Context, Content, and Epistemic Transparency

Mahrad Almotahari
Massachusetts Institute of Technology
mahrad@mit.edu

Ephraim Glick
Massachusetts Institute of Technology
ephraim@mit.edu

We motivate the idea that presupposition is a transparent attitude. We then explain why epistemic opacity is not a serious problem for Robert Stalnaker's theory of content and conversation. We conclude with critical remarks about John Hawthorne and Ofra Magidor's alternative theory.

1. Introduction

In 'Assertion, Context, and Epistemic Accessibility', John Hawthorne and Ofra Magidor present a challenge for Robert Stalnaker's theory of conversation as set out in his paper 'Assertion' (Stalnaker 1978). Their discussion takes advantage of recent trends in the theory of knowledge to show that a central thesis of Stalnaker's framework, which they call 'Uniformity', is problematic. The source of the problem, according to Hawthorne and Magidor, is the interaction between Uniformity and the fact — if it is a fact — that knowledge and presupposition are not transparent attitudes. Hawthorne and Magidor take the upshot of their discussion to be that Stalnaker's account of assertion cannot satisfy all of its own constraints. If they are right, then the framework of 'Assertion' is internally incoherent.

In this paper, we quickly present Stalnaker's theory of conversation, summarize Hawthorne and Magidor's worry in our own way, and argue that Stalnaker is able to address their worry without much difficulty. We will show that the case they take to be a counterexample to the method of diagonalization can be dealt with even on the assumption that presupposition is nontransparent, and that the 'toy model' they propose as an alternative to Stalnaker's framework does not explain how we interpret and understand utterances.

Mind, Vol. 119 . 476 . October 2010
doi:10.1093/mind/fzq087 Advance Access publication 17 January 2011

2. Reviewing the framework of 'Assertion'

The central notion in Stalnaker's theory of discourse is the notion of *context*. Contexts play two distinct roles. First, they supply the information relative to which one's audience interprets one's speech act. This role is probably most clear when we consider the interpretation of indexical expressions. To fix the semantic value of an indexical, we rely on information from the context in which the expression occurs. Second, contexts are the objects on which speech acts act. Participants in a conversation begin with a body of information which is, or which is thought to be, available to all, and it is this body of information that their speech acts are meant to affect. If a speaker's speech act is successful, then it will change the common stock of information that defines the context. Stalnaker formally represents the context in which a speech act occurs in terms of the set of possible worlds consistent with what the speaker and her audience mutually knows, believes, or assumes. Stalnaker calls this set of worlds the *context set*.

Of course, each speaker/auditor represents for herself what is mutually known, believed, or assumed. When one speaker/auditor's representation of the 'common ground' differs from that of her conversational partners, the context is said to be defective. Defective contexts, according to Stalnaker, are unstable; they tend to correct themselves. One mechanism that allows the context to reach an 'equilibrium state' is presupposition accommodation.

> One way for a defective context to reveal itself is for the speaker to say something that shows that she believes that it is (or will be) common belief that φ where the addressee does not believe that φ, even after recognizing that the speaker is presupposing it. For example, Alice says to Bob, who is holding his baby daughter, 'how old is he?' … Bob recognizes that Alice is taking something to be common belief that he knows to be false. How should he respond? … Bob might decide to ignore the matter, tacitly *accepting* what Alice is manifestly presupposing for the purpose of facilitating communication without disrupting the conversation with a distracting correction. That is, Bob accommodates, not by coming to *believe* the false presupposition that Alice is presupposing, but by accepting it as part of the common ground. (Stalnaker 2002, p. 717)

We raise this point in order to highlight an aspect of Stalnaker's framework which we will employ later in our discussion.

Assertions are meant to act on and update the context. Update occurs when participants in the conversation revise what they believe or assume. To formally represent how assertions update the context,

Stalnaker takes the content of an assertion to be the set of possible worlds in the context set at which the sentence one asserts is true. This conception of asserted content allows Stalnaker to explain how assertions revise the common stock of information, and thus update the context set, in terms of set-theoretic intersection. To derive the new context set, one simply intersects the old context set with the asserted content of the speaker's utterance. Although update often occurs when audience members accept an assertion, this is not the only way by which one can affect the context set. More on this point below.

