**Epistemic Modals, Relativism and Assertion**[1]
Andy Egan
University of Michigan/Australian National University

*Introduction*

I think that there are good reasons to adopt a *relativist* semantics for epistemic

modal claims such as "the treasure might be under the palm tree", according to which

such utterances determine a truth value relative to something finer-grained than just a

world (or a <world, time> pair).[2]  Others have argued for relativist semantics in other

areas.[3]  Anyone who is inclined to relativise truth to more than just worlds and times

faces a problem about assertion.  It's easy to be puzzled about just what purpose would

be served by assertions of this kind, and how to understand what we'd be up to in our *use*

of sentences like "the treasure might be under the palm tree", if they have such peculiar

truth conditions.

In the first section of the paper, I provide a very quick argument to motivate a

relativist view of epistemic modals.  I'll be talking about 'might', but nothing much

hangs on this choice of examples.  In fact, the intuitions that I'm appealing to are

probably stronger for 'probably'.  So if you think I might be wrong about 'might', you'll

probably be happier to go along if you think about the parallel argument for 'probably'

instead.

[2] See Egan, Hawthorne, and Weatherson (forthcoming) for more detailed arguments for this than I will present here.  See also MacFarlane (forthcoming a), to which both that paper and this one owe a great deal.
[3] For an argument for relativism about knowledge attributions, see MacFarlane (forthcoming b).  For a relativist semantics of predicates of personal taste, see Lasersohn (MS).

I won't be trying to establish an airtight case for relativism – my only goal here is to show why such a view might be attractive, and worth looking at fairly closely. (I'll do that by pointing to a somewhat puzzling phenomenon that relativism would provide a nice explanation of if it were true.) I'll then sketch my preferred relativist theory of epistemic modals in a bit of detail. The last order of business will be to bring out, and resolve, the problem that relativists face in making sense of the role of assertions with relativist truth conditions. Solving this problem should be helpful in two ways: first, it eliminates an apparently forceful objection to relativism, and second, spelling out the relativist account of assertion and communication will help to make clear just what the relativist position is, exactly, and why it's interesting.

## 1. Motivation for Relativism: Eavesdroppers

My goal in this section is to provide some motivation for a relativist theory of epistemic modals, according to which a single utterance can be true relative to one evaluator and false relative to another. My own favorite motivation for relativism comes from a family of eavesdropping cases, such as the one that follows:

James Bond has just returned to London after a long day of infiltrating SPECTRE's secret base in the Swiss Alps, planting a bug in the main conference room, and slipping out by night after leaving persuasive but misleading evidence of his presence in Zurich. Sipping martinis in MI6 headquarters while monitoring the newly placed bug, Bond and his CIA colleague Felix Leiter overhear a conversation between Blofeld and his second in command, Number 2.

Some facts to bear in mind about everyone's epistemic situation: Both Bond and Felix, obviously, know that Bond is in London. It is compatible with everything that Blofeld and Number 2 know that Bond is in Zurich.

In the course of a discussion of the pros and cons of various nefarious plans, Number 2 says to Blofeld, "Bond might be in Zurich". Since we will be referring back to it a fair bit in what follows, let us name this utterance ZURICH. (Note – we're naming the particular, dated, context-bound utterance, not the sentence type.)

Consider two of our characters' responses after hearing ZURICH. Don't theorize yet. We'll do that in a moment – right now we're just gathering data, not making an argument. So for now, just notice your intuitive judgments about the appropriateness of Blofeld's and Felix's responses:

In SPECTRE's alpine headquarters, Blofeld turns to Number 2 and says, "that's true". Blofeld's response is clearly appropriate. It would be equally appropriate for Blofeld to say, "you're right", "you have just expressed a truth", "ZURICH is true" or "what you have just said is true". It would be equally appropriate for Blofeld to say such things if he were listening in from a great distance while Number 2 was talking to someone else, rather than sitting in the room participating in the conversation.

Now suppose that, in London, upon hearing Number 2's utterance of "Bond might be in Zurich", Felix turns to Bond and says, "that's true." Is Felix's response appropriate? I think it's pretty clear that it is not. It sounds, to my ear at least, pretty bad for Felix to respond to ZURICH with this sort of agreement, and Bond would be right to react with some measure of surprise and puzzlement. Felix should, if challenged, backtrack and say instead that ZURICH is false. (It would have been just as bad for Felix to say any of the

following: "what Number 2 just said is true", "he's right, you know", "ZURICH is true", or "Number 2 has just expressed a truth".  Note also that it would have been just as bad to say all of these things if Felix and Bond had been hiding under the table in the conference room rather than listening in from London.)

Not everyone shares this intuition.  More on this later.  For now, let me just reiterate something I mentioned in the introduction: most people's intuitions are stronger about the case where Number 2 says "Bond is probably in Zurich".  Call this utterance P-ZURICH.  If you have the relevant intuition about Felix's saying "that's true" in response to P-ZURICH, but not in response to ZURICH, then think about the P-ZURICH case instead.  (My hope is that, in addition to providing an alternative case to use to motivate relativism, thinking about P-ZURICH will help to pump your intuitions about ZURICH.)

In short: Felix ought not to say that ZURICH is true.  In fact, Felix ought to say that ZURICH is false.  Blofeld ought not to say that ZURICH is false.  In fact, he ought to say that ZURICH is true.[4]

So much for gathering intuitions.  On to the argument.

*Appropriate* attributions of truth values to utterances are, in general, also *accurate*. This isn't a claim about the right *analysis* of appropriateness.  All I'm relying on here is that the appropriate attributions of truth and falsity to utterances are pretty nearly

---

[4] Have I cheated by insisting that 'ZURICH' name Number 2's particular utterance, and then appealing to intuitions about the appropriate attribution of truth and falsity that are really about whether the sentence-type, "Bond might be in Zurich", would express a truth if uttered in the attributor's context? I don't think so.  We are, in general, pretty good at distinguishing between the truth of a particular utterance and the truth of an utterance of a sentence of the same type in our own context.  Consider paradigm indexicals like "I" and "here".  If, seeing Jaws walk into the room, Blofeld says "Jaws is here", it's appropriate for Felix to say "that's true", "what Blofeld said is true", etc., inappropriate for him to say "that's false", "what Blofeld said is false", etc.  This despite the fact that the sentence type, "Jaws is here" would express a falsehood if uttered in Felix's context.   Similarly if Blofeld says "I am in Switzerland".  In these cases, our intuitions about the propriety of attributions of truth or falsity really do track the truth and falsity of the *particular utterance*, not whether an utterance of a sentence of the same type would be true on the attributor's lips.  If we're going to say that something different is going on in the case of ZURICH, we will need to provide some reason why this should be so.

coextensive with the accurate attributions, and that this isn't just some weird coincidence: there are some systematic reasons why that's so. The cases where we find appropriateness without accuracy are cases where something has gone wrong. So if we're going to say, in some particular case, that we have appropriateness without accuracy, we need to provide some account of just what's gone wrong.

Here, then, are some plausible-looking claims:

1) It's appropriate for Felix to say that ZURICH is false.

2) It's appropriate for Blofeld to say that ZURICH is true.

3) Appropriate attributions of truth and falsity are, absent some defeater, accurate.[5]

4) There are no defeaters in these cases.

