Mark Schroeder
University of Southern California
July 18, 2011

is semantics formal?

In this paper I will be concerned with the question of the extent to which semantics can be thought of as a purely formal exercise, which we can engage in in a way that is neutral with respect to how our formal system is to be interpreted. I will be arguing, to the contrary, that the features of the formal systems which we use to do semantics are closely linked, in several different ways, to the interpretation that we give to those formal systems. The occasion for this question, and the main example that I will use to illustrate my answer to it, is the close relationship between the formal systems employed in recent statements of apparently competing accounts of epistemic modals with the dynamic, expressivist, and relativist theoretical paradigms.

The structure of the paper will be straightforward. In part 1, I will briefly introduce four theories of epistemic modals – one dynamic theory, two expressivist theories, and one relativist theory. Then in part 2 I'll show that one expressivist theory is formally equivalent to the dynamic theory, that the other is formally equivalent to the relativist theory, and that the two expressivist theories are themselves essentially notational variants. I'll use these facts to pose our central question: if these theories have so much formally in common, then doesn't that suggest that we can separate the task of constructing a formal semantics from the task of deciding between competing interpretations of it? Finally, in part 3 I'll answer that question in the negative. There are at least three reasons why formal semantics cannot be separated from questions of interpretation that are illustrated by the theories I introduce in part 1.

1 four theories

In the following four subsections I will be introducing four simple theories of the epistemic modal 'might'. All four theories are nondescriptivist, in the sense that they do not assign absolute truth-conditions to sentences involving wide-scope 'might' relative to contexts of utterance. It is important to distinguish the particular *theories* that I will introduce from the *paradigms* for semanic theorizing to which they belong. The three paradigms that I will be discussing in this paper are the *dynamic*, the *expressivist*, and the *relativist*. Each theory that I will introduce belongs to one of these paradigms, but is just one of many possible theories

that fall under the same paradigm. Indeed, I will introduce two different theories which fall under the expressivist paradigm. Though the two expressivist theories that I discuss in the main body of this paper are very similar in structure and commitments, it is important to realize that there are other possible expressivist theories which differ from them in important ways. (I'll return to this point in part 3.)

1.1 simple update semantics

The first nondescriptivist theory of 'might' that I want to introduce is borrowed from Frank Veltman [1996], and I will call it *simple update semantics* or just *update semantics*, for short. Update semantics belongs to the *dynamic* paradigm for semantic theorizing. The basic idea of the dynamic paradigm is that the job of semantics is to tell us how the successful utterance of a sentence should change the conversational context. The claim of the dynamic paradigm to be able to tell us something about *meaning* is due to the fact that the way in which successful utterances of sentences do change the conversational context clearly does seem to depend on what they mean. The dynamic theorist holds that this dependence is constitutive. Dynamic semantics is at least in principle a *non-descriptive* paradigm, because it leaves open the possibility that some sentences do not change the context by adding any information to that shared by the speakers in the context, but instead do something else. It is this possibility that is exploited by simple update semantics to offer a nondescriptivist account of 'might'.

To get the framework used by simple update semantics, we need just three steps. The first two steps are simplifications for purposes of the model, and the third step just applies the general idea of dynamic semantics, on the assumption of these two simplifications. First, we simplify by modeling contexts as states of information, interpretated as the information shared by the conversational participants. This assumption leaves out other possible parameters of the context. Second, we simplify again by modeling states of information by sets of possible worlds. This assumption washes out any differences in how confident the conversational participants are in different claims, and requires that the conversational participants' information is consistent and closed under logical consequence. Finally, since the job of semantics is to tell us how the successful utterance of a sentence should change the context, under these simplifying assumptions what we want is for our semantics to give us, for each sentence, a function from sets of worlds to sets of worlds. Such a function will tell us, for any context in which that sentence might be uttered, what the context will be like after it is successfully uttered.

With the help of a little bit of notation, we can now state the simple update semantics. Let i and j be the information states associated with arbitrary contexts, 'α' an arbitrary atomic sentence, $|\alpha|$ the

intension of 'α' (we'll assume for simplicity that no atomic sentence requires a nondescriptivist account), and p and q arbitrary sentences. We write '$\uparrow$p' for the function from contexts to contexts that is the semantic value of 'p', and use postfix rather than prefix notation. So '$i\uparrow p = j$' means that the semantic value of 'p' maps context i to context j.

simple update semantics

Atomic	$i\uparrow\alpha = i \cap \lvert\alpha\rvert$
Negation	$i\uparrow\sim p = i \setminus i\uparrow p$
Conjunction	$i\uparrow p\&q = (i\uparrow p)\cap(i\uparrow q)$
Might	$i\uparrow\Diamond p = \{w \in i: i\uparrow\sim p \neq i\}$