Stalnaker distinguishes the semantic content of an utterance, which one can compositionally derive from the semantic values of the words that a speaker utters in a context, from its asserted content, which depends in part on principles of rational communication. In any case where the identification of semantic content and asserted content would violate a principle of rational communication, Stalnaker thinks the two kinds of content come apart. When they do, the audience members *diagonalize*; they identify the asserted content of the speaker's utterance with its *diagonal proposition*.[1] The diagonal proposition of (an utterance of) a sentence *s* is the proposition which is true at a world in the context set, *w*, if and only if the semantic content of *s* at *w* is true at *w*.

Discourse that involves the violation of a principle of rational communication is thought to be abnormal. The default case is, therefore, one where the semantic content of a sentence *is* its asserted content.

(Default) Unless the principles of rational communication are violated, the asserted content is the semantic content

Hawthorne and Magidor's worry depends on this point, so we borrow their formulation of the principle.

One kind of case in which the semantic content and the asserted content of an utterance come apart occurs when the semantic content of an utterance is different at different possible worlds in the context set. The worlds in the context set are worlds that, for all the participants in the conversation know or presuppose, *are* the actual world.

[1] According to Stalnaker (1978), utterances determine a *propositional concept*, which is a function from worlds in the context set to propositions, or, alternatively, a function from pairs of worlds to truth-values. One can represent a propositional concept as a two-dimensional matrix; the diagonal proposition is so called because it is depicted by the set of truth-assignments that run from the upper left-hand corner of the matrix to the lower right-hand corner.

So, if the asserted content of an utterance were its semantic content in cases where the semantic content of the utterance differs across worlds in the context set, audience members would not know how to update the context. Hence the need for re-interpretation.

> Take a simple example. Suppose I know there is only one person in a certain room, but I do not 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 do not 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). (Hawthorne and Magidor 2009, p. 379)

If, in the case above, the semantic content of the speaker's utterance — 'He is on fire' — were its asserted content, we would not know whether to remove from the context set worlds in which Bill is in the room and on fire, or worlds in which Ben is in the room and on fire. Diagonalization resolves the difficulty; audience members take the asserted content of 'He is on fire' to be its diagonal proposition, which maps a world in the context set on to the value True iff the person in the room at that world is on fire at that world. Thus one removes from the context set all and only the worlds at which the person in the room at that world is not on fire.

It should be clear from our brief summary of Stalnaker's theory that the notion of a *possible world* plays a key role in the formal implementation of the framework, but make no mistake: the substantive point that this technology models does not depend on a particular metaphysics of modality, nor does it depend on Stalnaker's conception of what propositions are. (One should distinguish the representational tools that a theory employs to model a particular phenomenon from the phenomenon itself.) The substantive point (the phenomenon that Stalnaker wants to model) is simply that sometimes conversational participants lack the kind of knowledge that is necessary for recovering the semantic content of an utterance, and yet they somehow manage to recover information from the speech act. One need not think that there are possible worlds, or take

propositions to be sets thereof, to acknowledge this point. John Perry, for example, has drawn attention to the very same phenomenon in a slightly different way.

> Ellsworth goes to Hawaii and sends me a postcard. Unfortunately, it gets a bit wet before I receive it. The postmark, return address, and signature are all illegible. The message stays dry: 'I am having a good time.'
>
> If I am a competent speaker of English, I will understand the meaning of the sentence written on the postcard and hence the truth conditions of the utterance that produced it. It is true, if the person who wrote the postcard was having a good time at the time he or she wrote it. …
>
> I want to emphasize that the postcard is only a somewhat dramatic example of a common phenomenon. Language is a tool for communication, and the artful speaker takes care not to rely on contextual factors that the intended listener will not be able to use. …
>
> The reason that such cases strike us as a bit odd, however common, is that our paradigm is the case of successful communication. The speaker wants the listener to believe a certain proposition; the skillful speaker does not rely on contextual items in expressing that proposition that the listener cannot use in grasping it. (Perry 1988, p. 197)

It is clear from the surrounding discussion in Perry's article that by 'the truth conditions of the utterance' he does not mean the semantic content of the utterance, but rather the proposition that the author of the message was having a good time when he wrote it. Perry intends to highlight the fact that one can recover that proposition even if one does not know who the author of the message was or when the message was written. Perry acknowledges the similarity between his idea and Stalnaker's in the 'Afterword' (Perry 1988, p. 205).

