

Assertion, Context, and Epistemic Accessibility

John Hawthorne and Ofra Magidor

Introduction

In this paper we discuss a problem for the meta-semantic framework proposed by Robert Stalnaker. In his seminal paper ‘Assertion’ (Stalnaker 1978), Stalnaker distinguishes between the semantic content of a sentence on an occasion of use and the content asserted by an utterance of that sentence on that occasion. In general, the default option for assertoric content is the semantic content of the utterance, but as Stalnaker sees it, the mechanisms of conversation sometimes force the two apart. In this connection, he outlines some principles for asserted content that systematize cases of divergence. Central to Stalnaker’s framework is the notion of a context of communication, which is roughly the set of live options at each point in a conversation, and a conception of the role of assertion as that of cutting down the live options and thus narrowing the context. Of special interest in this connection is one of the principles governing assertoric content in the framework, one according to which the asserted content ought to be identical at each world in the context set (as we call it, the ‘Uniformity’ principle). Indeed it is Uniformity that explains, for many cases, why assertoric content comes apart from semantic content.

In this paper we challenge the plausibility of this principle by reflecting on its interaction with the facts of epistemic accessibility. In particular, we argue that the fact that agents can know propositions without knowing that they know them and fail to know propositions without knowing that they fail to know them causes serious problems for that principle and thus, more generally, for Stalnaker’s framework. In section 1, we briefly introduce Stalnaker’s framework. In section 2, we explain why failures of epistemic transparency cause a serious problem to the principle of Uniformity. In section 3, we elaborate on the notion of a diagonal content, and suggest that in the light of this discussion, Stalnaker might plausibly revert to a weaker version of the uniformity principle – the principle we call ‘Weak Uniformity’. However, we show that our argument in section 2 can be generalised to an argument against Weak Uniformity. In section 4, we show how the issues in question are particularly pressing in the light of the

presence of vague expressions. In section 5, we discuss various responses that a proponent of Stalnaker's framework can suggest to our arguments, and claim that they will not do. In section 6, we begin by noting our arguments so far have relied on a knowledge-theoretic construal of the context set (roughly, the worlds in the context set are those compatible with the participants' knowledge). We thus consider whether the arguments can be defused by moving from a knowledge-theoretic to a presupposition-theoretic notion of context (roughly, one according to which the worlds in the context set are those compatible with the participants' presuppositions). We argue that such a move will not help because presuppositions also suffer from transparency failures, and thus our arguments can be generalised against the presupposition-theoretic framework as well. This concludes the core of the paper, and our central case against Stalnaker's framework. In a final section, section 7, we offer some cursory remarks regarding the underlying rationale for the uniformity principle and argue that it is unpersuasive.

1. Context and assertion

We begin by outlining the essentials of the framework. According to Stalnaker, conversation takes place against a set of possible worlds called 'the context set'. Intuitively, the context set at a time represents those worlds that are considered live options in the conversation at that time. The role of assertion on this picture is to update the context set by ruling out from the set all those worlds that are incompatible with the asserted content. The content of an assertion is treated as a set of worlds, and update proceeds by intersection of asserted content (which, on Stalnaker's view, is a set of possible worlds) with the context set, thus producing a new context set.

Standardly, when a sentence is uttered, its asserted content is its semantic content – which is in turn determined by the standard compositional mechanisms of the language. Indeed, the default asserted content for an utterance is its semantic content:

Default: Unless the principles of rational communication are violated, the asserted content is the semantic content.

However, Stalnaker notices that there are settings where, if the asserted content were always the semantic content, assertions that are in fact intuitively felicitous would be unacceptable. One such case is where the semantic content is true across the context set or false across the context set (the simplest cases being ones where the semantic content is either necessarily true or necessarily false *simpliciter*). If semantic content were asserted content then it would be pointless to make assertions using sentences with necessarily true contents (because the assertion would have no impact on the context set). Meanwhile assertions with necessarily false semantic contents would be self-defeating (because they would remove all worlds from the context set and make further communication impossible)

Stalnaker thinks that asserted content may come apart from semantic content even in cases where the semantic content is contingent across the context set. This happens when the semantic content of an utterance is different at different worlds in the context set. Stalnaker's basic worry is that in such cases, if semantic content were asserted content, an audience would not know how to update the context set in the light of the assertion. Suppose that at some of the live options the semantic content rules out a certain world w but at others it does not. Since all options are live, each of the semantic contents is live, and hence the participants in the conversation would not know whether to keep world w in the context set or remove it: they would not be able to discern which worlds were being knocked out by the assertion. Take a simple example. Suppose I know there is only one person in a certain room, but I don't know whether it is Bill or Ben. Someone points to whoever is in the room and asserts 'He is on fire'. Assuming semantic orthodoxy, there is at least one world in the context set where the semantic content is the proposition that Ben is on fire, and at least one world where it is the proposition that Bill is on fire. Prior to the assertion, there are four relevant classes of live options – ones where Bill is in the room and on fire, ones where Bill is in the room and not on fire, ones where Ben is in the room and on fire, and ones where Ben is in the room and not on fire. If asserted content is semantic content, then three of these classes are ruled out, but while I know two of these three classes (those where the person in the room – be it Bill or Ben - is not on fire), I don't know which is the third class (the one where Bill is in the room and on fire, or the

one where Ben is in the room and on fire). Hence if asserted content is semantic content, the audience will be unable to appropriately update the context set and so the conversational setting will be defective.