It is worth making several observations before we go on. First, note that all sentences that are free of '$\Diamond$' are assigned to functions that reduce the context set by some fixed set of worlds. This corresponds to the idea that sentences free of '$\Diamond$' are *descriptive*. In contrast, sentences of the form '$\Diamond$p' perform a *test* – they either return the original context set, or they return the empty set. Such sentences are *nondescriptive*. Rather than being associated with any piece of information that is communicated, they perform an alternative communicative function.[1]

1.2 first expressivist semantics: modeled on gibbard

On the face of it, expressivist is a very different paradigm for semantic theorizing than the dynamic framework. The main idea of the expressivist paradigm is that the job of semantics is to tell us, for each sentence 'p', what it is to believe that p – or perhaps more carefully, if the object language contains expressions the theorist would prefer not to use (such as slurs), to tell us, for each sentence 'p', what it take to accept 'p' (I'll stick to the first formulation in what follows). Expressivism's claim to tell us something about *meaning* is that what it takes to believe that p clearly does depend on the meaning of 'p'. Expressivism is at least in principle a nondescriptive paradigm, because it leaves open the possibility that for some sentences 'p', believing that p does not involve being confident in any particular piece of information. The expressivist views I'll be introducing in this section and the next exploit this feature by identifying believing that it might be that p with having failed to rule p out.

[1] It is also worth noting in passing that modeling information states by sets of worlds is very coarse-grained – indeed, too coarse-grained to deal with modals like 'probably'. But the compositional rule for negation relies heavily on modeling information by sets of worlds. This is a severe limitation on the prospects for generalizing the theory once the simplifying assumptions are dropped. For discussion of this problem see Schmitt [unpublished].

To get to the expressivist theory modeled on Gibbard, we again need just three steps. This theory is modeled on Gibbard, in the sense that it follows his method of world-norm pairs as described in Gibbard [1990], simply replacing *norms* with *information states*. In the first step, we model ordinary descriptive belief states by sets of worlds. In the second step, we model epistemic modal belief states by sets of something else – for us, that something else will be information-states, understood as sets of worlds. This encapsulates the idea that epistemic modal beliefs are holistic properties of one's belief state. Finally, in the third step we combine these two by modeling arbitrary states of mind by sets of world/information-state pairs. (I'll comment further on this in a moment – this is the hard step.)

We can now state the Gibbard-inspired expressivist theory. The notation I'll use is as before, but with the addition that we'll now use $[p]$ to denote what I'll call the *Y-expressivist* semantic value of 'p' – the appropriate set of $<w,i>$ pairs to model the state of mind that 'p' expresses – and let U be the set of all $<w,i>$ pairs.

Y-expressivist semantics

Atomic	$[\alpha] = \{<w,i> \in U: w \in \lvert\alpha\rvert\}$
Negation	$[\sim p] = U \setminus ([p])$
Conjunction	$[p\&q] = [p] \cap [q]$
Might	$[\Diamond p] = \{<w,i> \in U: \exists v \in i\ (<v,i> \in [p])\}$

Again, a few observations are in order – more this time, since it's still not quite clear exactly how a set of $<w,i>$ pairs represents a state of mind. First, note that whenever 'p' is free of '$\Diamond$', $[p]$ is *i-invariant*, in the sense that every world that gets paired with any information state by $[p]$ is also paired with every other information state by $[p]$. *i-invariant* sets of $<w,i>$ pairs are equivalent to sets of worlds, and so it is a consequence of this that sentences that are free of '$\Diamond$' correspond to sets of worlds. This is what makes them descriptive, in accordance with our first step.

In contrast, $[\Diamond p]$ is always *w-invariant*, in the sense that every information state that gets paired with some world by $[\Diamond p]$ also gets paired with every other world by $[\Diamond p]$. *w-invariant* sets of $<w,i>$ pairs are equivalent to sets of information states, and hence correspond to coarse-grained properties of information states – the kinds of state of mind associated with epistemic modals, in accordance with our second step. It is these two facts which explain why this theory can use a single common currency – sets of $<w,i>$ pairs – to represent both sets of worlds (and thereby descriptive beliefs) and also sets of information states (and thereby epistemic modal beliefs), in accordance with our third step. These are the facts exploited by Gibbard [1990] in order to make his system work.

This still, however, does not give us any way of interpreting what sort of state of mind is associated with an *arbitrary* set of <w,i> pairs – it only tells us what is going on with sets that are either w-invariant or i-invariant. To see how to give such an interpretation, we need two more observations. First, observe that if [p] is i-invariant (and thus equivalent to a set of worlds |p|, then an arbitrary information state i excludes all of the worlds excluded by |p| (that is, $i \subseteq |p|$) just in case $\forall w \in i$ (<w,i>∈[p]). So if an agent's belief state is characterized by i and [p] is i-invariant, then the agent believes that p just in case $\forall w \in i$ (<w,i>∈[p]).