If all of 1-4 are true, then it looks like we had better be relativists – the only way that all of this makes sense is if ZURICH has different truth values relative to Felix and Blofeld. We could react to this in one of two ways: we could take it as an argument for relativism, or we could take it as an argument against the conjunction of 1-4.

If we're going to take the fact that 1-4 jointly commit us to thinking that utterances can have different truth values relative to different evaluators to be an argument that not all of 1-4 are true, the best candidates for rejection seem to be (1) and (4).

We could reject (1). The intuitions here probably aren't as secure as the ones that support (2), so perhaps we ought to reject them. This response is less plausible for

---

[5] I don't mean anything fancy or technical by 'defeater' here. All I mean is that, absent some story about what's gone wrong, we ought to take appropriate attributions of truth and falsity to be accurate.

"probably" than for "might". So this doesn't seem to be a good *general* strategy for avoiding eavesdropping-based arguments for relativism.[6]

We could reject (4), by providing some defeater that explains why one or the other of the attributions is appropriate but not accurate. It's tempting to think that, in this sort of case, there *must* be a defeater. In order to say that Felix's attribution of falsity to ZURICH and Blofeld's attribution of truth to ZURICH are not only appropriate, but also accurate, we would have to say that ZURICH – the dated, context-bound utterance, not the sentence type – is true for Blofeld, but false for Felix. And that doesn't even make sense (or so it might appear, anyway). So whatever fancy footwork we have to do in order to make it turn out that one or the other of the attributions is appropriate, but not accurate, we'd better do it.

I am not going to try to argue against every maneuver one could make to explain what's gone wrong in one or the other of the cases. But all of these moves are going to

---

[6] The intuitions here are actually more nuanced than I've been letting on. I think it's pretty clear that, in Felix's situation, *agreement* with ZURICH is inappropriate – it's bad for Felix to say, "that's true", "that's absolutely right", etc. In my initial statement of the argument, I moved very quickly from this fact to the *appropriateness* of saying "that's false", etc. But the intuitions here are less clear. There is a certain hesitancy to attribute *error* to Number 2, and saying "that's false" sounds a lot like an attribution of error. (On a 'thin' conception of error, where all it takes to be in error is believing something that's false – there needn't be any bad reasoning, etc. going on.)

Normally, disagreement and attribution of error go hand in hand. But in this case they seem to pull in opposite directions. This is, I think, the source of some people's ambivalence about (1). They're hesitant to say that it's appropriate for Felix to say "that's false" because it seems *inappropriate* for Felix to accuse Number 2 of having made any sort of mistake, and the attribution of falsity seems like such an accusation. So when we think of Felix's attribution of falsity as indicating some criticism of Number 2, we're hesitant to say that it's appropriate. When we think of Felix's attribution as indicating a firm refusal to agree, we have no such hesitation. And we could restate 1-4 in terms of appropriate agreement, and a defeasible connection between appropriateness of agreement and truth, and between inappropriateness of agreement and falsity.

(This conflict of intuitions based on the pulling apart of appropriate agreement and appropriate accusation of error is, I think, further evidence for relativism. The only way for appropriate disagreement and appropriate attribution of error (on the thin conception of error) to come apart in any systematic way is for some utterances to be false relative to the evaluator and true relative to the speaker. Since it's false relative to the evaluator, it's appropriate to disagree. Since it's true relative to the speaker, it's inappropriate to attribute error.)

carry theoretical costs and commitments. Accepting these costs and commitments will be well motivated if relativism is bad enough.

The remainder of the paper is devoted to showing that relativism turns out not to be that bad. To show this, I will spell out a relativist view in some detail, and address what seems at first glance to be a very serious problem. With the view spelled out and the problem dealt with, I hope that the sort of relativism that I advocate will not seem like a deeply terrible thing, to be embraced only as a very last resort. None of what follows is intended as a positive argument for relativism - the only positive argument for relativism I'm going to make in this paper is the eavesdropping argument above. From now on, my task is purely explanatory and defensive.

If relativism turns out not to be such a bad thing after all, then it's not so clear that we really need to go in for the sort of fancy footwork that's required to undermine the connection between appropriateness and accuracy for Felix's and Blofeld's attributions. If it's not really so bad just to take Felix's and Blofeld's conflicting attributions of truth values to be not only appropriate, but also accurate, then we'll lose much of the motivation to reject relativism.

## 2. Self-Locating Content and Relativism

Cases like the eavesdropping example above provide some (admittedly non-conclusive) motivation for a relativist theory of epistemic modals. But all the motivation in the world will be no help if there is no coherent way to spell out a relativist theory. In this section I will present my preferred way of implementing relativism, and my preferred relativist semantics for epistemic modals.

The sort of relativist account that I favor is one on which some of our utterances express a kind of proposition that determines a truth-value not just relative to a world, but relative to a <world, time, individual> triple. (Call these *centered worlds*, since they single out not just a world, but also a privileged position – a *center* – within that world.) Such a theory of content is *relativist* because it says that some of our utterances have contents which can, for two individuals x and y, simultaneously determine different truth-values relative to x and y, even though x and y are worldmates.

Let us say that, when a particular centered world is the one that corresponds to my current position, I am *correctly located by* – in the sense that my (current, actual) location is fixed by – that centered world, and let us call whichever centered world I am correctly located by my *correct location*. No one is ever correctly located by more than one centered world at a time, and no one is ever correctly located by any one centered world for very long. I am, as I write this, correctly located by the centered world <@, Egan, 5:58pm>. You are, as you read this, correctly located by <@, you, whenever you're reading>. And as you read *this*, you're correctly located by some other centered world, whose time element is very slightly later.

A few words about how I'm thinking of *propositions*: Propositions are whatever kinds of things play the proposition role. That role has two parts: first, propositions are supposed to be the objects of propositional attitudes, and second, they're supposed to be the semantic values of declarative sentences.[7]

---

[7] A potential source of difficulty: Maybe there isn't any single kind of entity that does both of those things. This would, I think, be at least a little bit surprising, given the role that declarative sentences play in giving voice to our beliefs. Anyway, for purposes of this paper, I'm going to assume that the same things play both parts of the role. (Though the last sections of the paper can be read as, in part, a response to an argument that this assumption is mistaken.)

One standard view either identifies propositions with functions from worlds to truth values (or, equivalently, with sets of possible worlds) or, more modestly, takes propositions to be entities that determine such functions.

It's important to notice, though, that this is not plausibly a *definitional* claim. It's a bit of philosophical and linguistic theory.[8] To claim that propositions are, or determine, functions from worlds to truth values (or sets of worlds) is to claim that such functions or sets (or the entities that determine them) are the objects of belief, desire, etc., and the semantic values of declarative natural language sentences.[9] This is, at least at a first glance, a very well-motivated bit of theory. Propositional attitudes, and declarative sentences, are concerned with making distinctions between possibilities. Our beliefs distinguish between the possibilities we rule out and those that we take to be candidates for actuality. Our desires distinguish between the possibilities we favor and those that we despise. Our declarative utterances distinguish between the possibilities in which things

---

[8] It might be that, in some circles, it really is definitional of propositions that they determine functions from possible worlds to truth values. If you move in those circles, take me to be claiming that the contents of some natural language utterances are not propositions, but proposition-like entities that instead determine functions from <world, individual, time> triples to truth values. One can define one's technical terms in whatever way seems most useful, so it's perfectly legitimate to insist on the definitional claim about propositions. (Lewis does this in Lewis (1979).) What's important is that it's *not* legitimate to insist on *both* the definitional claim that propositions are things that determine functions from possible worlds to truth values, *and* the claim that propositions are the objects of propositional attitudes and the semantic values of (utterances of) declarative sentences. If you insist on the definitional claim, it's an open question in semantics and the philosophy of mind whether propositions play what I've called the proposition role. (Hence Lewis's argument that propositions aren't the objects of belief and desire.) If you insist that propositions are whatever plays the proposition role, then it's an open question in semantics and the philosophy of mind whether propositions determine functions from worlds to truth-values. Either will do for my purposes – the question I'm concerned with here is whether the contents of various bits of thought and language determine functions from worlds to truth values or from <world, individual, time> triples to truth values, and that question remains open however we decide to anchor our use of 'proposition'.