To solve this problem, Stalnaker introduces a mechanism for selecting an alternative asserted content where simply going with the semantic content will result in a defect of this sort. The mechanism is via an operation that Stalnaker calls ‘diagonalization’. The diagonal content of an utterance is a proposition that is true at a world of the context set if the semantic content of the utterance at that world is true and false if the semantic content is false. The case of Bill and Ben is a paradigm case where asserted content is diagonal content rather than semantic content. In that scenario, the semantic content of ‘He is on fire’ is the proposition that Ben is on fire at each world in the context set that Ben is in the room and the proposition that Bill is on fire at each world in the context set that Bill is in the room. The diagonal proposition thus delivers the value ‘true’ at each world in the context set where the person in the room is on fire and ‘false’ at the remaining worlds in the context set. It is thus clear that what the audience ought to do, on this model, is to remove from the context set all and only the worlds where the person in the room is not on fire.

The general principle underlying the move to diagonalization in such cases is the following principle, which we call ‘Uniformity’:

Uniformity: In cases of rational communication, the same proposition is asserted at each world in the context set.¹

2. Epistemic accessibility and Uniformity failures

One subtle issue for the Stalnaker framework concerns the determination of the context set. What makes a world a live option in the relevant sense? In this section we shall operate with an account of context determination according to which a world is live in a

¹ Stalnaker 1978, p. 88. This is his third principle, modified to make explicit his claim that his principles are an account of rational communication rather than an empirical generalization. We have replaced ‘asserted’ with ‘expressed’ to make the content of the principle maximally clear.

conversation if it compatible with those propositions that are known by all parties to the conversation. Having raised a problem for the framework with context understood in this way, we shall later discuss whether some other conceptions of context will evade the problem (including iterated knowledge and presupposition based proposals).

The crux of the problem is this. Stalnaker holds that Uniformity governs rational communication. In cases where the semantic content will not respect Uniformity, Stalnaker reverts to the diagonal content. However, this implicitly assumes that moving to the diagonal content will ensure that Uniformity is satisfied. But, as we now argue, this assumption is unwarranted.

Our argument will rely on failures of transparency for knowledge. That is, it will rely on cases where one knows p , but one does not know that one knows p , and on cases where one does not know p , but one does not know that one does not know p .² Call a world w_2 ‘accessible relative to conversation C ’ from a world w_1 if and only if what is known by all participants in C at w_1 is compatible with w_2 . If knowledge is not transparent, there could be cases where accessibility relations of this sort are asymmetric, and also cases where they are intransitive.^{3, 4} Failures of transitivity make for cases like this: For some

² The claim that one sometimes does not know without knowing that one does not know is uncontroversial. (For example, in standard cases of illusion, one thinks one knows some proposition p that one does not in fact know). As for the claim that one sometimes knows without knowing that one knows: the possibility of such cases has been argued for extensively in Williamson 2000. The core line of thought is that one cannot know that exact limits of what one knows because one knows in general that the cases just beyond the limits of what one knows have to match in truth value with the proposition that one knows. If one knew the limits of what one knows, they would not then, after all, be the limits of what one knows since one could use those general considerations about the limits to push them further. We will return to Williamson’s argument in more detail in section 6.

³ We note in passing that even if (per impossible) knowledge were transparent, then as we set things up, it is fairly straightforward to construct cases of transitivity failure. Suppose A knows p and knows that he knows p , and that B knows p and knows that she knows p . But suppose that A does not know that B knows p and B does not know that A knows p . In this setting there will be accessible worlds where A does not know p (since that is not a proposition that both conversational participants know) and accessible worlds where B does not know p (for the same reason). From those worlds certain worlds where not- P holds will be accessible (since p is not known by both participants at those worlds). One way of trying to handle this is by crafting a conception of the context set in terms of mutual iterated knowledge (more on this later). Another is to treat such contexts as defective on account of a mismatch between what the participants know that relevant to the discussion. (Relevant here is Stalnaker 1978, p. 85).

⁴ More accurately, failures of the principle that says that if one knows p then one knows that one knows p ensure that the accessibility relation is not transitive. Failures of the principle that if one does not know p then one knows that one does not know p entail that the accessibility relation is not Euclidean, i.e. that there

conversation C, there is a world w_1 , such that the context set of C in w_1 rules out some world w_3 . However, there is a world w_2 in the context set of C in w_1 , such that, according to w_2 , C does include w_3 in its context set. Suppose q is incompatible with what is known at w_1 but is not incompatible with anything that is known to be known at w_1 . There will then be a q -world w_3 that is inaccessible from w_1 but is accessible from a world w_2 that is accessible from w_1 . In this scenario, w_3 will be in the context set of C at w_2 but not in that of w_1 .

Failures of symmetry make for cases like this: There is a world w_1 such that a world w_2 is part of the context set of w_1 , but such that from the perspective of w_2 , w_1 is not part of the context set. In such a case, w_2 will be epistemically accessible from w_1 (relative to the conversation) but w_1 will not be accessible from w_2 .