Second, observe that if [p] is w-invariant, and hence equivalent to a set of information states, an arbitrary information state i is a member of this set just in case $\forall w \in i$ (<w,i>∈[p]) (note that the restriction on the quantifier is redundant). So if an agent's belief state is characterized by and [p] is w-invariant, then the agent believes that p just in case $\forall w \in i$ (<w,i>∈[p]). Since the same conditional allows us to characterize both descriptive and epistemic modal belief, we can generalize and say that in general, believing that p is just having one's information state belong to the class, $\{i: \forall w \in i$ (<w,i>∈[p])$\}$. This gives us a general way of mapping from sets of <w,i> pairs to states of mind, thus making good on the expressivist idea that what a semantic theory must do is tell us, for each sentence 'p', what it is to believe that p. Although this mapping may look *ad hoc*, it at least gets the right results in the special cases we've observed.

Although I've described this theory as 'Gibbard-inspired', careful readers may now recognize it as the first, preliminary, version of the theory stated in Yalcin [2007]. The *ad hoc*-looking mapping which we've constructed which tells us how to interpret an arbitrary set of <w,i> pairs as a belief state is just the metalanguage reflection of Yalcin's semantics for the object-language verb 'believes'. That's why I've called it the *Y-expressivist* theory.

I.3 an alternative expressivist semantics

Now that we have a way of interpreting arbitrary sets of <w,i> pairs in terms of constraints on an agent's overall belief state, it is easy to see that there is something strikingly arbitrary about the assignments of sets of <w,i> pairs to sentences in the Y-expressivist theory. That is because the mapping from [p] to the corresponding constraint on belief-states only pays attention to members of [p] for which $w \in i$. Consequently, whenever $w \notin i$, it makes no difference to which belief state is assigned to [p] whether <w,i> is included in [p] or not. But according to the Y-expressivist theory, *nearly every* sentence is assigned to a set

of <w,i> pairs which includes members such that $w \notin i$.[2] That is, nearly every value for [p] has redundant members which do no work in the semantic theory.

Since from the point of view of the interpretation of this framework it makes no difference which <w,i> pairs for which $w \notin i$ are included in [p], this makes all of the assignments of semantic values under the Y-expressivist theory look arbitrary. There are many other ways of assigning sets of <w,i> pairs which differ from it only for pairs for which $w \notin i$, and which are equally good, in that they generate the very same states of mind expressed by each sentence. Of all of these, the least arbitrary would seem to be to include *no* pairs for which $w \notin i$ in [p]. That is the idea of our *alternative* expressivist semantics.

Letting $\langle p \rangle$ denote the *Alt-expressivist* semantic value for 'p' and taking $\hat{U} = \{<w,i> \in U: w \in i\}$, there are two equivalent ways in which we can state the alternative expressivist semantics. First, we could define it derivatively, by letting $\langle p \rangle = [p] \cap \hat{U}$. Alternatively, we could simply replace 'U' by '$\hat{U}$' in the definitions for [p], above.

Alt-expressivist semantics

Atomic	$\langle \alpha \rangle = \{<w,i> \in \hat{U}: w \in \lvert \alpha \rvert\}$
Negation	$\langle \sim p \rangle = \hat{U} \setminus (\langle p \rangle)$
Conjunction	$\langle p \& q \rangle = \langle p \rangle \cap \langle q \rangle$
Might	$\langle \Diamond p \rangle = \{<w,i> \in \hat{U}: \exists v \in i\ (<v,i> \in \langle p \rangle)\}$

This alternative expressivist semantics is not formally equivalent to the Y-expressivist semantics. Although it is easy to construct a mapping from the Y-expressivist semantics to the alt-expressivist semantics, the Y-expressivist semantics contains extraneous information that plays no role in the interpretation of the system in terms of constraints on an agent's overall belief state. But though they are not formally equivalent systems, they are essentially notational variants, from the point of view of the expressivist interpretation. This is because on both views, the only cash value of a set of <w,i> pairs is to characterize a constraint on an agent's overall belief state, and both views agree exactly on this: $\{i: \forall w \in i\ (<w,i> \in [p])\} = \{i: \forall w \in i\ (<w,i> \in \langle p \rangle)\}$. So the restriction to $\hat{U}$ makes no difference whatsoever for the states of mind expressed by sentences, on this view. The Alt-expressivist theory is therefore just a less arbitrary version of the Y-expressivist theory.

[2] The simplest exceptions to this principle are of the form, 'p&~◊~p'.

1.4 simple relativist semantics

Finally, the fourth theory of epistemic modals that I want to compare in what follows is a kind of relativism. In general, basic idea of the relativist paradigm is that the job of semantics is to assign truth-conditions (or propositions that in turn determine truth-conditions), but that truth is sensitive to the context from which an utterance (or proposition) is assessed. The relativist's claim to tell us something about meaning is supposed to be the same as that of a non-relativist truth-conditional or propositional semantic paradigm, based on some conception of the relationship between truth and meaning. Like the dynamic and expressivist paradigms, relativism is in principle a nondescriptive paradigm, because it allows that some sentences may not be associated with any fixed piece of information, even in context.