We are now in a position to clearly state the principle whose apparent violation, in the cases we have been considering, triggers diagonalization. This principle is meant to express the intuitive thought that Perry and Stalnaker share, that is, even in cases where the audience is semantically ignorant — where they do not know what the referent of 'he' or 'I' is — they can recover some information from the speaker's speech act. Although the principle is formulated in terms of the possible worlds theory of context and content, the phenomenon it is meant to explain is, we think, a pre-theoretical datum.

(Uniformity) In cases of rational communication, the same proposition is asserted at each world in the context set[2]

[2] Hawthorne and Magidor distinguish Uniformity from 'Weak Uniformity', which says that, in cases of rational communication, the propositions asserted at each world in the

In the case involving Bill and Ben, diagonalization brings the conversation back in line with Uniformity and thus allows update to proceed as usual. But for Stalnaker's strategy to work generally, diagonalization has to guarantee that the speaker's utterance respects Uniformity. In other words, Stalnaker assumes that if one takes the asserted content of an utterance to be its diagonal proposition, then Uniformity will not be violated. Hawthorne and Magidor argue that this assumption is false; they think that, in contexts where epistemic transparency fails, Stalnaker's repair strategy will not work.

3. Reconstructing the objection

Our reconstruction of Hawthorne and Magidor's argument will differ from their presentation in two ways. First, although Hawthorne and Magidor originally present their objection in terms of knowledge, we will present it in terms of the more general notion of presupposition, which in the context of Stalnaker's theory of discourse amounts to treating a proposition as true for the purpose of conversation. We do not think that this difference will unfairly distort Hawthorne and Magidor's worry, since they claim that their argument works either way. Additionally, when the objection is put in terms of presupposition it side-steps a particular reply that friends of Stalnaker's theory will no doubt want to give in response, specifically, that the objection rests on a mischaracterization of Stalnaker's notion of common ground.

Second, we will present Hawthorne and Magidor's argument in two stages. In the first stage we take them to argue that presupposition is not transparent. One can presuppose *p* and not presuppose that one presupposes *p*. Similarly, one can fail to presuppose *p* and not presuppose that one fails to presuppose *p*. In the second stage they argue that, in cases where one's higher-order presuppositions are not in line with one's first-order presuppositions, the method of diagonalization breaks down, because it cannot bring the conversation back in line with Uniformity.

We present their objection in two stages because we do not think that Hawthorne and Magidor have fully described a case

context set agree in truth value with respect to all worlds in the context set. Stalnaker originally had the weaker principle in mind, but we ignore the distinction here since, as we will argue shortly, an even weaker principle allows Stalnaker to accomplish his aims.

where transparency does in fact fail. Therefore, unless the first stage of their argument succeeds, we have no reason to think that there are problematic cases of the sort they rely on in the second stage of their argument.

Hawthorne and Magidor ask us to imagine a case in which the conversational participants presuppose that Bill is in the room even though they fail to presuppose that they presuppose Bill is in the room. They ask us to imagine a similar case in which the conversational participants fail to presuppose that Bill is in the room and, further, that they fail to presuppose that they do not presuppose he is. However, they do not fully describe in pre-theoretical, intuitive terms a case in which transparency fails. They simply ask us to take the above story about Bill and Ben and add the supposition that transparency fails.

We find it very difficult to imagine such cases in any detail. Why would reflective, rational audience members fail to have access to their first-order presuppositions? Normally, we take ourselves to have direct, reliable first-person access to our mental states. And furthermore, we are often able to draw conclusions about our own mental states simply by reflecting on the subject matter of those states. Suppose the audience members presuppose that Bill is in the room on the basis of reliable testimony. Since the audience members are reflective, they should be aware that they have evidence from testimony, and they should judge that testimony to be sufficient to guide their presuppositions. So why would they not take themselves to presuppose exactly what the evidence (which they are perfectly aware of) tells them to presuppose? On the other hand, suppose they fail to presuppose that Bill is in the room because they have no evidence at all that it is Bill rather than Ben. Why should they be unable to reflect on their first-order evidence as follows? 'I cannot see who is in the room. I have not been told who is in the room. I cannot think of any good reason to suppose that is Bill rather than Ben in the room, or vice versa. So rationality dictates that I should not presuppose either that it is Bill or Ben in the room.' Hawthorne and Magidor do not address these questions.