Let us see how these phenomena make trouble for Uniformity. Return to our example of Bill and Ben. First, let us exploit symmetry violations. Suppose we don't know whether it is Bill or Ben in the room, but for all we know, we know that it is Bill in the room. Then there is an epistemically accessible world w_2 , i.e., a world in our context set, such that in all worlds in the context set of w_2 it is Bill in the room. Suppose a participant in the conversation now says 'He is on fire', referring to the person in the room. For reasons given above, the asserted content is not the semantic content but the diagonal content: Uniformity would be violated by asserting the semantic content in this case. Consider now the asserted content of the same utterance at w_2 . As opposed the actual world, relative to w_2 , the semantic content of the utterance is uniform across all world in its context set: It is always the proposition that Bill is on fire. Given Default, this means that the asserted content at w_2 is simply the proposition that Bill is on fire. But this in turn entails that the asserted content at the actual world is different from the asserted content at w_2 . Thus we violate Uniformity even by asserting the Diagonal proposition since at

are three words w_1 , w_2 , and w_3 , such that w_2 and w_3 are both accessible from w_1 , but such that one of w_2 and w_3 is not accessible from the other. Given the reflexivity of knowledge, symmetry and transitivity failures will each entail a Euclidean failure, but not necessarily vice versa. It's worth noting, however, that any kind of Euclidean failure has the structure that we need for our argument: namely, that there are three worlds w_1 , w_2 , and w_3 , such that w_2 and w_3 are both in the context set of w_1 but one of w_2 and w_3 is not in the context set of the other.

other worlds in the context set, Default will not be trumped and hence the Semantic content of the utterance at that world will be asserted. It thus turns out that whether we assert the semantic content or the diagonal content we will violate Uniformity.

Next, let us exploit transitivity violations. Suppose we know it is Bill in the room but do not know that we know. There is some epistemically accessible world w_2 -- i.e. a world in our context set -- where it is unknown whether Bill or Ben is in the room. Someone actually utters 'He is on fire'. Given that all worlds in the context set are such that Bill is in the room, the semantic content will be uniform across the context set and hence, given that there are no other salient countervailing considerations to Default, the asserted content will be the semantic content. Consider now the asserted content of the utterance at w_2 . The context set at w_2 includes some worlds where the semantic content of the utterance concerns Bill and also some worlds where the semantic content of the utterance concerns Ben. Hence, by Stalnaker's proposed mechanism, the asserted content at w_2 will be the diagonal content. But now, once again, Uniformity is violated for the asserted content in the actual world is the proposition that Bill is on fire, while at certain worlds in the context set, the asserted content is a different, Diagonal, proposition.

3. Diagonal propositions and Weak Uniformity

In our initial presentation, we followed Stalnaker in talking of 'the Diagonal proposition' of an utterance. However, this is strictly speaking misleading. As we have laid things out thus far, we have only specified how diagonal content behaves with respect to worlds within the context set: at each of those worlds, the diagonal content is true if the semantic content expressed at that world is true and false otherwise. But this specification is completely silent as to how the so-called diagonal is to behave outside the context set. Thus, strictly speaking, we should talk about *a* diagonal proposition, not *the* diagonal proposition. A diagonal proposition is one that conforms to the diagonal recipe for the worlds in the context set. Now Stalnaker's own attitude seems to be that since it never matters what the truth value is outside the context set, it never matters which diagonal proposition one is asserting:

Don't Care: In cases of rational communication where the asserted content is not the semantic content, it is unimportant which diagonal proposition one asserts.⁵

This relaxed attitude shows that Stalnaker should not be fully committed to Uniformity but rather to a weaker version of it:⁶

Weak Uniformity: In cases of rational communication, the propositions asserted at each world in the context set agree in truth-value with respect to all worlds in the context set.

But the structural features of epistemic accessibility that we have been working with render this combination unstable. In connection with symmetry violations, take a version of the first case described in section 2, where as a matter of fact Ben is in the room. Then there will be a world in the context set where the proposition asserted – that Bill is in the room – is false at the actual world, but where the proposition actually asserted – a diagonal content – is true at the actual world. Here, then, there is a clash between Default and Weak Uniformity.

In connection with transitivity violations, consider a case where I know either Bill or Ben is in the room, but I don't know which and further, that for all I know that I know, I know it is Ben or Tim that is in the room. I cannot assert the semantic content at the actual world, for familiar reasons. Stalnaker recommends that I assert a diagonal proposition but reassures me that it doesn't matter which. Suppose then I actually assert the proposition that Bill or Ben is on fire by 'He is on fire'. There is world w_2 in my context set where the context set at w_2 includes worlds where either Ben or Tim is in the room, and where for familiar reasons the asserted content of my utterance relative to w_2 cannot be the semantic content. Stalnaker recommends that at w_2 the assertor expresses a diagonal proposition but that it doesn't matter which. Suppose the proposition asserted at w_2 is that Ben or Tim is on fire. Finally let us suppose that in the actual world, Bill is in the room

⁵ Note that where the semantic content is the asserted content, the semantic content is a diagonal proposition. But Default tells us that in this case it does matter which proposition one asserts.

⁶ Stalnaker's preference for Weak Uniformity is confirmed by a footnote in 'Assertion' (see Stalnaker 1978, p. 91, f.n. 13) and is reiterated in Stalnaker 2006, p. 297.

and on fire. Now Weak Uniformity is violated, since the proposition asserted in the actual world is true at the actual world but the proposition asserted at the Ben-Tim world w_2 is false at the actual world.

4. Vagueness

The reflections of the previous sections have an interesting application to cases where one makes an assertion containing a vague expression. To see how Stalnaker's framework interacts with vague utterances, we first need to fix on a theory of the semantic content for vague predicates. We shall consider the two most prominent theories – epistemicism and supervaluationism – and show that the framework runs into trouble for reasons tightly connected to the structural considerations discussed above.