To get to the simple form of relativism about 'might' that I will introduce here, there are again three main steps. First, we relativize truth-conditions to contexts of assessment. Second, we model the contexts of assessment by information states, and third, as before we model information states coarsely, by sets of possible worlds. We can use the same notation as before, but now we write $[\![p]\!]^{w,i}$ for the truth value of 'p' relative to world w and information state i.

relativist semantics

Atomic	$[\![\alpha]\!]^{w,i}$ = true iff $w \in \lvert\alpha\rvert$
Negation	$[\![\sim p]\!]^{w,i}$ = true iff $[\![p]\!]^{w,i} \neq$ true
Conjunction	$[\![p\&q]\!]^{w,i}$ = true iff $[\![p]\!]^{w,i}$ = true and $[\![q]\!]^{w,i}$ = true
Might	$[\![\Diamond p]\!]^{w,i}$ = true iff $\exists v \in i$, $[\![p]\!]^{v,i}$ = true

As before, a few quick observations are in order, although because I take it that the relativist framework is more familiar I will be brief. Note that all sentences free of '$\Diamond$' are assigned the same truth value relative to every information state, and hence that such sentences count as 'descriptive' in the same sense as before, in that they are associated with fixed pieces of information with absolute truth-values. In contrast, sentences of the form '$\Diamond$p' are assigned the same truth value relative to every world; this makes good on the idea that they are nondescriptive, in that they do not constrain the world in any way.

2 is semantics formal?

So far, all that I've done is to lay out the simplest versions of four nondescriptivist theories of epistemic modals. In section 2.1 I'll point out how much these theories have in common, and use that observation in section 2.2 to pose a natural question that seems to be on the tips of many tongues of those acquainted

with these systems: doesn't it look like we can separate the task of formal semantics from the question of how that system is to be interpreted?

2.1 translations

In part I I described four theories, belonging to three semantic paradigms. But these theories have much in common. At a formal level, all four theories treat the semantic values of sentences as relations between worlds and information states (understood as sets of worlds). And all four theories apply the same compositional principles to those relations. The only difference among the theories, at the formal level, is over *which* relations between worlds and information states are assigned to each sentence. One answer to that question is given by the Y-expressivist theory and the relativist theory, which are therefore formally equivalent, and a different but closely related answer is given by the Alt-expressivist theory and the simple update theory, which are therefore formally equivalent. So between the four theories there are only two formal systems, and those two systems agree entirely over their compositional principles.

Let's take these observations in turn. First, we may observe that it is easy to translate between the Y-expressivist theory and the relativist theory:

first translation

$[p] = \{\langle w,i\rangle : [\![p]\!]^{w,i} = \text{true}\}$
$[\![p]\!]^{w,i} = \text{true iff } \langle w,i\rangle \in [p]$

It is easy to verify that each of these translations are correct, by showing that they are correct for atomic sentences, and that the rules for '$\sim$', '&', and '$\Diamond$' all preserve the translation.

This shows that the Y-expressivist theory and the relativist theory use exactly the same formal system, but merely represent this system in different ways. This should not be a surprise, because as I noted earlier, the Y-expressivist theory is just the simpler theory presented in Yalcin [2007], but Yalcin uses the relativist's notation. The intertranslation also shows us that both of these theories can be thought of as assigning sentences to *relations* between worlds and information states, as their semantic values, and as employing the same composition principles on those relations. When such a relation is represented as a set of <w,i> pairs, the compositional principles are *complementation* for negation, *intersection* for conjunction, and *union* (by implication) for disjunction.

Similarly, it is easy to translate between the Alt-expressivist semantics and the simple update semantics:

second translation

$\langle p \rangle = \{<w,i>: w \in i\uparrow p\}$
$i\uparrow p = \{w: <w,i> \in \langle p \rangle\}$

As before, each of these translations can be straightforwardly verified by showing that they preserve the clauses for atomic sentences, and that each of the compositional rules (for '~', '&', and '◊') preserves the correctness of the translation.

This shows that the Alt-expressivist theory and the simple update theory employ the very same formal system, but merely represent that system in different ways. Again, this should not be terribly surprising. To see why not, recall our formula for mapping sets of <w,i> pairs to constraints on belief states: $\{i: \forall w \in i\ (<w,i> \in \langle p \rangle)\}$. When we introduced this formula in section 1.2, we did so by observing that it gets the right results for two of its special cases, but we weren't able to offer any intuitive gloss on why this is a meaningful formula, and so it looked somewhat *ad hoc*. However, this formula originally didn't come from the expressivist theory; it was borrowed by Yalcin [2007] directly from Veltman [1996]. In the framework of update semantics, what it says is $\{i: i\uparrow p = i\}$, which just means that you believe that p just in case finding out p would not change your existing information. From the point of view of update semantics, rather than looking *ad hoc*, this formula therefore actually looks highly natural.

