language. So I think we need to explain why we don't see it. Metasemantic relationism isn't irrelevant to this question, but I also don't see that it settles the question.

B: Obviously, it'd be very cumbersome to have an ordinary conversation in subtractive English. You'd probably end up constantly negating things. It's

hard to see how anaphora and presupposition would work. Not to mention rhetorical relations.

A: For sure. Superficially at least, it seems like subtractive English would be a comparatively inefficient or costly language to use. That feels like the obvious explanation for why we don't observe subtractive update in the wild (*modulo* of course your abstract relationist point that if any language has a compositional dual language, it can be described as subtractive). But we should separate two kinds of suboptimality. Subtractive update could be suboptimal *for humans and other creatures relevantly like us*, merely owing to contingent features our cognitive and/or social organization—maybe call that *empirically* suboptimal. Or it could be *ideally* suboptimal, meaning that there is something comparatively inferior (inefficient, costly) about this mode of update even when we restrict attention to ideal agents, agents who lack the peculiarities of human cognition and are free of limitations of reasoning or memory. Is subtractive English empirically suboptimal, ideally suboptimal, or both?

B: I'd like it to turn out to be ideally suboptimal, so that the assumption that we're rational is all that's really needed to explain why we don't observe languages like subtractive English. Hopefully I'll have an argument for that conclusion by the next time we talk.

References

Baker, D. 2012. "'The Experience of Left and Right' Meets the Physics of Left and Right." *Nous* 46(3): 483–498. https://doi.org/10.1111/j.1468-0068.2010.00814.x

Casati, R., and A. Varzi. 2000. "True and False: An Exchange." In *Circularity, Definition, and Truth*, ed. A. Gupta and A. Chapuis, 365–370. New Delhi: Indian Council of Philosophical Research.

Davidson, D. 1967. "Truth and Meaning." *Synthese* 17: 304–323. (Reprinted in Davidson 1984, 17–36.) https://doi.org/10.1007/BF00485035

Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Dummett, M. 1959. "Truth." *Proceedings of the Aristotelian Society* 59(1): 141–162. https://doi.org/10.1093/aristotelian/59.1.141

Fine, K. 2007. *Semantic Relationism*. Malden, MA: Wiley Blackwell. https://doi.org/10.1002/9780470690826

Geach, P. T. 1972. "Czemu zdanie nie jest nazwa." *Studia Semiotyczne* 3: 13–21.

Geach, P. T. 1982. "Truth and God." *Proceedings of the Aristotelian Society* 56: 83–115. https://doi.org/10.1093/aristoteliansupp/56.1.83

Groenendijk, J., and M. Stokhof. 1991. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14(1): 39–100. https://doi.org/10.1007/BF00628304

Heim, I. 1982. "The Semantics of Definite and Indefinite Noun Phrases." Ph.D. thesis, University of Massachusetts.

Lee, G. 2006. "The Experience of Left and Right." In *Perceptual Experience*, ed. T. Gendler and J. Hawthorne, 291–315. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199289769.003.0009

Leibniz, G. W. 1716. "Fourth letter to Clarke." (Reprinted in Leibniz and Clarke 2000, 22–28.)

Leibniz, G. W., and S. Clarke. 2000. *Correspondence: G. W. Leibniz and Samuel Clarke*. Indianapolis: Hackett Publishing.

Lewis, D. K. 1975. "Languages and Language. In *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science*, ed. K. Gunderson, vol. 7, pp. 3–35. Minneapolis: University of Minnesota Press.

Lewis, D. K. 1980. "Index, Context, and Content." In *Philosophy and Grammar*, ed. S. Kanger and S. Ohman, 79–100. Dordrecht: D. Reidel. https://doi.org/10.1007/978-94-009-9012-8_6

Lewis, D. K. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61(4): 343–377. https://doi.org/10.1080/00048408312341131

Massey, G. J. 1978. "Indeterminacy, Inscrutability, and Ontological Relativity." In *Studies in Ontology*, ed. N. Rescher, 43–56. American Philosophical Quarterly Monograph Series, no. 12. Malden, MA: Basil Blackwell.

McGee, V. 2004. "Ramsey's Dialetheism." In *The Law of Non-Contradiction: New Philosophical Essays*, ed. G. Priest, J. Beall, and B. Armour-Garb, 276–291. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199265176.003.0018

Merrill, G. H. 1980. "The Model-Theoretic Argument against Realism." *Philosophy of Science* 47(1): 69–81. https://doi.org/10.1086/288910

Ninan, D. 2010. "Semantics and the Objects of Assertion." *Linguistics and Philosophy* 33(5): 355–380. https://doi.org/10.1007/s10988-011-9084-7

Quine, W. V. O. 1992. *Pursuit of Truth*, rev. ed. Cambridge, MA: Harvard University Press.

Rabern, B. 2012. "Against the Identification of Assertoric Content with Compositional Value." *Synthese* 189(1): 75–96. https://doi.org/10.1007/s11229-012-0096-9

Stalnaker, R. 1978. "Assertion." In *Pragmatics*, ed. P. Cole, 315–332. Syntax and Semantics, vol. 9: Pragmatics, ed. P. Cole, 315–332. New York: Academic Press. https://doi.org/10.1163/9789004368873_013

Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*. London: Routledge.

Wittgenstein, L. 1953. *Philosophical Investigations*, 2nd ed., translated by G. E. M. Anscombe. Malden, MA: Blackwell.

Yalcin, S. 2014. "Semantics and Metasemantics in the Context of Generative Grammar." In *Metasemantics*, ed. A. Burgess and B. Sherman, 17–54. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199669592.003.0002

Yalcin, S. 2018. "Semantics as Model-Based Science." In *The Science of Meaning*, ed. D. Ball and B. Rabern, 334–360. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198739548.003.0012

Yalcin, S. 2021. "Context-Sensitivity by Way of Parameters." In *Oxford Studies in the Philosophy of Language*, vol. 2, ed. E. Lepore and E. Sosa, 267–304. Oxford: Oxford University Press.