Let us first consider a standard epistemicist story. The actual semantic content of a vague utterance such as 'Bill is tall' is a content that satisfies both excluded middle and bivalence. In a borderline case, we do not know whether the content is true, and this ignorance has a distinctive source: at close worlds, imperceptibly different from the actual world, the utterance expresses a slightly different content that disagrees in truth value with regard to the case at hand. (In non-borderline cases there is this variation in content across close worlds as well, but the contents need not disagree in truth-value).

As a toy model let us assume that in each world, if the sharp cut-off point for 'tall' is N then in all epistemically accessible worlds the cut-off is either N or N plus one cm or N minus one cm. A model of this sort explains unknowability in borderline cases and generates a margin for error principle for vague predicates.⁷ Suppose that the (unknown) sharp cut-off for 'tall' in the actual world is 1.8m. Given our model, this means that in epistemically possible worlds, the cut-off point for 'tall' has one of the three values: 1.79m, 1.8m, and 1.81m. Suppose the context set prior to my assertion is one where Bill's height is unknown. I discover he is 1.82m and exclaim 'Bill is tall'. If the content of my assertion is the semantic content then Weak Uniformity is violated. So, by Stalnaker's lights, I assert a diagonal proposition here. That is to say, I assert a

⁷ See Williamson 1994, particularly ch. 8.

proposition that it true at a world in the context set, if and only if the semantic content expressed at that world is true at that world. Thus at the worlds in which the cut-off for ‘tall’ is 1.79, the diagonal proposition will be true if and only if Bill is at least 1.79m in height, in the worlds in which the cut-off for ‘tall’ is 1.8m, the diagonal proposition will be true if and only if Bill is at least 1.8m in height, and so forth. Of course, as usual, there are many diagonal propositions - differing in their truth-values over worlds which are not in the context set – but given Don’t Care, it doesn’t matter which of those diagonal propositions I assert.

Now consider a world in the context set where the sharp cut-off for ‘tall’ is 1.79m. Given our model, epistemically accessible worlds from that world will have a profile different from the actual case: they will include worlds where the cut-off is restricted to the values 1.78m, 1.79m, and 1.8m. At such a world, for familiar reasons, asserted content will not be semantic content, but rather a diagonal content. However given that the set of worlds in the context set at that world do not include all of the worlds in the actual context set, the worlds that fall within the ambit of Don’t Care for inhabitants of that world will include some worlds (namely those in which the cut-off is 1.81m) that fall within the actual context set. It is now an easy matter to construct a diagonal proposition for members of that world that differs from the diagonal proposition actually asserted. Consider for example the following kind of diagonal proposition which we can assert at the 1.79m cut-off world. With respect the worlds in the context set of that world (i.e. the 1.78m, 1.79m, and 1.80m cut-off worlds) this diagonal is defined by the usual recipe. But given Don’t Care, we can define the behaviour of the diagonal proposition in the worlds outside the context set any way we like. In particular, since 1.81 cut-off worlds are outside of that context set, we can define our diagonal proposition so that it is true in those worlds if and only if Bill is above 1.9m. The diagonal thus defined is false in the 1.81 cut-off worlds (because Bill is 1.82m tall, which is less than 1.9m), but recall that our *actual* assertion of ‘Bill is tall’ is true at the 1.81m cut-off worlds. The upshot is that the proposition asserted in the actual world, and that asserted in the 1.79m cut-off world disagree in truth-value with respect to some worlds in the context set (the 1.81 cut-off worlds), and Weak Uniformity will be violated.

A similar story will apply to supervaluationism. Without going into details, that framework posits a range of acceptable precisifications for a predicate at a world, without privileging any one of them semantically. In this setting, let us think of the meaning of a predicate as given by its range of acceptable precisifications. Our use somehow determines a range of acceptable precisifications. It has to be allowed, course, that small, imperceptible differences in use sometimes makes for small differences in the range of acceptable precisifications (After all, one can construct a series of cases, with each adjacent case imperceptibly different but where the meanings of the cases at each end are wildly different. So small differences have to make for a difference in meaning). The upshot is that, within the context set, the meaning of a vague predicate will not be constant. Moreover, there will be worlds that are just far enough away from the actual world to be outside the context set but close enough to some worlds within the context set to be within the context set of inhabitants at that world. With this structure in place, the preceding considerations apply: Worlds that we care about in the actual world will count as don't cares relative to other worlds in the context set and hence Weak Uniformity will fail.

5. Variations on the knowledge-theoretic framework

In this section and the next, we will briefly consider various ways in which the proponent of a broadly Stalnakerian framework might try to respond to our argument. We assume, as before, that it is hopeless to defend the transparency of knowledge.

One response we have encountered is to use iterated knowledge to define the context set. Let us say that we know* p if and only if we know p and we know that we know p and so on. Let a world belong to the context set if and only if it is not incompatible with what we know*. Using know* to define the context set is a brute force method for responding to the cases discussed earlier that turn on transitivity failure.

This response is unsatisfying for a variety of reasons. First, if KK fails, there may be hardly anything that we know*, since each iteration brings fresh epistemic demands with it. Defining the context set by know* thus risks making the context set enormous,

extending far beyond what intuitively count as the ‘live options’ in a conversation. Second, and relatedly, it is not plausible that when someone asserts something we come to know* the proposition that they assert. Even if we come to know that content, know* is (assuming the failure of KK) so demanding that we may well fail to know* what is asserted. If the context set is defined by knowing* then assertion will rarely do the job that the Stalnakerian asks of it, namely that of narrowing the context set. Third, while the move to knowing* speaks to the original cases that turn on failures of transitivity, it does not address cases that turn on failures of symmetry.⁸

In response, the Stalnakerian might defend himself by conceding that we hardly ever know* anything but maintain that his description of rational communication is a permissible idealization. He might remind us that, after all, idealizations are quite common in decision theory and game theory (for example, theorizing in that setting often assumes logical omniscience), and presumably that is often quite justifiable for the purposes at hand. Can’t an invocation of knowledge* be justified in a similar way?