In addition, our translation also shows that the formal system shared between the Alt-expressivist theory and the simple update theory can be thought of, like the formal system shared between the Y-expressivist theory and the relativist theory, as assigning sentences to relations between worlds and information states for their semantic values, and as employing the very same compositional principles on these relations – *complementation* for negation, *intersection* for conjunction, and *union* (by implication) for disjunction. The only difference between the two formal systems is that they assign *different* relations to sentences from one another. But even this difference is small; as we saw, the Alt-expressivist/update system assigns relations that are simply *restrictions* of the relations assigned by the Y-expressivist/relativist system to the set $\hat{U}$.

Add to this the fact that as we've already seen, from the expressivist point of view there are no possible grounds to prefer one formal system to the other except on the basis of the kind of aesthetic considerations that we used to motivate the Alt-expressivist theory, and it's clear that there is a very strong convergence, from a formal point of view, between the four nondescriptivist theories outlined in part I. It's hard not to find this convergence striking.

2.2 the natural question

Convergences like the ones I've been describing up to this point are the kind of thing which lead some theorists to suppose that at some level of description, the four views that I've described really articulate the *same* semantic theory, and that it is a separate task – perhaps for pragmatics – to tell us how to interpret the significance of that theory for linguistic communication. According to strong versions of this idea, formal semantics is *autonomous* from questions of interpretation. On this view, the relevant questions of formal semantics are just how many and which relata must be included in the relations assigned to sentences as their semantic values, and which relations get assigned to which sentences. This is sometimes equivalently[3] described in other terms as the view that the job for semantics is just to assign 'truth' values relative to an 'index', where the only questions for semantics are which parameters to include in the 'index', and to say how the 'truth' value of each sentence depends on the 'index'.

3 the heteronomy of formal semantics

As tempting as the formal similarities between the views described in part I of this paper might make the autonomy thesis, more careful observation shows that it is false. It is important to understand *why* it is false, however, and so in the next three sections I will lay out three different reasons why formal semantics needs to be informed by philosophical work on how those formal systems are to be interpreted. Together these three reasons paint a picture of an intimate relationship between formal semantics and conceptual questions about the interpretation of formal systems. All three reasons will be illustrated by the very same theories that we've been comparing from part I.

3.1 assignments of values

It is important not to be overly impressed by the fact that from the point of view of the expressivist interpretation, there is nothing to decide between the Y-expressivist theory and the Alt-expressivist theory. This fact makes it look like their respective formal frameworks have a great deal in common, and indeed they do. But they are still not the *same* formal framework. And that is important. For though there is nothing to choose between them from the point of view of the expressivist interpretation, there is everything to choose between them both from the point of view of the dynamic interpretation, and from the point of view of the relativist interpretation.

[3] And misleadingly, I would add, since there is nothing that such assignments have to do with our ordinary notion of truth.

Let's take the dynamic interpretation first. By applying our second translation scheme to the Y-expressivist theory instead of the Alt-expressivist theory, we can construct an alternative dynamic theory which is formally equivalent to the Y-expressivist and relativist theories. It looks like this:

butchered update semantics

Atomic	$i\uparrow_b \alpha = \lvert\alpha\rvert$
Negation	$i\uparrow_b \sim p = i \backslash i\uparrow_b p$
Conjunction	$i\uparrow_b p\&q = (i\uparrow_b p) \cap (i\uparrow q)$
Might	$i\uparrow_b \Diamond p = \{w: i\uparrow_b \sim p \neq i\}$

It is a simple exercise to verify that butchered update semantics employs the same formal system as the Y-expressivist theory, by using the translations

third translation

$[p] = \{\langle w,i\rangle: w \in i\uparrow_b p\}$
$i\uparrow_b p = \{w: \langle w,i\rangle \in [p]\}$

But butchered update semantics deserves its name. According to butchered update semantics, the result of a successful assertion of any sentence 'p' free of '◇' is to erase any information previously shared by the conversational participants, so that the only information they have is that p. Similarly, according to butchered update semantics, if 'p' is consistent with the previous information available in the context, then a successful utterance of '◇p' erases all of the previous information available in the context, so that nothing at all has been established in the conversation. Neither of these claims is remotely plausible. That is why from the dynamic point of view, it makes a very big difference *which* relation between worlds and information states is associated with each sentence, even in cases for which that does not make a big difference from the expressivist point of view.