What Hawthorne and Magidor have done instead is offer an outline of a case where we are to suppose that the accessibility relation which defines the context set behaves as it would if the outline were filled in so as to imply the failure of transparency. The terms with which Hawthorne and Magidor describe their cases are too theoretical, and far too much is left unsaid for these cases to be thought of as

counterexamples. Hawthorne and Magidor's case, therefore, does not threaten the method of diagonalization in the same way that Gettier cases threaten the Platonic conception of knowledge: Gettier described cases in which it was intuitively clear that a subject does not have knowledge, but Hawthorne and Magidor have not described a case in which it is intuitively clear that a subject lacks the relevant higher-order knowledge or presupposition. The second stage of their objection to Stalnaker's theory critically depends, therefore, on the general considerations that they use to show that there must be cases in which transparency fails. We will briefly assess the merits of those considerations, though our central focus will be on the second stage of Hawthorne and Magidor's argument. The two stages can be assessed independently, and even if the first stage is questionable, it will be worthwhile to reflect on how non-transparency would interact with Stalnaker's framework.

4. Stage one

Assume for the sake of argument that presupposition is transparent.

(PP) If one presupposes q, one presupposes that one presupposes q

Next, assume that presupposition is closed under entailment.

(Closure) If one presupposes each of $q_1, \ldots, q_m$, and $q_1, \ldots, q_n$ entail r, then one presupposes r

(The closure principle is, as Hawthorne and Magidor note, an unattractive commitment of Stalnaker's theory of content. So even if one were sceptical of Closure, it is not open to Stalnaker to reject this premiss of the argument.) Finally, Hawthorne and Magidor ask us to

> 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 they will be rational in this way. (Hawthorne and Magidor 2009, p. 392)

Hawthorne and Magidor think that, in these conditions, it will be true that

(MEP) A reflective, rational agent will presuppose that if the tree is n cm tall, she will not presuppose that it is less than $n+1$ cm tall.

And from these assumptions it follows that, if a reflective, rational agent were to presuppose (in the setting Hawthorne and Magidor ask us to consider) that a tree is 300 cm tall, then she would be led to presuppose that the tree has no height at all, which is absurd. The source of the problem, according to Hawthorne and Magidor, is (PP). They then conclude that (PP) is false.

Why think (MEP) is true? Hawthorne and Magidor write that 'rational agents will not presuppose height propositions whose negation cannot be ruled out by the available visual evidence' (p. 392), but we see no reason to think this is true. What a rational agent presupposes not to be ruled out by the visual evidence may come apart from what the visual evidence in fact rules out. It is no part of being rational that one be able to avoid this kind of mistake; it is simply part of being good at tree-height estimation. There may be some other principle similar to but more attractive than (MEP), one which would allow Hawthorne and Magidor to derive the non-transparency of presupposition, but we cannot see what it is.[3] And, in any event, we will show that non-transparency does not undermine the core of Stalnaker's framework.

5. Stage two

We think Stalnaker can respond to the second stage of Hawthorne and Magidor's argument in two different ways. The first way involves a minor revision to the method of diagonalization, but the account we end up with is able to accomplish the same theoretical goals that originally motivated Stalnaker's framework. The second response does not involve any revisions at all; it merely draws on tools and techniques that are independently attractive. The upshot of either strategy is that cases where one's higher-order presuppositions are not in line with one's first-order presuppositions do not thwart the

[3] An alternative principle is the following: if x is a reflective, rational agent, then x will not presuppose that the tree is less than n cm tall for any n such that x presupposes the available visual evidence to be compatible with the tree's being n cm tall or greater. But it is not obvious whether this principle would suit Hawthorne and Magidor's purpose.

method of diagonalization. Before we set out our reply, we will summarize the second half of Hawthorne and Magidor's argument.