A blanket appeal to the acceptability of idealization is not satisfying here. Idealizations are most saliently permissible where real world rational agents approximate the situation described in the idealization in a way that makes the decision theory a good guide to how real world rational agents ought to proceed. But it is not clear that the Stalnakerian framework is useful to real world communicants in this way. The central line of thought demanding uniformity – one that claimed that hearers would not know how to proceed if uniformity were violated – would not be assuaged if speakers only approximated

⁸ There are other ways of trying to use iterated knowledge to solve the problem. But they all fail for reasons that should, by now, be unsurprising. Suppose we took the context set to be given by what we know we know and defined the context set by what is compatible with that second level knowledge. This helps with one problem with know*: while we hardly know* anything, there is arguably quite a bit that participants in a conversation know that they know. But it doesn’t help with the problem of symmetry failures. Moreover, and importantly, the main arguments of the paper will apply straightforwardly to cases where second level knowledge is not transparent. Suppose instead we took second level knowledge to be crucial, but defined the context set not by what is compatible with second level knowledge but instead by stipulating that the context set for a world w (relative to a conversation and time) are those worlds where the second level knowledge is exactly the same as in w . This has the drawback of detaching the notion of context set from the intuitive notion of what is a live option. For in those cases where there is second level knowledge but it is not known whether or not it is present, it will be a live option, intuitively, that there is no such second level knowledge.

uniformity. More generally, Stalnakerian arguments that divorce asserted content from semantic content assume strict compliance with uniformity (at least on its weaker version), not merely approximate compliance.⁹ Meanwhile that framework is utterly silent about what to do in cases where both semantic content and diagonal content violate uniformity. If the norm ‘approximate uniformity’ is to guide us in those cases, it needs further articulation and motivation.

A second possible response is to give up Don’t Care: in those cases where the rules governing rational conversation mean that the asserted content is a diagonal proposition, it is not just any diagonal proposition that we assert, but a particular one. The most plausible proposal of this sort recommends that asserted content in such cases ought to be the ‘superdiagonal proposition’, one that is true at a world if and only if that utterance expresses a true semantic content at that world and is false at a world if and only if the utterance expresses a false semantic content at that world.¹⁰ (Preference for superdiagonals is motivated by the need to ensure that the asserted content is uniform even where the facts of epistemic accessibility differ.) The proposal needs refinement. After all, there are worlds where the utterance does not exist. What does the so-called superdiagonal of the actual utterance have to say about those? Ignoring certain complications, let us allow a don’t care attitude about those worlds.¹¹

⁹ While we cannot enter at length into the relevant debates, it bears emphasis that even the idealizations of standard decision theories may in some cases be deeply problematic. If failures of transparency are a pervasive feature of our epistemic lives, normative guidance that is insensitive to such failures may in some cases turn out to be poor guidance. A decision theoretic framework that was altogether inapplicable to beings for whom transparency failed would not provide a good guide for epistemic updating. Given that an account of the dynamics of rational epistemic lives in the individual case ought not to assume transparency, wouldn’t it be rather strange to insist on it in the communal case?

¹⁰ Note that it will not help to replace Don’t Care with an injunction to always assert the diagonal that goes undefined at worlds outside the context set. For consider a case where a world w_2 is in my context set but outside the context set of a world w_1 in my context set. Then what I assert will be defined at w_2 , but what is asserted at w_1 will be undefined at w_2 . No further weakening of Weak Uniformity will preserve its original spirit here. As what I actually assert will be true in some cases, false in others, the situation described would, by Stalnaker’s lights, generate an ‘I don’t know what to do’ dilemma of exactly the type that his framework was trying to avoid.

¹¹ One such complication is if there are cases where a sentence is uttered but it is not known whether it is, or if a sentence is not uttered but this is not known. (see also Stalnaker 2006, p. 301 for a short note on this issue).

In response, note first that this does not help with all of our cases. Recall cases of symmetry violation where in the actual world we assert a diagonal but at an epistemically accessible world we assert the semantic content. In such cases, refined advice about which diagonal to assert will not restore Weak Uniformity. Second, insofar as we are permitted any intuitive fix at all on the asserted content, the proposal will lead to wildly unintuitive consequence. If I assert ‘He is on fire’ there is some possible world where the sounds I produce have the semantic content that there are three goldfish in the room.¹² According to the superdiagonal proposal, my asserted content has no more to do with fire than it has to do with goldfish. Sure enough, the proposal can explain why I get fire-theoretic information – namely by intersecting the assertion with the context set. Nevertheless, our intuitive sense that the assertion has itself something to do with fire rather than goldfish is blatantly violated.