[9] In order to avoid having to write any more sentences as ugly as the preceding one, I will from now on be systematically sloppy about two distinctions that will not matter for our purposes. The first of these is the distinction between the claim that propositions *are* functions from worlds to truth values (or sets of worlds), and the claim that propositions merely *determine* such functions. Solely because it makes exposition easier, I will almost always talk in terms of identity rather than determination in what follows. The second distinction that I propose to be sloppy about is the distinction between functions from worlds to truth values and sets of worlds. I will move opportunistically between set-talk and function-talk depending on which makes it easier to express the point at hand.

are as we say and those in which they are otherwise. So the claim that the contents of propositional attitudes and declarative sentences are sets of worlds seems quite plausible.

Nonetheless it turns out to be false. (At least, it's false on its natural, strong reading, where the claim is that the objects of *all* propositional attitudes, and the semantic values of *all* declarative sentences of natural language, are sets of worlds.)[10]

Talk and belief about the time give us one reason to relativise the truth of propositions to something more than just a world. When I believe or assert that it's 12:00, the accuracy of my belief or assertion doesn't just depend on which world is actual – it also depends on which time is *present*. Since fixing which world is actual doesn't fix which time is present, the way in which my beliefs and assertions about the time distinguish between possibilities is not well-captured by thinking of their contents as sets of possible worlds. Things go better if we think of the contents of such beliefs and assertions as sets of <world, time> pairs. (We'll call these *tensed propositions*, and we'll call sets of worlds *possible worlds propositions*.) To adopt such a belief, or to make such an assertion, is to take a stand not just on which world is actual, but on which time is present. These considerations about belief and assertion provide a strong motivation for countenancing propositions whose truth is relative not just to worlds, but also to times.[11]

---

[10] The phenomena and examples that follow will be familiar to readers of Lewis (1979) and Perry (1979).

[11] Another motivation is that claims about the time exhibit a similar sort of behavior to what we saw from ZURICH in *Eavesdroppers* – it can be appropriate for listeners at different times to attribute different truth values to one and the same utterance.

One possible example of this is when Blofeld says, at noon, "it's lunchtime". Number 2, in the room at the time of utterance, should agree that Blofeld's utterance is true. Bond, listening to the recorded conversation later that night, should probably – though I think the intuitions here are less clear - say that Blofeld's utterance is false. (I don't think Bond should say that it *was* false – only that it *is* false, at the time when he's evaluating it. It would, for example, be appropriate for Bond to say "that's false" on hearing Blofeld's recorded utterance. Certainly it would be appropriate for him to respond "no" if Felix asked him, "is that true?" or "is what Blofeld said true?".)

I don't want to rest any weight on this – I only point it out as an interesting case that might be an instance of the same sort of phenomenon.

Cases like *Eavesdroppers* motivate relativisation to more than just times. Felix and Blofeld are both evaluating ZURICH at the same time, and so *time* relativity won't help us get the result that Felix is right when he says that ZURICH is false and Blofeld is right when he says that ZURICH is true. To get that result, we need the truth of epistemic modal claims to be relative not just to worlds and times, but also to individuals.

If we take ZURICH's semantic value to be a function from *centered worlds* to truth values, then we can say that ZURICH really is true relative to Blofeld and false relative to Felix. Here is how:

Pretty much everyone (at least, pretty much everyone who thinks that epistemic modals have truth conditions at all) thinks that the truth or falsity of epistemic modal claims turns, somehow or other, on what's known by some person or group. Or at least, that it turns on the epistemic state of some person or group - probably there's more to characterizing the relevant epistemic states than fixing what people *know*.

If we're looking for a possible worlds proposition – a set of worlds – to be the semantic value of a sentence like "Bond might be in Zurich", it will be something like: *the set of all worlds w such that it's compatible with all of the facts that group G stands in epistemic relation R to in w that Bond is in Zurich*.

There doesn't seem to be any single, universally appropriate way of picking out just which group G is. It's pretty clear that no single group will be appropriate in every context.

There also doesn't seem to be any single, universally appropriate epistemic relation R. In some cases, the fact that there's a map that, had I looked at it, *would* have told me that the treasure is in the mountains seems to be enough to make my claim that

*the treasure might be on the beach* false, and in some cases not. (This sort of case is due to Hacking (1967), and discussed in DeRose (1991).) These sorts of cases make it quite plausible that there is no single relation R that is always appropriate – different relations matter in different contexts. The idea, though, is pretty clear: *It might be the case that P* is true iff it's compatible with all of the facts that are within some group's *epistemic reach* that P, where what it takes to be within one's epistemic reach can vary across contexts.

If what we're looking for is not a possible worlds proposition, but a *centered worlds proposition* – a set of centered worlds – we will say something pretty similar. Here is the relativist proposal for the semantic value of "Bond might be in Zurich" that I favor: it's the set of all <w,t,i> triples such that it's compatible with all of the facts that are within i's epistemic reach at t in w that Bond is in Zurich.

In general, the proposal is:


RELATIVIST 'MIGHT': *It might be the case that P* is true relative to a centered world

<w,t,i> iff it's compatible with everything that's within i's

epistemic reach at t in w that P.[12]


So the proposition expressed by ZURICH is the proposition that is true of an individual (at a time, in a world) just in case they're in a certain sort of epistemic situation – one where all of the evidence that's accessible to them is compatible with Bond's being in Zurich. So, believers of this proposition have a belief about their own

---

[12] For 'probably', we might, as a first pass, try something like: *Probably P* is true relative to a centered world <w,t,i> iff the evidence within i's epistemic reach at t in w supports a greater than .5 credence in P.

epistemic circumstances: they take themselves to be in a situation in which none of the evidence that's within their reach rules out Bond's presence in Zurich. (Note that this is the sort of belief that one can pretty easily be wrong about, since there will very often be things that are within my epistemic reach that I'm not yet aware of – such as information on the treasure map in the Hacking-type case described above.)