Exactly analogous points go for the relativist interpretation. By applying our first translation scheme to the Alt-expressivist theory instead of the Y-expressivist theory, we can construct an alternative relativist theory which is formally equivalent to the simple update and Alt-expressivist theories. It looks like this:

butchered relativist semantics

Atomic	$[\![\alpha]\!]_b^{w,i}$ = true iff $w \in i \cap \lvert\alpha\rvert$
Negation	$[\![\sim p]\!]_b^{w,i}$ = true iff $[\![p]\!]_b^{w,i} \neq$ true
Conjunction	$[\![p\&q]\!]_b^{w,i}$ = true iff $[\![p]\!]_b^{w,i}$ = true and $[\![q]\!]_b^{w,i}$ = true
Might	$[\![\Diamond p]\!]_b^{w,i}$ = true iff $w \in i$ and $\exists v \in i$, $[\![p]\!]_b^{v,i}$ = true

Again, it is a simple exercise to verify that butchered relativist semantics employs the same formal system as the Alt-expressivist theory, by using the translations

fourth translation

$\langle p \rangle = \{\langle w,i\rangle : [\![p]\!]_b^{w,i} = \text{true}\}$
$[\![p]\!]_b^{w,i}$ = true iff $\langle w,i\rangle \in \langle p \rangle$

However, again I think it is clear that the butchered relativist theory deserves its name. It is an important feature of the standard relativist theory that most sentences – including those used to state the theory itself – make claims whose truth is invariant across contexts of assessment. This is important, because it is what grounds the theory's claim not to amount to an incoherent kind of perspectivalism. But according to butchered relativism, *every* sentence is false when assessed from the point of view of an information state that rules it out – even the sentence that says that butchered relativism is true. This means that even proponents of butchered relativism must admit that it is false from the point of view of anyone who thinks that it is false. So again, it's clear that from the perspective of the relativist, it matters which relations are assigned to individual sentences, even in cases for which that does not matter for the expressivist interpretation.

3.2 compositional principles

In the last section I showed that even though our expressivist interpretation is indifferent between our two formal systems, the choice of formal system is very important from the point of view of both the dynamic and relativist paradigms, and that each of these paradigms requires a different answer to which values are assigned to individual sentences. This shows that even among a class of views which agree perfectly about what *kind* of semantic value is assigned, and agree perfectly about the compositional principles for sentential connectives, our choice of formal system still needs to be informed by our understanding of how that formal system is to be interpreted. In this section I'll provide a further illustration of why formal semantics is not autonomous from the question of how our formal system is to be interpreted, by arguing that our

choice of interpretation also influences our conception of which compositional principles it makes sense to employ.

Because the four theories that I introduced in part I all employ the same compositional principles, we know that the choice between the dynamic, expressivist, and relativist paradigms does not force an answer to the question of how our compositional principles are to work – or at least, does not force that answer to be other than the one that is shared among these four theories. But while these compositional principles *may* be used within each semantic paradigm, in this section I'll argue that whereas they are a particularly *good* fit for the relativist, there are *other* compositional principles that look particularly natural from the dynamic perspective, and the ones we've been using are actually a *poor* fit, from the expressivist point of view.

The first of these points should be familiar to anyone acquainted with work within the dynamic paradigm. For whereas simple update semantics employs a *commutative* rule for conjunction, the normal way of treating conjunction in the dynamic paradigm is *sequential* and therefore *non-commutative*:

Dynamic Conjunction $i\uparrow p\&q = (i\uparrow p)\uparrow q$

When p and q are both free of '$\Diamond$', this compositional rule for '&' and the principle included in our simple update theory agree. But for sentences involving '$\Diamond$', they can disagree. So if we substitute *Dynamic Conjunction* for the *Conjunction* rule in simple update semantics, we no longer get a theory that employs the same formal system as the Alt-expressivist theory, because it requires different compositional principles.

Dynamic Conjunction does not make a lot of sense from a relativist point of view. After all, the relativist's theory appeals to truth, and truth *does* obey the principles of complementation, intersection, and union. The sentence '$\sim$p' is true just in case 'p' is not true, 'p&q' is true just in case 'p' and 'q' are both true, and 'p$\vee$q' is true just in case either 'p' is true or 'q' is true. So the relativist interpretation makes the compositional principles of complementation, intersection, and union look very compelling. But from the dynamic point of view, the commutative *Conjunction* rule just looks like one possibility among others, and *Dynamic Conjunction* is an extremely natural-looking competitor. It is natural, because on the dynamic picture, the semantic role of a sentence is a matter of how it changes the context, and *Dynamic Conjunction* treats conjunctions in the same way as sequential utterances of their conjuncts (which, intuitively, is what they are). Indeed, as Gillies [2004] has shown, there are some very striking virtues to a dynamic treatment of epistemic modals and conditionals that incorporates *Dynamic Conjunction*.

What we've just seen is that the interpretation that we give to our formal semantic system can affect which compositional principles are natural. The sequential rule for conjunction is extremely natural in the dynamic framework, but not at all natural in the relativist framework. This shows both that there are alternative compositional rules to those shared by the theories in part I, and that whether these are visible to us will depend on which interpretative paradigm we adopt. We can illustrate this same point, and strengthen it, by consideration of the expressivist paradigm. Though both of the expressivist theories described in part I employ the compositional principles of complementation, intersection, and union, the core of the difficulties involved with the traditional Frege-Geach problem for expressivism have to do with the poor fit between these compositional principles and the expressivist paradigm.