Another option is to drop Default. But it should be clear that this move will not help with cases that turn on transitivity violation, since in these cases all the asserted contents were non-semantic, and thus no appeal was made to Default. In order to be responsive to the arguments, one would need to drop both Default and Don’t Care, instituting a general policy to superdiagonalize. This proposal would address both symmetry and transitivity failures. But it is still highly problematic. First, the concern that superdiagonal content is counterintuitive remains. But there is a second, perhaps deeper worry. Superdiagonal content is constructed out of semantic content. It thus assumes the reality of semantic content. But on the current proposal, semantic content is never asserted. It is surely rather troubling to insist on the reality of semantic content but detach it altogether from the facts about what is asserted.¹³

6. Replacing knowledge with presuppositions

¹² One might try to insist that while similar noises are made, it is not the same utterance. Note that it will not do here to insist that difference in semantic content makes for difference in utterance, since the whole framework relies on the epistemic accessibility of cases where the semantic content varies. Note also that insofar as one takes seriously the numerical identity and difference of utterances, one will be forced by Sorites-style considerations to allow that there are cases where some worlds in the context set contain a numerically different utterance to the actual utterance.

¹³ This objection to dropping Default is supported by Stalnaker. See for example Stalnaker 2006, p. 297.

One might think that the problems that we have been discussing can be solved by dropping our simplifying assumption that the context set is delivered by the facts about what is known. This is particularly pertinent, since Stalnaker himself explicitly distances himself from the thesis that a world belongs to a context set if and only if it is incompatible with what is known. He tells us that we regularly assume certain propositions for the purposes of a conversation even though those propositions are not known. In such cases, he wants to treat worlds incompatible with the assumptions as outside the context set even if they are compatible with what is known.

The core attitudinal state in Stalnaker's own presentation is that of presupposition. He does not define it in terms of more primitive ideology. Rather he takes it to be a fundamental mental state, constituted ultimately by speakers' dispositions. The central idea is that to presuppose something in a conversation is to take it for granted in that conversation.¹⁴ Can Stalnaker avoid our problems by defining the context set in terms of presupposition and insisting that while knowledge is not transparent presupposition is? If the transparency of a state is taken to require that one always knows whether or not one is in that state then one hardly avoids the problem. As Williamson indicates, the sorts of considerations that show that knowledge is not transparent in this sense show that no non-trivial condition is such that one is in a position to know one is in it whenever one is in it, and is such that one is in a position to know one is not in it whenever one is not in it. For his part Stalnaker assumes that presupposition is transparent in exactly this sense.¹⁵ (That Stalnaker confidently makes this assumption in some places is especially surprising given that in others he talks in language that strongly suggests that presupposition is not transparent. For example: "A person's presuppositions are the presuppositions that he

¹⁴ There are special cases where for the purposes of conversation we assume something we know to be false. As for the more standard cases, Stalnaker sometimes glosses our presuppositions as what each of us believe all of us believe (on his terminology, 'common belief'), sometimes as what each of us believes to be the common beliefs, sometimes as what each of us know that all of us know ('common knowledge'), sometimes what each of us believe is common knowledge (believed common knowledge). (See for example Stalnaker 1973. Note that this vacillation is very understandable since in the normal case our beliefs about what others believe is based in large part on what we believe others know.) Whichever is ultimately his preferred gloss, it is clear that it is supposed to line up with what intuitively count as the live options in a conversation. Once anti-luminosity arguments are conceded, none of these complex conditions are going to be transparent in the knowledge-theoretic sense. Moreover, as we shall see soon, it is implausible to think these states are transparent in the presupposition-theoretic sense.

¹⁵ This assumption is made explicit in Stalnaker 1998, pp. 99-100.

takes for granted, often unconsciously.... They are the background assumptions that may be used without being spoken – sometimes without being noticed.”¹⁶) Hence, pending some adequate response to the relevant arguments, an assumption of transparency is unwarranted, and the concerns of this paper for Stalnaker’s framework remain when the context set is defined in terms of presupposition rather than knowledge.

One might try to avoid the problem by trying to cash out the relevant notion of transparency in terms of second level presuppositions. If the context set is defined as those worlds compatible with our presuppositions in a context, then perhaps it is not important that the state of presupposition is transparent in the sense that its presence or absence be knowable, but rather that it be transparent in the sense that its presence or absence be presupposed. Granted it may be initially tempting to think that if I presuppose something, I presuppose that I presuppose it. But it is also initially tempting to think that if I know something then I know that I know it. As it turns out, both initially tempting thoughts cannot be true and for similar reasons.

We presented Williamson’s line of thought in informal terms earlier. Let us look at a more careful (and faithful) version of the argument. Suppose someone knows that a tree is less than 300 centimetres tall by exercising visual discrimination from a distance. Given known limitations in human vision, the following margin for error principle seems to be something that the person knows about himself in that situation:

(MEK) If the tree is N centimetres tall, the person is unable to know (from his current position, no matter how much reasoning he indulges in) that the tree is less than $N+1$ centimetres tall.

Suppose further that the KK principle (roughly, if one knows p then one is in a position to know that one knows p) were true and that knowledge can be always extended by deduction. Given KK, the person knows that he knows that the tree is less than 300 centimetres tall (we can assume that the person has reflected on the question of whether

¹⁶ Stalnaker 1973, p. 447.

he knows, and that all other such conditions that an application of the KK principle requires are satisfied in this case). He can then use the premise that he knows that the tree is less than 300 centimetres tall along with the margin for error principle in a deduction from known premises to the conclusion that the tree is less than 299 centimetres tall. Given KK he knows that he knows that the tree is less than 299 centimetres tall. Now he can repeat the procedure. We are led to the absurd conclusion that he can know that the tree has no height at all.

Let us see how a very similar style of argument applies to presupposition. Assume a PP principle:

PP: If one presupposes q , one presupposes that one presupposes that q .