This formulation of the truth conditions for ZURICH differs from the preceding, non-relativist one in three ways. First, and most obviously (but also most importantly), it's a specification of a set of *centered worlds* – of $\langle w,t,i \rangle$ triples – not a set of possible worlds. Second, and least importantly, it's stated in terms of *epistemic reach* rather than the relation variable R. Finally, there is no mention of a group G. More on this later.[13]

The thing to notice is that this lets us get the right result for *Eavesdroppers*. Suppose that what's in one's epistemic reach is just what one knows, which seems likely to be at least a pretty good approximation of the truth in this case. It is compatible with all of the facts that Number 2 knows (at the relevant time, in the relevant world) that Bond is in Zurich. So ZURICH is true relative to Number 2's context – relative to the

---

[13] Here is a rough preview: we don't need to include anything about G in this definition because we can do the relevant work with epistemic reach. Some individuals – notably those with whom I'm engaged in a cooperative conversation – are such that everything, or at any rate *pretty much* everything, that they know counts as within my epistemic reach (because, roughly, it would be sufficiently easy for them to tell me things). In fact, such individuals are likely to be such that whatever's within their epistemic reach is also within mine, since it's such a trivial matter for them to inform me about the things they find out about. Call such a person a *source* for me. (Note that I'll count as a source for myself.) The definition above, according to which "Bond might be in Zurich" expresses a proposition that's true relative to $\langle @, me, now \rangle$ iff it's compatible with everything that's within my (present, actual) epistemic reach that Bond is in Zurich, is equivalent to one that says that "Bond might be in Zurich" expresses a proposition that's true relative to $\langle @, me, now \rangle$ iff it's compatible with everything that's within the epistemic reach of any of my sources that Bond might be in Zurich. (Since something's being within reach of a source puts it within my reach, as well.) Claim: there's no extra work that we need groups to do that we can't do by exploiting this phenomenon of the extension of epistemic reach through one's sources. (We'll actually want something more complicated than being a source *simpliciter*. One reason for this is that some people will be such that they'll share information with me about topic A, but not about topic B. Such a person will be a source with respect to A, but not with respect to B. What we'll probably want to rely on is something like the following notion of being a *source with respect to P*: x is a source with respect to P for y iff x is disposed to share all of her P-relevant information with y. This may not be exactly the right notion, but it will be something in this neighborhood.)

centered world that correctly locates Number 2.  For short, it's true relative to Number 2.

It's not compatible with all the facts that Felix knows (at the relevant time, in the relevant

world) that Bond is in Zurich, since one of the things Felix knows is that Bond is in

London.  So ZURICH is false relative to Felix.  We get the same result even if we extend

our subjects' epistemic reach quite a bit.  (Though not, of course, if we extend it to *all* of

the facts.)

So what kinds of propositions do epistemic modal sentences express, on this

account?  They express propositions about our *epistemic situation*.  When I believe that it

might be that P, I believe something about my particular situation relative to the evidence

– namely, that none of the evidence that's within my epistemic reach rules out P.


*3. Self-Locating Assertion*

Now we have a relativist theory of epistemic modals in hand, which gives us, as

desired, the result that ZURICH is true for Blofeld and false for Felix.   In this section, I

am going to present a problem for this relativist theory – actually, a problem for relativist

theories in general.  The problem is that centered worlds propositions do not seem to be

well suited to play the role in assertion that the contents of declarative sentences ought to

play.  Making assertions with centered worlds contents seems, at least at a first glance, to

lead to disaster.

In this section, I will first review what I take to be an extremely plausible theory

of assertion, due to Robert Stalnaker (1978).  The theory, as formulated by Stalnaker,

does not allow for the assertion of centered-worlds propositions, but it can easily be

modified to do so, in a way that I will outline at the end of the section.

In the next section, I will present the problem: that we have good reason to think that centered worlds propositions would be systematically disastrous to assert, and that this gives us good reason to think that no declarative sentences of natural language express centered worlds propositions.

It's extremely attractive to think of what's going on in standard cases of assertion as adding the content of the assertive utterance to the stock of propositions that all of the parties to the conversation accept, take each other to accept, take each other to take each other to accept, etc. – in short, to what the parties to the conversation *presuppose*.

In general, what's asserted will be something that the asserter believes, and which the other parties to the conversation come to believe after they accept the assertion. So, when Alice makes an assertion, she utters some sentence the content of which she believes, and when Bob accepts her assertion, he comes to share the belief that Alice expressed with her assertion. Call this model the *belief-transfer model* of assertion. (See Stalnaker 1978.)

(I am abstracting away from ways of accepting assertions that fall short of belief, as when one accepts some assertion for present purposes, or in the course of a reductio, etc. But 'the belief-transfer-in-paradigm-cases model' is a bit unwieldy, and the complications don't really matter for our purposes. What's important about this Stalnakerian picture is that, in the core cases of successful assertion, assertion serves to transfer beliefs from the assertor to the members of her audience.)

Here is a more formal characterization of the Stalnaker framework: Successful assertions – assertions that are accepted by all parties to the conversation – distinguish between possibilities in a very specific sort of way. On the original framework, the sort

of possibilities we're distinguishing between are possibilities for the world. As the conversation progresses, if things go well, we collectively narrow down the range of live options for how the world might be. More propositions come to be presupposed by all parties to the conversation, and so the class of worlds that are compatible with the conversation's presuppositions gets smaller. Some terminology:

A proposition is *presupposed* by a speaker iff the speaker takes it that all parties to the conversation accept the proposition, take each other to accept it, take each other to know that everyone else accepts it, etc.

A presupposition is *shared* iff it's presupposed by all parties to the conversation.

A speaker's *context set* is the set of worlds compatible with all of a speaker's presuppositions.

When there are no unshared presuppositions – when everyone is presupposing all the same things – we can also speak of the *conversation's* context set, as the set of all of the worlds compatible with the shared presuppositions.

It's important to keep in mind – it will be particularly important later – that conversants typically presuppose a lot less than they believe. In many conversations, there isn't a big gap between acceptance and belief. But there's almost always going to be a very substantial gap between acceptance and presupposition. To presuppose some proposition P, it's not enough that I accept it – I've got to take you to accept it too. So if I know that P is something that's in dispute, or a point on which we differ, I *don't* presuppose it, even though I believe it. Presuppositions are (taken to be) common ground. Typically, what's common ground will allow for a lot more possibilities than would any of the conversants' beliefs – the various individuals who are parties to a conversation will

each have their own much more specific, and probably divergent, views. For example: Everybody agrees that there's a goat in the room. There's disagreement, however, about just where it came from. Since it's presupposed that there's a goat, the context set won't include any worlds where the room is completely goat-free. Since we don't all accept – and all know that we don't all accept – any particular hypothesis about its origin, the context set will include worlds in which each of the hypotheses that we haven't agreed to rule out is true. Many of those worlds will be ruled out by one or another party to the conversation's *beliefs*, but they won't be ruled out by the *presuppositions*.

Assertions of sentences that express self-locating propositions are problematic on this picture. The first problem is that the formal framework only talks about worlds, not about centered worlds, and so the model can't, without modification, accommodate the assertion of self-locating propositions. The second problem is that, even once we've made the obvious modifications, such assertions seem to have the wrong results, such that they would be crazy things to assert.

The natural way to modify the formal account of assertion in order to accommodate self-locating assertion is just to say all of the same things as before, except to everywhere replace "possible world" with "centered world". The resulting account is as follows: We take context sets to be sets of *centered* worlds. We take the contents of assertions to be centered worlds propositions, (which we'll continue to treat as sets of centered worlds). Successful assertions of some proposition P eliminate all the centered worlds that are not members of P from the context set. Our accounts of presupposition, and of shared presupposition, remain unchanged.

(This account can still duplicate everything the original account could do. We can, in particular, still model possible-worlds assertions, by restricting our attention to the centered-worlds propositions the truth-values of which are only sensitive to worlds – that is, the ones that, for any world x, include either all or none of the <w,t,i> triples with x in the world position. Call these the *boring* centered-worlds propositions, since they serve only to distinguish between worlds, and not between locations within them. Call the others the *interesting* centered-worlds propositions.)