To begin, consider the brief story above concerning Bill and Ben. Suppose we do not presuppose whether Bill is in the room or whether Ben is. But for all we presuppose, we presuppose that Bill is in the room. (Similarly, for all we presuppose, we presuppose that Ben is in the room.) Then there is at least one world, $w^\star$, in the context set such that in all worlds in the relevant context set of $w^\star$ Bill is in the room. (There will also be at least one world in the context set such that in all worlds in the relevant context set of *that* world Ben is in the room.) Now suppose that someone points to whoever is in the room and says 'He is on fire'. Since the semantic content of the sentence the speaker utters differs across worlds in the actual context set, the asserted content must be the diagonal proposition. Uniformity would be violated otherwise. Now Hawthorne and Magidor ask us to consider what the asserted content of the speaker's utterance is at $w^\star$. Since Bill is in the room at every world in the context set of $w^\star$, the semantic content of 'He is on fire' does not vary across worlds in this context set; it is always uniform. Default tells us, then, that the asserted content of 'He is on fire' at $w^\star$ just is its semantic content. But this means that the asserted content of 'He is on fire' at the actual world — namely, the diagonal proposition that the person in the room is on fire — differs from the asserted content of 'He is on fire' at $w^\star$, which happens to be the proposition that Bill is on fire. Therefore, we cannot respect Uniformity by way of diagonalization. Whether we take the speaker to assert the semantic content of 'He is on fire' or its diagonal proposition, Uniformity will be violated.

Hawthorne and Magidor describe another kind of situation which would yield a similar result. Assume that the conversational participants presuppose that Bill is in the room, but do not presuppose that they presuppose he is. Then it will follow that all of the worlds in the actual context set are worlds where Bill is in the room. But there is at least one world, $w\dagger$, in the actual context set such that in some but not all worlds in the relevant context set of $w\dagger$ Bill is in the room. Someone points to whoever is in the room and utters the sentence 'He is on fire'. Since every world in the actual context set is one where Bill is in the room, the semantic content of the sentence the speaker utters is uniform across worlds in the actual context set. It follows, by Default, that the asserted content of the actual utterance is its semantic content, namely, the proposition that Bill is on fire. But the semantic content of 'He is on fire' at $w\dagger$ is not uniform across worlds in the

context set of *w*†. So Uniformity would be violated if conversational participants took the asserted content of 'He is on fire' at *w*† to be its semantic content. The participants thus diagonalize in an effort to respect Uniformity. But their doing so at *w*† means that Uniformity is violated in the actual world. The asserted content of the speaker's utterance at the actual world is its semantic content, which is the proposition that Bill is on fire, but the asserted content of her speech act in *w*†, which is a member of the actual context set, is the diagonal proposition that the person in the room is on fire. The upshot, then, seems to be that Stalnaker's theory can neither work with Uniformity nor without it.

The first response we want to explore admits that Uniformity is problematic, but maintains that the problem does not challenge the intuition that motivates it. The motivation for Uniformity, one will no doubt recall, is that, in normal cases, an utterance is a correct move in the conversation only if auditors have the knowledge necessary to determine the proposition being asserted. But sometimes auditors are ignorant of the semantic facts necessary to determine the semantic content of the sentence the speaker utters; they might not know, for example, what the referent of a particular pronoun is. And in such cases, auditors can still derive information from the utterance. Uniformity is meant to explain how this is possible. If one can specify a rule of conversation that respects the data and plays the sort of explanatory role that Uniformity plays in Stalnaker's original framework, but which does not give rise to the problems that Hawthorne and Magidor describe, then the objection will lose its force, because diagonalization will still be a device that brings the conversation back in line with principles of rational communication. The problem with Uniformity, according to Hawthorne and Magidor, is that in the cases where transparency fails it is violated whether or not we assume that audience members diagonalize. What we need, then, is a principle that directs auditors to diagonalize in just the cases that Uniformity does, but which is not violated in the problem cases when one assumes that auditors do identify the asserted content of the speaker's utterance with the diagonal proposition. We think the principle below is just such a rule.