We will also assume the following principles. First,

Closure: Presupposition is closed under entailment. If one presupposes each of $q_1 \dots q_n$, and q_1, \dots, q_n entail r then one presupposes r .

Note that closure is uncontroversial in the Stalnakerian framework: it simply follows from Stalnaker's representation of the contents of presuppositions as sets of possible worlds.

Second, we need a suitable margin for error principle for presuppositions. Consider a setting where serious empirical inquiry into a tree is being conducted from a perceptual distance, where it is clear that one's only available sources of information are human vision from a distance and reflection, and where one presupposes that one's visual faculties are working normally. In such a setting rational agents will not presuppose height propositions whose negation cannot be ruled out by the available visual information. Moreover, reflective rational agents will presuppose that that they will be rational in this way. Thus in the case where the tree is 300 centimetres tall and an agent's visual system is working normally (from a distance), a rational agent will not presuppose

that the tree is less than 301 centimetres tall, since it would be reckless to take such a proposition for granted given the visual information available. Moreover, a reflective rational agent realizes this and thus presupposes that if the tree is 300 centimetres tall, he will not presuppose that it is less than 301 centimetres tall. More generally:

MEP A reflective rational agent will presuppose that if the tree is N centimetres tall, he will not presuppose that it is less than $N+1$ centimetres tall.¹⁷

Now suppose a reflective rational agent in the kind of setting described above presupposes that a tree is less than 300 centimetres tall. By PP, that agent presupposes that she presupposes that the tree is less than 300 centimetres tall. By Closure and MEP, she presupposes that it is not the case that the tree is exactly 299 centimetres tall. By closure again, she presupposes that the tree is less than 299 centimetres tall. But now again the argument generalizes to an absurd conclusion: a rational agent is led to presuppose that the tree has no height at all.

Our argument assumed a principle according to which all reflective rational agents in the kind of setting described will conform to MEP. But actually one can make trouble for PP using a much weaker assumption. It seems perfectly obvious that at least *some* reflective rational agents in *some* cases of the sort described will presuppose that if the tree is N centimetres tall, they will not presuppose that it is less than $N + 1$. If PP is correct, any such agent will be led to the conclusion that the tree has no height at all (by the same reasoning as above).

We conclude that just as KK is unacceptable, it is equally unacceptable to endorse an analogous principle of iteration for presupposition. Given the failure of PP, our arguments cannot be blocked simply by shifting to a presupposition-theoretic definition of context.

¹⁷ As with MEK, the principle is not intended to hold for settings in which one has sources of information other than human vision from a distance plus reflection. Also, quite obviously, the principle is only intended to apply to settings in which one is conducting serious empirical inquiry. If one is telling a story about a tree rather than trying to ascertain its height, one may well think that the actual height of the tree is irrelevant to what one presupposes about it.

One further critical point is in order. Recall that there are two aspects of a condition's transparency – detecting its presence and detecting its absence. We have been focussing on the former, and argued that there is no defensible version of the thesis that when one presupposes something, the presence of that presupposition is taken for granted (i.e. PP). Now in comparison to PP, there is even less temptation to think that the absence of presupposition is always presupposed, that is, there is even less temptation to think that whenever someone doesn't take p for granted, they take for granted that they don't take p for granted. It bears emphasis that that even were one to cling to PP, this latter type of transparency failure will generate symmetry failures of the sort already discussed.

One might try to define 'presupposition' in a way that evades the arguments presented above. But we should emphasise that it is not a stipulated term but one that is supposed to line up with our intuitive idea of what we take for granted in a conversation. Once we realize that the boundaries of taking for granting are not transparent to us, we realize that rational speakers will not even form a belief either way around those boundaries. But then there is no recognizable sense of 'taking for granted' according to which we always take it for granted that we take p for granted whenever we take p for granted, and according to which we always take it for granted that we don't take p for granted whenever we don't take p for granted.

7. Reflection on Weak Uniformity

Our case against Stalnaker is now complete: the principle of uniformity (in any of its versions), so central to Stalnaker's framework, has proven unsustainable. This naturally raises the question as to how well-motivated the principle was in the first place. In this concluding section we query the cogency of Stalnaker's motivations and also indicate how those motivations connect to some broader issues in the foundations of semantics.

The rough and ready argument which is taken to motivate uniformity goes like this: "Suppose an assertor utters a sentence in a conversational context where, prior to that utterance some live possibilities entailed that the utterance has one semantic value, other

live possibilities entail that it has some other. If the assertoric content is identical to the semantic value, the interlocutor wouldn't know what to do with the assertion.”

The quick and dirty argument is far from immediately compelling. Suppose I don't know whether 'he' means Bill or Ben. I am told 'He is on fire'. The obvious response to the quick and dirty argument is that I do know what to do in face of the utterance: I should accept the semantic content of the assertion.

The objector might reply that if this is what one ought to do, then one would not know how to do it. But this reply is not particularly compelling either. Here is one toy model of what is going on: We do have a way of 'uploading' the semantic content of the sentence, no matter what world we are in. We simply accept the sentence 'He is on fire'. In doing this, we thereby accept the proposition that Bill is on fire in some worlds and the proposition that Ben is on fire in others. Of course the thinker will not thereby be able to select between certain descriptions of which thought he is thinking – but, on the account we are entertaining, there is nothing especially disturbing about this fact. Suppose I sequentially glimpse two people, knowing one of them to be Bill, and accept on independent grounds 'Bill is happy', I will not in this case be able to select between two descriptions of my thought ('Are you thinking of the person you glimpsed first or the person you glimpsed second that he is happy?'). But that is not especially disturbing.