I am going to assume, for the duration of this paper, that the Stalnakerian belief-transfer model of assertion presented above is correct. This is, I think, a pretty plausible assumption. It's also an assumption that makes my life as a relativist a bit awkward, as it is a key premise in the anti-relativist argument we'll be discussing in the next section.


*4. A Problem: Disastrous Assertions*

It's very easy to modify Stalnaker's model of assertion in order to allow for the assertion of self-locating propositions. But should we? Is self-locating assertion really something that we want to allow for? If it's not, then skepticism about self-locating linguistic content is in order.

If there's self-locating linguistic content (or at least, if there's self-locating content for declarative sentences), then there's self-locating assertion. So if there's no self-locating assertion, there's no self-locating linguistic content. And there are reasons to think that self-locating assertion is such a disaster that we should say that it never happens.

Following is a case that provides such a reason.

Believing that my pants are on fire is one of the paradigm cases of a self-locating belief. But it is not widely accepted that the English sentence, "my pants are on fire" expresses a self-locating proposition. Instead, it's standardly thought to express different *possible worlds* propositions on different occasions of utterance, depending on who says it. This is, I think, as it should be. Here is why:

Suppose that "my pants are on fire" expresses the self-locating proposition true in all and only the <w,t,i> triples such that i's pants are on fire at t in w. That is, it expresses the proposition that someone believes iff they take themselves to be a member of the unfortunate class of the burning-pantsed. (This is important. It will be *extremely* important in what follows to keep it in mind that to believe some self-locating proposition is to believe something about *your own* situation – it's to believe that you are correctly located by one of the <w,t,i> triples that is a member of the proposition.) Call this proposition *PANTS*.

(A notational stipulation: STRAIGHT CAPS are used to name utterances, such as ZURICH, *ITALIC CAPS* to name propositions, such as *PANTS*.)

Suppose that that's the right story about "my pants are on fire". Suppose also that the belief-transfer model of assertion is correct. Remember what the belief-transfer model says: in ordinary cases of successful assertion, there is a single proposition P which is (a) believed by the person making the assertion, (b) the content of the utterance used to make the assertion, and (c) comes to be believed, as a result of the assertion, by the other parties to the conversation. (The crucial troublemaking feature here will be that what the other parties to the conversation come to believe is the very same proposition that's expressed by the assertive utterance.) Given these assumptions, we should expect

the following sort of exchange when Bond hits Blofeld with an incendiary dart, setting Blofeld's pants on fire:

*Blofeld's Pants*

Blofeld: My pants are on fire!

Number 2: Aaagh! (Turns fire extinguisher on self.)

Of course, this isn't what happens. The conclusion that I want to draw from this is that utterances of "my pants are on fire" do not express *PANTS*. Here is a little argument to that effect:

(5) If utterances of "my pants are on fire" express *PANTS*, then, following a successful assertion of "my pants are on fire", all of the parties to the conversation will believe (or at least accept) *PANTS*.

(6) It's not the case that, following successful assertions of "my pants are on fire", all of the parties to the conversation believe (or at least accept) *PANTS*.

(C) Utterances of "my pants are on fire" do not express *PANTS*.

Premise (5) is a consequence of the belief-transfer model of assertion. Premise (6) is a plain fact about our actual practices of assertion. So, if we're going to accept the belief-transfer model of assertion, we had better reject the semantic theory that says that utterances of "my pants are on fire" express *PANTS*.

Another option here, of course, is to reject the belief-transfer model of assertion. I am no longer as convinced that this is a bad idea as I once was, but I still find the belief-transfer model very compelling, and so I'm inclined to try to stick to it if possible. (I also think that there are some principled things to say on its behalf as against the alternatives. Addressing these issues, though, would take us too far afield.) So having flagged this possibility, I will proceed by accepting the belief-transfer model as a constraint for purposes this paper, and explore the prospects for relativism, on the assumption that the belief-transfer model of assertion is correct.

We might think that it's not just the theory that says that "my pants are on fire" expresses *PANTS* that we ought to reject; we should reject any theory that says that *any* natural language utterance expresses *PANTS*. *PANTS* is, after all, a *systematically* bad thing to assert, because a *PANTS*-belief is, given the ways that burning pants actually tend to be distributed, a systematically bad belief to transfer to one's conversation partners.

We might also, looking at examples like this one, conclude that the same goes for self-locating propositions in general. They'll *all* be systematically bad things to assert, because transferring self-locating beliefs will *always* tend to systematically lead our conversation partners into predictable error in just the way that happens in *Blofeld's Pants*. If this is right, then we'll want to reject relativism across the board – since self-locating assertion is systematically disastrous, there's no call for self-locating linguistic content. And if there's no call for it, we should say that there isn't any.

I've just presented the argument relatively informally. In the next section, I present a more formal version of the same argument. Going through the formal details

will help to make clear just exactly what's gone wrong in *Blofeld's Pants*, and what's supposed to go wrong across the board in self-locating assertion.  This will help to clear the ground for a response.

*5. The Problem, Formally*

Before jumping in to the formal characterization of the problem, a bit of stage setting:

In the familiar case of possible-worlds assertion, things go bad when the context set doesn't include the actual world – the parties to the conversation are all led into error. Assertions that predictably exclude the actual world are defective.  That's what's wrong with assertions of "Goldfinger's yacht is longer than it actually is", for example.  There is something bad about assertions of contents that, predictably, would eliminate the actual world from the context set if they were accepted, and cooperative conversation partners will not make them.

In the case of self-locating assertion – assertion of centered-worlds propositions – things pretty clearly go bad when the context set excludes the correct locations of all of the parties to the conversation.  Again, the parties to the conversation are all led into error. Again, assertions of centered worlds propositions that, predictably, would exclude the correct locations of every party to the conversation if they were accepted, are to be avoided.  More interestingly, things *also* go bad when the context set includes the correct locations of some, but not all, parties to the conversation.  In this case, at least *some* party to the conversation is being led into error.  So we have a stronger principle: assertions of

centered worlds propositions that, predictably, would exclude the correct location of some party to the conversation if they were accepted, are to be avoided.

The Exclusion Constraint: *Don't make assertions that would, if accepted, exclude the correct location of some party to the conversation.*

Cooperative speakers will attempt to comply with the exclusion constraint. Successful assertions that violate the Exclusion Constraint will leave some party to the conversation 'stranded' – the context set after the assertion won't contain their correct location.

This is what goes wrong in *Blofeld's Pants*.  Let's characterize the context set before Blofeld's utterance – when it's an open question whose, if anybody's, pants are on fire – this way (for ease of presentation, times are suppressed in what follows):

In @, Blofeld's pants are on fire and Number 2's are not.

In $w_2$, Number 2's pants are on fire and Blofeld's are not.

In $w_3$, both Blofeld's and Number 2's pants are on fire.

In $w_4$, nobody's pants are on fire.