(Recovery) In cases of rational communication, the asserted content of the speaker's utterance is recoverable

To see that Recovery plays the same explanatory role that Uniformity plays in Stalnaker's original framework, note that the very same

considerations motivate both principles. Consider situations like Hawthorne and Magidor's original Bill and Ben case, in which auditors do not know who is in the room, but they *know* that they do not know this. Uniformity explains how the auditors determine the asserted content of 'He is on fire' as follows: they know that they do not know the semantic content of the utterance, so they know that at different worlds in the context set, the utterance has a different meaning. Since they do not know which of these worlds they are actually in, they conclude, in line with Uniformity, that the asserted content of 'He is on fire' is not the semantic content, but the diagonal proposition. Now compare the role of Recovery in the same sort of case. Auditors know that their knowledge does not allow them to recover the semantic content of the utterance, but they take the conversation to be a rational exchange. So, since they cannot determine the semantic content of the speaker's utterance, the auditors conclude, in line with Recovery, that the asserted content is not the semantic content, but the diagonal proposition.

Situations in which audience members cannot recover the semantic content of an utterance, and in which they do not know how to update the context with the semantic content, are situations in which Uniformity would be violated if the asserted content were the semantic content. The reason why audience members cannot recover the semantic content of the speaker's utterance in such cases is that they are semantically ignorant; they do not know the semantic value of at least one term in the speaker's utterance. But one does not need to know what the semantic value of each term in the speaker's utterance is in order to know what the diagonal proposition is. So auditors recover the diagonal proposition rather than the semantic content. Similarly, situations in which auditors are semantically ignorant are situations in which, if the asserted content were the semantic content, Recovery would be violated. Auditors' knowledge would be insufficient to recover the semantic content of the utterance. If the asserted content were the diagonal proposition, however, Recovery would not be violated. This illustrates why apparent violations of Recovery have the same significance as apparent violations of Uniformity.

Hawthorne and Magidor think that the cases involving $w^\star$ and $w\dagger$ imply that whether the audience diagonalizes or not, Uniformity will be violated. We claim that these cases do not pose the same problem for the revised framework — the framework with Uniformity replaced by Recovery. Take the first case, in which audience members do not presuppose whether Bill or Ben is in the room, but also fail to

presuppose that they do not presuppose whether Bill or Ben is in the room. Thus, the audience members do not presuppose any particular person to be the referent of 'he' in 'He is on fire', and, hence, do not presuppose enough to determine what the semantic content of the speaker's utterance is. Therefore, the audience members cannot recover the semantic content of the speaker's utterance. Recovery predicts, then, that the asserted content of the speaker's utterance cannot be its semantic content. What *can* the asserted content be? Well, since audience members can determine what the diagonal proposition of the speaker's utterance is even if they do not know the referent of 'he', they should identify the asserted content with the diagonal proposition. This, of course, is just what Uniformity tells us. But, whereas Uniformity appears to be violated whether or not one assumes that audience members actually diagonalize, Recovery does not.

Recall that at every world in the relevant context set at $w^\star$, Bill is in the room. So, at $w^\star$, audience members are not semantically ignorant in the way they actually are; they can determine the semantic content of 'He is on fire'. Recovery and Default predict, then, that at $w^\star$ the asserted content of the speaker's utterance is the proposition that Bill is on fire. But the revised framework, unlike the original framework, does not require that the same proposition be asserted at every world in the actual context set. For even if audience members would recover the semantic content at $w^\star$, it does not follow that, at the actual world, the audience is unable to recover the diagonal proposition. In fact, they do know enough to determine the diagonal proposition of the speaker's utterance. They know that the diagonal is true iff the referent of 'he' is on fire. So even if the speaker asserts different propositions at different worlds in the context set, the audience can still keep the conversation in line with Recovery by treating the speaker as asserting the diagonal proposition rather than the semantic content. Therefore, the fact that the asserted content of the speaker's utterance 'He is on fire' differs at $w^\star$ from its asserted content at the actual world does not generate the problem that it apparently does for the original framework.

One can apply the same line of thought to the second case involving $w\dagger$: in the actual world, audience members presuppose that Bill is in the room, so they presuppose enough to determine the semantic content of 'He is on fire'. Of course, at $w\dagger$ the situation is different; the audience there does not presuppose whether Bill or Ben is in the room, so they have to diagonalize in order to respect Recovery. But since the framework minus Uniformity does not require that the very

same proposition be asserted in every world in the context set, the kind of incoherence that apparently arises within the original framework does not arise within the revised framework.