I'll illustrate this poor fit here in two steps; interested readers should consult Schroeder [unpublished] for further discussion and detail. The first step is simple. For the expressivist, the goal of a semantic theory is to assign each sentence 'p' to what it is to believe that p. But belief states do not obey the principles of complementation, intersection, and union: believing that ~p is not the same as not believing that p, believing that p&q is not the same as believing that p and believing that q, and believing that p∨q is not the same as either believing that p or believing that q. This is the core of the tension between the expressivist paradigm and compositional principles that work by complementation, intersection, and union.

Of course, if we apply those compositional principles *indirectly*, to values which determine belief states, rather than to the belief states themselves, then we can get around some of the most egregious problems. Both of our expressivist theories from part I do exactly this. But – and this is the second step – both of the expressivist theories in part I still predict that believing that p&q is the same as both believing that p and believing that q; both predict that believing that ~◊p is the same as not believing that ◊p, and both predict that believing that p∨◊q is the same as either believing p or believing ◊q. These predictions are residue of the compositional principles that they apply, and evidence of the poor fit of those principles with the expressivist paradigm.[4]

So although there are *versions* of expressivism that do employ these principles, I believe that these particular expressivist theories are unnatural within the expressivist paradigm. In contrast, I believe that the formal framework of Biforcated Attitude Semantics, developed in Schroeder [2008], *is* natural from the expressivist point of view. For discussion of why, see Schroeder [2008], [2010], and [unpublished].

[4] For further discussion and proofs, see Schroeder [unpublished].

3.3 augmented object languages

In the last two sections I've argued, first, that even holding fixed a conception of what semantic values must be like and a view about how the compositional principles work, the choice of *which* values are associated with which sentences is deeply constrained by our philosophical interpretation of the formal framework, and second, that our choice of interpretation also affects both the availability and also the naturalness of difference choices for the compositional principles themselves. In this section I want to make one more important, but neglected, observation. And that is that even when our formal systems do overlap, that is often because these formal systems are incomplete.

As we've seen, the Y-expressivist theory and the relativist theory, described in section 1.2 and 1.4, share a single formal system, down to their specification of the exact semantic values for individual sentences. But the object languages over which these two systems are defined are limited in a particularly striking way. Neither object language is defined over the vocabulary that is used to state either the expressivist or the relativist interpretation. This fact is important for generating the apparent overlap between the two formal systems. For once we add the metasemantic vocabulary to the object language, we will get formal systems which diverge.

Assuming that we are engaged in an attempt to give a semantic treatment of natural language, the very language we speak, any addition of metasemantic vocabulary to our formal treatment of our object language must obey an important constraint:

key constraint

> Any formal semantic theory, when paired with an interpretation of that formal system, must allow us to consistently state that formal theory, that interpretation, and other independent important truths, using sentences in ways that are allowed by that interpretation given the values assigned by that formal theory.

In other words, it is a problem with either a formal theory or an interpretation of its significance, if using the language in ways that are licensed by that theory does not allow us to consistently state the theory without rejecting important independent truths.

It is because of the key constraint that we know something about how the semantic clauses for the verb 'believes' must go, if added to the Y-expressivist theory. Since our metalanguage interpretation of the significance of the semantic value [p] is that an agent believes that p just in case her belief state is characterized by an information state contained in $\{i: \forall w \in i\ (\langle w,i\rangle \in [p])\}$, there is only one clause for

'believes' that respects this metalanguage interpretation. For ease of stating it, I'll use i_S^w to denote the information state that characterizes S's total belief state at world w:

Belief $[\text{S believes that p}] = \{<w,i> \in U: \forall v \in i_S^w\ (<v,i_S^w> \in [p])\}$

As should be expected, this clause characterizes [S believes that p] as i-invariant, and so as equivalent to a set of worlds – the worlds at which S's belief state satisfies the condition for believing that p, according to our mapping from values of [p] to belief states. It is important to note that this *has* to be the Y-expressivist's clause for 'believes'; any other clause would violate the key constraint.

It should be no surprise, therefore, that *Belief* is exactly the treatment for the object-language attitude verb 'believes' by Yalcin [2007]. This object-language treatment and our metalanguage gloss on the relationship between the value of [p] and what it is to believe that p necessarily go in hand, for theorists who aspire to be giving an account of the very language they speak. However, although Yalcin [2007] implies that he is sympathetic to the expressivist interpretation of his formal system, he does not insist on it, leaving open that it might be given a relativist interpretation. So how well does *Belief* fit together with the characteristic metasemantic claims of the relativist?