Let the initial context set be:

$\{<@, \text{Blofeld}>, <@, \text{Number 2}>, <w_2, \text{Blofeld}>,  <w_2, \text{Number 2}>,  <w_3, \text{Blofeld}>, <w_3, \text{Number 2}>, <w_4, \text{Blofeld}>, <w_4, \text{Number 2}>\}$

Now suppose Blofeld asserts *PANTS*, the proposition that includes (continuing to ignore times), all and only the <w,i> pairs such that *i*'s pants are on fire in *w*. Asserting *PANTS* will eliminate from the context set all of the <w,i> pairs that are not members of *PANTS*. So if Blofeld's assertion is successful, the resulting, post-assertion context set will be:

{<@, Blofeld>, <$w_2$, Number 2>, <$w_3$, Blofeld>, <$w_3$, Number 2>}

Number 2 has been *stranded*. His correct location - <@, Number 2> - has been eliminated from the context set. Further, this was *predictable*. Given Blofeld's situation – knowing that Number 2's pants were *not* on fire – he could have predicted that a successful assertion would eliminate Number 2's correct location from the context set. And cooperative speakers do not lead their conversational partners into predictable error.

*PANTS* has the same effects across a wide variety of contexts – the contexts in which we could assert *PANTS* without disaster are few, far between, and extremely strange. So *PANTS* is, in general, a bad thing to assert, not just a bad thing to associate with "my pants are on fire". After looking at this case, we might think that what goes for *PANTS* goes across the board. If that's so, then this sort of unsuitability for assertion will be a general feature of self-locating propositions – they will *all* be systematically bad things to assert, because assertions of self-locating propositions are always liable to leave some party to the conversation stranded. And if self-locating propositions are, in general, systematically bad things to assert, then they seem to be, in general, bad candidates to be

the contents of assertive utterances.  Which makes the prospects for any theory of self-locating *linguistic* content look pretty bad.

*6. The Response, Informally*

The problem that we're faced with is this: looking at *PANTS* gives us reason to suspect that assertions of self-locating propositions will, in general, be communicative disasters.  They're attempts to transfer self-locating beliefs, and self-locating beliefs are precisely the wrong sorts of beliefs to transfer, since their contents are liable to be true relative to the speaker but false relative to her audience.  (Or true relative to some, but not all, members of the audience.)

Happily, *PANTS* is not representative of self-locating propositions in general.  Though it would not often be useful to transfer belief in *PANTS*, it is often useful to transfer other self-locating beliefs to our conversation partners.  Think about, for example, the situation of a group of people lost in the woods, or in the Stanford library.  It would be extremely useful for the members of such a group to be able to collectively deliberate not just about which world they occupy, but also about their location within it.  And in order to do this, it looks like they'll need to be able to do some transferring of self-locating beliefs.

Consider the self-locating proposition that includes all and only the $<w,t,i>$ triples such that that $i$ is near Sydney at $t$ in $w$.  Call this proposition *SYDNEY*.  (And let's simplify in two ways: first, by ignoring the context-sensitivity of 'near' – suppose that the correct standards of nearness are fixed, and not in dispute – and second, by suppressing times from now on.  So I will, from now on, talk about centered worlds as $<w,i>$ pairs.)

Let me be clear about the role that our discussion of *SYDNEY* is meant to play.

First, some things that I am *not* going to be attempting to do in discussing *SYDNEY*. I am not going to argue that any utterance of natural language ever has *SYDNEY* as its content. I am, in particular, *not* about to argue that the English sentence "Sydney is nearby" ever expresses *SYDNEY*. Nor am I about to argue for any claims whatsoever about the semantics of the English expression "nearby".

What I want to establish with *SYDNEY* is only that not all centered-worlds propositions are systematically disastrous things to assert. There turns out to be a quite large, natural, commonly occurring class of situations in which asserting *SYDNEY* would, far from being disastrous, be an extremely useful thing to do. There is also a large, natural, commonly occurring class of situations in which asserting *SYDNEY* would lead to exactly the sort of communicative disaster that we found in *Blofeld's Pants*. Perhaps this means that no natural language sentence ever expresses *SYDNEY*. Perhaps not. The issue of whether any utterance of, say, "Sydney is nearby" ever expresses *SYDNEY* needn't detain us. What will be useful about discussing *SYDNEY* is that our discussion will reveal (a) that self-locating assertion *is*, at least in some circumstances, a potentially useful thing to go in for, and (b) just what conditions have to be met in order for it to be possible to assert some self-locating proposition without communicative disaster.

So, just when *would* it be useful and appropriate to assert *SYDNEY*?

There are certainly cases where it would be inappropriate to assert *SYDNEY* – it would be bad, for example, for Bond to assert *SYDNEY* when he and Felix both know that one of them is in Australia and the other is in North America. In general, it would be

bad if they're presupposing that they're far enough from each other that nothing that's near the one is near the other.

But there are other cases where it *would* be useful to assert *SYDNEY*: when, for example, Bond and Felix are piloting a mini-sub along the Australian coast. Since mini-subs are small, and the differences in location that matter to *nearness* in navigational contexts are large, it's overwhelmingly likely that, if something is near Bond, it's also near Felix, and vice versa.

In general, *SYDNEY* will be a potentially useful thing to assert when it's presupposed that the parties to the conversation are all close enough together that what's near one is also near the others. This happens pretty often. In *lots* of conversations the various parties are all pretty close together. (Until the fairly recent advent of fast long-distance communication, this was true in almost *every* conversation)

The difference between the cases in which *SYDNEY* is a potentially helpful thing to assert and the cases in which it's not is that in the good cases, there's a *presupposition of relevant similarity*, and in the bad cases there isn't. It's potentially helpful to assert *SYDNEY* when it's presupposed that all of the parties to the conversation are relevantly similar in their geographical location. It's not helpful to assert *SYDNEY* in the absence of such a presupposition. (Except insofar as making the assertion might bring the presupposition into effect.)

If that's right, then *SYDNEY* does quite a lot better than *PANTS* as a thing to assert – it very often serves our communicative interests to get our conversation partners to share our beliefs about our geographical location, and we're very often in conversations where it's presupposed that everybody is pretty close together.

The payoff of this is that it's not systematically bad to assert *SYDNEY*. So, it's not systematically bad to assert self-locating propositions in general. So, the very general argument against self-locating linguistic content, which says that since *all* self-locating propositions are communicative disasters, no self-locating proposition is a fit candidate to be the semantic value of a declarative sentence of natural language, fails.

It does not follow from this that any natural language sentence expresses *SYDNEY*. In particular, it does not follow that the English sentence, "Sydney is nearby" expresses *SYDNEY*. Is this a problem? No. The point of discussing *SYDNEY* was not to provide an example of a self-locating proposition that is in fact expressed in natural language. The point was twofold: first, to show that there is a market for self-locating assertion, by providing some cases in which it would be useful to transfer self-locating beliefs, and second, to draw out the importance of the presupposition of relevant similarity by contrasting the cases in which assertions of *SYDNEY* are disastrous and those in which they are not. Neither of these points depends on *SYDNEY* ever actually being expressed by any utterance in natural language.

## 7. *The Response, Formally*

Here is a formal characterization of what's happening with *SYDNEY*: Examination of the formal details will allow us to get a clearer picture of just what goes wrong in the bad cases, and what preconditions need to be met for the felicitous assertion of self-locating propositions.

There are some contexts in which asserting *SYDNEY* would lead to a violation of the Exclusion Constraint, and others in which it would not. The ones in which asserting

*SYDNEY* would not violate the Exclusion Constraint are those in which the

presupposition of relevant similarity (which will be given a formal characterization

shortly) is in place.

Here is a bad case:

*Bad SYDNEY*

It's presupposed that the parties to the conversation are far from each other.

The relevant worlds are:

In $w_1$, Bond is near Sydney and Felix is far away.