We now turn to two possible challenges to the position sketched thus far. First, there are cases which it is natural to describe as being ones where I don't know who is being talked about. In such cases, an assertion seems infelicitous even if I have the ability to upload its content by accepting an anaphorically linked sentence. Thus suppose I cannot see into a room and you say out of the blue 'He is on fire'. Your assertion seems straightforwardly infelicitous. Doesn't our account predict felicity where there is infelicity?

Now one thing we should point out here is that Stalnaker's framework does not easily explain the infelicity of your assertion either. His framework does recommend that the assertor not assert the semantic content of the sentence in the case described. But it does

not prohibit a diagonal content being asserted. His framework, as we outlined it, does not explain why on the one hand it is felicitous to assert ‘He is on fire’ when someone comes into view and one does not know whether it is Bill or Ben, but infelicitous when there is no salient person in view.

As far as the simplified model is concerned, we do not deny that there are detailed felicity-condition generating mechanisms about which our account is silent. If I say ‘The person behind you is trying to get past you’ to you in a setting where you are previously unaware that there is someone behind you, that is far less grating than an utterance ‘He is trying to get past you’. I can ‘accommodate’ the former much easier than the latter. Such facts are not to be explained at all by anything like the Weak Uniformity principle, nor indeed by merely appealing to general principles of rational communication. They are to be explained by specific conventional details regarding certain kinds of constructions: use of a deictic expression requires that the referent be already salient to the audience, whereas use of a definite description does not. In sum, there is no general principle governing conversation approximated by Weak Uniformity, and cases of assertoric infelicity tied to ‘not knowing who or what is being talked about’ are to be explained by fine-grained principles governing referential devices, not general principles about rational communication.

We turn to a second challenge, one that turns on more foundational issues. It is important to Stalnaker that, in a given setting, we (i) represent someone’s cognitive state simply by which propositions he or she accepts or rejects and (ii) we think of propositions in a coarse grained way, as sets of possible worlds. He wants, as a theorist, to represent ignorance about whether Bill or Ben is in the room by a context set that includes some Bill-in-the-room worlds and some Ben-in-the-room worlds. Intuitively, that ignorance persists even after ‘He is on fire’ has been uttered in a context where Bill is in the room. So if we represent ignorance in the way prescribed, it seems that we should, as theorists, represent the post-assertion context set as including some Bill-in-the-room worlds and some Ben-in-the-room worlds. But if this is right, then we cannot represent the update as the result of intersecting the semantic content with the original context set – for that

would knock out all the Ben worlds. In sum, Stalnaker insists on coarse-grained contents and insists on using contents rather than, say, guise/content pairs as the means of representing intuitive facts about ignorance. This puts severe constraints on which contents can be ascribed to thinkers in a particular explanatory context.

We recognize that such a perspective might lead one to have reservations about our rough and ready toy account. Certainly, our aim in discussing the second challenge is largely to show how uniformity connects to broader issues and not to resolve the latter. But one conclusion of the preceding discussion does bear emphasis. We have argued that Stalnaker cannot respect all the constraints that he places on himself. True enough, those constraints flow from his desire to represent ignorance using coarse grained contents – he subtly appreciates certain constraints which would have to be satisfied for ignorance to be adequately represented within the confines of his approach. That the constraints that flow from those confines cannot jointly be satisfied strongly suggests that we need to break out of these confines, and move to a different meta-semantic framework.^{18, 19}

John Hawthorne
Magdalen College, Oxford
OX1 4AU, UK
john.hawthorne@philosophy.ox.ac.uk

Ofra Magidor
Balliol College, Oxford
OX1 3BJ, UK
ofra.magidor@philosophy.ox.ac.uk

References

Stalnaker, R., 1973: ‘Presuppositions’, *Journal of Philosophical Logic* 2, pp. 447-457.

¹⁸ Note that if one accepts that it is the semantic content asserted and absorbed in the Bill-Ben case then the use of diagonalization to explain the informativeness of sentences that true across the context set or false across the context set (e.g. necessary truths) will be thrown into question too. For assume the effect of ‘He is on fire’ (in the presence of Bill) is to shrink the context set to worlds where Bill is on fire. Now suppose someone says ‘He is Bill’ (demonstrating Bill with ‘He’). That is intuitively informative. But this cannot be explained by diagonalization, since if the context set has already shrunk in the way described, then even the diagonal proposition will have a constant truth value across the context set.

¹⁹ We are grateful for conversations and comments from Brian Ball, Robert Stalnaker, and Timothy Williamson.

Stalnaker, R., 1978: 'Assertion', in his *Context and Content*, Oxford University Press, 1999, pp. 78-95. Originally published *Syntax and Semantics* 9 (1978).

Stalnaker, R., 1998: 'On the representation of context', in his *Context and Content*, Oxford University Press, 1999, pp. 96-114. Originally published in *Journal of Logic, Language and Information* 7 (1998).

Stalnaker, R., 2006: 'Assertion revisited: on the interpretation of two-dimensional modal semantics', in Garcia-Caripintero, M. and Macia, J. (eds.), *Two Dimensional Semantics*, Oxford University Press.

Williamson, T., 1994: *Vagueness*, Routledge.

Williamson, T., 2002: *Knowledge and its Limits*, Oxford University Press, 2000.