According to the relativist's metasemantic claims, each declarative sentence expresses a single fixed content in context – a single object of belief or assertion whose truth value nevertheless depends on the point of view from which it is assessed. So, for example, the idea is that '◇p' has a single content that may be believed by thinkers who occupy contexts of assessment with different information states, but this same belief may be true when evaluated with respect to some of these contexts but false when evaluated with respect to others. The relativist translation of *Belief*, however, says:

Belief $[\![\text{S believes that p}]\!]^{w,i} = \text{true iff } \forall v \in i_S^w\ ([\![p]\!]^{v,iSw} = \text{true})$

So in particular, for '◇p' what this says is that

◇Belief $[\![\text{S believes that } \Diamond p]\!]^{w,i} = \text{true iff } \forall v \in i_S^w\ (\exists x \in i\ [\![p]\!]^{x,iSw} = \text{true})$

But the first quantifier is redundant, so what this really says is that

◇Belief $[\![\text{S believes that } \Diamond p]\!]^{w,i} = \text{true iff } \exists x \in i\ [\![p]\!]^{x,iSw} = \text{true}$

In other words, this says that S believes that ◇p just in case ◇p is true with respect to her existing information. So on this view, differences in truth value for the same proposition across different contexts of assessment simply track differences in belief. But this doesn't fit at all with the relativist's claim to be talking about relative *truth*, which should at least be more transcendent that actual belief.

There should be no puzzle about why the very semantic clause for 'believes' that is needed by Y-expressivism fits poorly with relativism. It is because expressivism and relativism make competing claims about the significance of the attribution of a particular semantic value to a particular sentence, even when they employ the same underlying formal system. It is a direct consequence of this that they need to assign different semantic values to the sentences that may be used to formulate their competing claims. Consequently, they can only employ the exact same formal system for fragments of the language that do not include their metasemantic vocabulary. This illustates yet a third way in which formal semantics cannot be autonomous. Any formal semantic treatment of vocabulary that is used in the metasemantic interpretation of formal framworks needs to inform and be informed by the choice of metasemantic interpretation.

4 wrapping up

In this paper I used four very simple nondescriptivist theories of epistemic modals in order to pose the question of the extent to which formal semantics is autonomous from the conceptual project of interpreting the formal framework in a particularly sharp way. I used those very same theories, however, to show that even on such fertile ground for optimism about the autonomy of formal semantics, there are a whole range of obstacles to conducting formal semantics in a way that is autonomous from interpretive questions.

One of the main features of the discussion in this paper is that it has been conducted entirely through discussion of what I've called *non-descriptive* semantic paradigms, using that term broadly to include relativism. It is important to note, however, that even if no non-descriptive paradigm is correct, and hence the *best* way to construct a semantic theory follows a common kind of structure, that doesn't show that formal semantics is free of interpretive commitments – it just shows that some interpretive commitments are better than others.

If you've found the main points that I've been making in this paper obvious, then congratulations – I believe that they should be. But in a climate in which different theorists have reached for familiar and similar-looking formal tools to develop theories with different interpretive commitments, it is important to

clearly understand the intimate relationship between formal semantics and its conceptual interpretation. A less than full appreciation of this intimate relationship can only narrow our understanding of the possibilities for both projects.[5]

references

Egan, Andy, John Hawthorne, and Brian Weatherson [2005]. 'Epistemic Modals in Context.' in G. Preyer and P. Peter, eds., *Contextualism in Philosophy*. Oxford: Oxford University Press.

Gibbard, Allan [1990]. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.

Gillies, Anthony [2004]. 'Epistemic Conditionals and Conditional Epistemics.' *Noûs* 38(4): 585-616.

MacFarlane, John [2011]. 'Epistemic Modals are Assessment-Sensitive.' In Egan and Weatherson, eds., *Epistemic Modality*. Oxford: Oxford University Press.

Schroeder, Mark [2008]. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.

_____ [2010]. 'How to Be an Expressivist About Truth.' In *New Waves in Truth*, edited by Nikolaj Jang Pedersen and Cory Wright. New York: Palgrave MacMillan, 282-298.

_____ [unpublished]. 'Attitudes and Epistemics.' Unpublished paper.

Unwin, Nicholas [1999]. '*Quasi*-Realism, Negation and the Frege-Geach Problem.' *The Philosophical Quarterly* 49(196): 337-352.

Veltman, Frank [1996]. 'Defaults in Update Semantics.' *Journal of Philosophical Logic* 25(3): 221-261.

Yalcin, Seth [2007]. 'Epistemic Modals.' *Mind* 116(4): 983-1026.

_____ [forthcoming]. 'Nonfactualism Concerning Epistemic Modality.' In Egan and Weatherson, eds., *Epistemic Modality*. Oxford: Oxford University Press.

[5] Special thanks to Scott Soames, Derek Ball, Ben Lennertz, and an audience for the talk this paper is based on at the Arché Workshop on Propositions and the Aims of Semantics at Ardtornish, Scotland, in May 2011.