In $w_2$, Felix is near Sydney and Bond is far away.

In $w_3$, both Bond and Felix are far from Sydney.

The initial context set is:

$\{<w_1, \text{Bond}>, <w_1, \text{Felix}>, <w_2, \text{Bond}>, <w_2, \text{Felix}>, <w_3, \text{Bond}>, <w_3, \text{Felix}>\}$

Suppose someone asserts *SYDNEY*.  This will eliminate all of the $<w,i>$ pairs

where $i$ is far from Sydney in $w$.  So the post-assertion context set will be:

$\{<w_1, \text{Bond}>, <w_2, \text{Felix}>\}$

This situation has two bad features.  First, no matter which of $w_1$ or $w_2$ turns out to

be actual, one of the parties to the conversation has been stranded.  Second, the parties to

conversation know this – each knows that, if they're correctly located by one of the centered worlds in the context set, the person they're talking to isn't. The post-assertion context set has the regrettable feature that it only contains centered worlds in which the individuals at the center are in conversations with people whose correct locations are ruled out by the conversation's presuppositions. The combination of the initial context set – where the parties to the conversation are presupposing that they're far away from each other – and an assertion of *SYDNEY* had this result predictably. So it's bad to assert *SYDNEY* in contexts where it's presupposed that the parties to the conversation are far away from each other.

Here is a good case (like the mini-sub case in the preceding section):

*Good SYDNEY*

It's presupposed that the parties to the conversation are close together.

The relevant worlds are:

In $w_1$, Bond and Felix are both near Sydney.

In $w_2$, Bond and Felix are both far from Sydney.

The initial context set is:

$\{<w_1, \text{Bond}>, <w_1, \text{Felix}>, <w_2, \text{Bond}>, <w_2, \text{Felix}>\}$

Suppose someone asserts *SYDNEY*. This will eliminate all of the $\langle w,i \rangle$ pairs where $i$ is far from Sydney in $w$. So the post-assertion context set will be:

$\{\langle w_1, \text{Bond} \rangle, \langle w_1, \text{Felix} \rangle\}$

       Here there is no danger of leaving any of the parties to the conversation stranded.

       However, it also looks like the self-locating bit of the assertion isn't doing any work – we could have just asserted the possible worlds proposition that Sydney is near *Bond and Felix*. This is an unfortunate side effect of making the example simple. In order to see what's really going on, we need to complicate things a bit.

       What we need to do is keep track of the presuppositions about the boundaries of the conversation. Some terminology ('CW' will be used hereafter to abbreviate 'centered world'):

POINTS AT: A CW $\langle w, x, t \rangle$ *points at* a CW $\langle w, y, t \rangle$ iff x and y are parties to the same
       conversation at t in w.[14]

BEREFT: A CW is *bereft* in a context set C iff it points at a CW that is not a member of
       C.

TRAGIC: A context set is *tragic* iff it contains some bereft CWs.

TOTALLY TRAGIC: A context set is *totally tragic* iff it contains only bereft CWs.

---

[14] We'll need to complicate things again to deal with people who are participating in more than one conversation at a time, but never mind that for now.

It's obviously bad to make assertions that would leave a totally tragic context set. We know in advance that such an assertion will leave some party to the conversation stranded, so we know in advance that we will have violated the exclusion constraint.  It's also bad to make assertions that will leave a tragic, though not *totally* tragic, context set. If the context set is tragic, then it's compatible with the presuppositions that somebody's been stranded – that is, that there's some party to the conversation whose correct location has been eliminated from the context set.  Assertions that leave a tragic context set run the risk of violating the exclusion constraint.

Some more terminology:


SIMILAR: A CW x and CW y are *similar* with respect to P iff either both x and y are
      members of P or neither is.

PRESUPPOSITION OF SIMILARITY: A context set C is one in which a *presupposition*
      *of similarity with respect to P* is in effect iff, for every CW x that's a member of C,
      either every CW that x points to is a member of P or none is.  (That is: within the
      context set, P doesn't cut across conversations by including the correct locations
      of some, but not all, parties to some conversation.)


A *presupposition of relevant similarity* is in effect for some assertion iff there's a presupposition of similarity with respect to the proposition asserted.  When a presupposition of relevant similarity is in place, there's no danger of ending up with a tragic post-assertion context set.  (Unless the pre-assertion context set was already tragic.

So, there's no danger of going from a non-tragic pre-assertion context set to a tragic post-assertion context set.)

Now we can provide the more complicated version of *Good SYDNEY* that illustrates the importance of the self-locatingness of *SYDNEY*.

*Better SYDNEY*

Again, it's presupposed that the parties to the conversation are close together.

This time, it's not presupposed that the parties to the conversation are Bond and Felix. It is, however, presupposed that Bond and Felix are in a conversation with each other (and only each other) and Blofeld and Number 2 are in a conversation with each other (and only each other).

The relevant worlds are:

In $w_1$, Bond and Felix are both are near Sydney, while Blofeld and Number 2 are far away.

In $w_2$, Bond and Felix are both far from Sydney, while Blofeld and Number 2 are near Sydney.

In $w_3$, everybody is far from Sydney.

In $w_4$, everybody is near Sydney.

The initial context set includes all of the $\langle w,i \rangle$ pairs you get by freely combining $w_1, w_2$, $w_3$, $w_4$ with Bond, Felix, Blofeld, and Number 2. For each world $w$, the $\langle w$, Bond$\rangle$ and

<*w*, Felix> CWs will point at each other, and the <*w,* Blofeld> and <*w,* Number 2> CWs
will point at each other.

Suppose someone asserts *SYDNEY*. This will eliminate all of the <w,i> pairs where *i* is
far from Sydney in *w*. So the post-assertion context set will be:

{<$w_1$, Bond>, <$w_1$, Felix>, <$w_2$, Blofeld>, <$w_2$, Number 2>, <$w_4$, Bond>, <$w_4$, Felix>,
<$w_4$, Blofeld>, <$w_4$, Number 2>}

Since there are no bereft CWs in the post-assertion context set, the assertion of *SYDNEY*
in this context is felicitous. And here, the self-locatingness of *SYDNEY* was important –
unlike in *Good SYDNEY*, there are worlds ($w_1$ and $w_2$) such that some, but not all, of the
positions within them were ruled out by the assertion.

Again, the upshot is that *SYDNEY* is not a systematically disastrous thing to assert.
There is a class of quite commonly occurring situations in which asserting *SYDNEY*
would be not only non-disastrous, but very useful. So the general argument against self-
locating linguistic content is undermined   This is still not enough to show that any
natural language sentence (or utterance) expresses *SYDNEY*, though.

Return for a moment to our discussion of *PANTS*. Theories that claim that some
sentence or utterance expresses *PANTS* are (more or less) ruled out by the fact that
*PANTS* is a systematically disastrous thing to assert. Theories that claim that some
sentence or utterance expresses *SYDNEY* are not ruled out on *those* grounds, but this is

not the only hurdle facing a theory that attributes self-locating content to some bit of natural language.

We can summarize the lessons of our discussion of *SYDNEY* in terms of a test for theories of content. Assertions of some self-locating proposition P are only felicitous when there's a presupposition of similarity with respect to P in place. Suppose your theory says that a sentence S expresses, in some context C, some self-locating proposition P. If it turns out that S can be, as a matter of fact, felicitously asserted in C, then it had better turn out that P can be felicitously asserted in C. This means, at least, that it had better turn out that a presupposition of similarity with respect to P is in effect in C.[15]

In other words, felicitous assertion for *sentences* had better track felicitous assertion for the contents that our theory assigns them. If some proposition P would be a disaster to assert in some context, and some sentence S would not be a disaster to assert in that context, then P isn't the content of utterances of S in that context.

One consequence of this is what I will call *the assertability test*:

The Assertability Test: Suppose that a proposition P is the content of utterances of a sentence S in a context C. Then S is not assertable in C unless C is a context in which a presupposition of similarity with respect to P is in effect.[16]

---

[15] What about all the things Stalnaker says in "Assertion"? Isn't one of the points of that paper that we can go in for various pragmatic repair strategies to make otherwise infelicitous assertions okay? Yes, and that's an option here, too. So what we actually get is that, if S is felicitously assertable in C, it had better turn out *either* that P is felicitously assertable in C, *or* that some plausible pragmatic repair story is available. If the pragmatic stories turn out to be systematically available, then we could attribute self-locating contents willy-nilly, and appeal to pragmatic repairs in the cases where things would otherwise go bad. I'm nervous about the prospects of this sort of move, though, and I'd rather not have to make it.

[16] See footnote 15 for the necessary qualification about pragmatic repair strategies. With that qualification in place, the test works in general, not just for self-locating assertion, because there's *always* an assumption of relevant similarity in place for possible worlds propositions. Contexts in which we're leaving it open that some parties to the conversation are not worldmates are pretty hard to come by.

The theory that says that "my pants are on fire" expresses, in every context, the proposition *PANTS*, fails this test. It's felicitous to assert "my pants are on fire" in plenty of contexts in which no presupposition of similarity with respect to burning-pantsedness is in effect.

The theory that says that "Sydney is nearby" expresses, in every context, the proposition *SYDNEY*, also fails this test. It's felicitous to assert "Sydney is nearby" in plenty of contexts in which no presupposition of similarity with respect to location (or even with respect to proximity to Sydney) is in effect. This shows that passing the assertability test is a stronger requirement than just a lack of systematic disastrousness. The assertability test imposes a further requirement, that the cases in which assertions of P would be felicitous *line up* in the right way with the cases in which assertive utterances that, according to the theory, express P, would be felicitous.

Suppose (as I am supposing for purposes of this paper) that the belief-transfer model of assertion is non-negotiable. Then the theory according to which "my pants are on fire" expresses *PANTS*, and the more general theory according to which sentences containing indexicals express, in general, centered-worlds propositions, are non-starters because assertions of the relevant contents will, in very many cases, be systematically disastrous. The theory according to which "Sydney is nearby" expresses *SYDNEY*, and the more general theory according to which sentences containing "nearby" express, in general, centered-worlds propositions, are ruled out because they fail the assertability test: There are plenty of circumstances in which "Sydney is nearby" is felicitously assertable,

but *SYDNEY* is not. How, then, does the relativist theory of epistemic modals fare with respect to the assertability test?

*7. Relativism about Epistemic Modals and the Assertability Test*

On the view I favor, ZURICH expresses the centered-worlds proposition – call it *ZURICH* – which includes all and only the <w,t,i> triples such that it's compatible with all of the facts that are within i's epistemic reach at t in w that Bond is in Zurich. This theory fares extremely well, I think, on the assertability test.

It's quite plausible that, with regard to the self-locating propositions in question, the presupposition of relevant similarity is almost always in effect – that we almost always presuppose that we are, in relevant respects, in the same epistemic situation as the other parties to the conversation.

Why is this so plausible? Because it's quite plausible that, when you and I are engaged in a conversation, what's within your epistemic reach is also within mine, and vice versa. We are, after all, in a situation where it's easy for us to tell each other things, so anything that I'm in a position to know, I'm also in a position to tell you about, and vice versa. (At least, to a pretty close approximation, ignoring limitations of vocabulary, etc.) So for each fact F that we might consider, it's extremely likely that F is within my epistemic reach iff it's within yours, because, at least to a close approximation, being within my epistemic reach *suffices* for being within yours, and vice versa.

So it seems very plausible that, at least in the usual, garden-variety sort of case, the presupposition of relevant similarity that's needed in order for self-locating assertion to be a good idea is going to be in place. Given the similarity in what's within each of

our epistemic reaches that comes with being cooperative conversation partners, it's quite

plausible that if some proposition P is compatible with everything that's within your

epistemic reach, then it's compatible with everything that's within my epistemic reach,

and vice versa. And it's quite plausible that we will presuppose that this is so.

According to the relativist theory sketched above, what we're up to when we're

making assertions using epistemic modals is characterizing our position relative to the

evidence. It's useful for the various parties to a conversation to have a way of

communicating this sort of self-locating information about their shared epistemic

situation, in the same way that it's useful to have a way of communicating self-locating

information about geographical locations.[17] This account seems to explain nicely the

ways that people in fact use, and respond to, epistemic modal claims.


*Conclusion*

We have a motive, based on the appropriateness of attributions of truth and falsity

in cases like *Eavesdroppers*, to give a relativist account of epistemic modal claims. The

relativist, though, then owes us an account of the role of these kinds of sentences, and the

peculiar truth conditions he associates with them, in communication. There is, as it turns

out, a communicative niche for the kinds of self-locating contents that the relativist

assigns to sentences like "Bond might be in Zurich" – it is often a good thing to be able to

share information about our collective situation that is not just information about which

world we all inhabit.

---

[17] It's worth noticing that this really is (often) useful information to exchange. Since it can happen that there are bits of evidence that we're not aware of, but are still within our epistemic reach (particularly when we're engaged in conversations with people who know different things than we do), we're not always going to have very good access to the facts about our evidential situation.

This account of the role of self-locating contents in communication is interesting because it provides some cash value – some fairly substantive, non-terminological consequences – for the claim of relativism. If a relativist semantics is right in some area, that has consequences for what we're up to when we make assertions in that area. We can test our relativist semantics by seeing whether it provides a *plausible* account of what we're up to. One such test is the following: If you're going to propose a relativist semantics that attributes self-locating contents to some bunch of declarative sentences that are commonly used without disastrous miscommunication, you need to make it plausible that the presupposition of relevant similarity is typically in place when those sentences are used to make assertions. The relativist semantics for epistemic 'might' passes this test.

References:

DeRose, K. (1991) Epistemic Possibilities, *Philosophical Review* **100**:581-605.
Egan, A., Hawthorne, J. and Weatherson, B. (forthcoming) Epistemic Modals in Context, *Contextualism in Philosophy*, Preyer and Peters, eds., Oxford University Press, Oxford.
Hacking, I. (1967) Possibility, *Philosophical Review* **76**:143-68.
Lasersohn, P. (MS) Context Dependence, Disagreement, and Predicates of Personal Taste.
Lewis, D. (1979) Attitudes *De Dicto* and *De Se*, *Philosophical Review* **88**:513-43.
MacFarlane, J. (forthcoming a) Epistemic Modalities and Relative Truth
MacFarlane, J. (forthcoming b) The Assessment Sensitivity of Knowledge Attributions, forthcoming in *Oxford Studies in Epistemology*, Oxford University Press.
Perry, J. (1979) The Problem of the Essential Indexical, *Nous* **13**:3-21.
Stalnaker, R. (1978) Assertion, *Syntax and Semantics* **9**: 315-332.