6. Repair without Recovery

We will now present our second response on behalf of Stalnaker. This response does not revise the original framework of 'Assertion'. Rather, it draws on a couple of Stalnaker's more recent observations concerning update and presupposition accommodation. We begin our discussion in this section by returning to Hawthorne and Magidor's objection.

One might worry that, in the contexts which Hawthorne and Magidor have in mind, the process of interpretation would not even get off the ground. This conclusion seems to follow naturally from Stalnaker's view, since the audience members do not know, in the relevant contexts, which worlds are members of the context set. To derive the asserted content of an utterance, one must first determine whether the utterance in question would express the same proposition in all worlds in the context set, and that is precisely what one cannot do if one does not know what the context set is. So in the first case, the audience cannot diagonalize, and in the second case they cannot know that the semantic content is the asserted content. This line of thought strongly suggests that, in the cases Hawthorne and Magidor have in mind, the audience members cannot interpret the speaker's utterance without some kind of repair strategy that allows them to first determine what the common ground is. We do not think that this conclusion is far-fetched; we will argue that a natural repair strategy of just this kind is available in such cases. The strategy we will describe merely extends some independently attractive presuppositions of Stalnaker's framework.

In our description of Stalnaker's theory, we said that update occurs when the audience accepts or rejects an assertion. We then said that this was not the only way update can occur. Participants in a conversation update the context set with facts about the environment that become manifest as the conversation evolves. Our first assumption is that facts of this kind are often available after an utterance has been made but before it has been accepted or rejected. 'The prior context that is relevant to the interpretation of a speech act is the context as it is changed by the fact that the speech act was made, but prior to the acceptance or rejection of the speech act' (Stalnaker 1998, p. 101).

It follows from this observation that the mere utterance of the sentence 'He is on fire' by the speaker conveys information that updates the context set. The audience can use this information to assist them in the process of interpretation.

Our second assumption is that diagonalization is not the only way to react to an apparent violation of Uniformity.

> As with the other principles [governing informative linguistic interaction], one may respond to apparent violations [of Uniformity] in different ways. One could take an apparent violation as evidence that the speaker's context set was smaller than it was thought to be, and eliminate possible worlds relative to which the utterance receives divergent interpretation. (Stalnaker 1978, p. 91)

This alternative way of responding to apparent violations of Uniformity will play a key role in our reply.

Now recall the worry that we began with: the context sets in Hawthorne and Magidor's problem cases will not allow interpretation to even get off the ground. This worry arose because conversational participants in such contexts will be unsure what the context set is, so they will not know what resources are available to interpret the speaker's utterance. Hence, they will be unable to determine what the speaker means to assert. Such a situation is one in which the kind of interpretive strategy that Stalnaker describes above would naturally be put to use. When members of the audience observe the speaker utter the sentence 'He is on fire' in the problem cases, they reason that the speaker must take the context set to be such as to allow the audience members to recover the utterance's asserted content and, hence, that the context set is smaller than it seemed. For if worlds relevantly like $w^\star$ or $w\dagger$ were members of the context set as the speaker took it to be, then the speaker would know or presuppose that her audience would be unable to interpret her utterance, and she would therefore not bother to utter the sentence.

On these grounds, the audience members then accommodate the speaker's presupposition and remove worlds, such as $w^\star$ and $w\dagger$, at which the context set diverges in such a way as to block interpretation. Therefore, accommodation guarantees that worlds at which the utterance would express divergent asserted contents, specifically $w^\star$ and $w\dagger$, are removed from the context set. Interpretation then proceeds as usual. Audience members determine whether to accept or reject the asserted content of the speaker's utterance 'He is on fire' relative to the context set as it is after the problematic worlds have been removed.

Compare the situation here with Stalnaker's case involving Alice and Bob. Prior to accommodation, reflective and cooperative audience members will think the context is defective. They will take the speaker to be assuming that her audience can interpret her utterance, but prior to accommodation, worlds relevantly like $w^\star$ or $w\dagger$ *are* part of the context set as the audience takes it to be. The appearance of mismatch between the speaker's representation of the context set and the audience members' representation motivates accommodation in just the way that it does in the case involving Alice and Bob. Cooperative audience members accommodate, and, as a result, the process of interpretation gets off the ground. Moreover, no violation of Uniformity can be derived in the way Hawthorne and Magidor suggest, because the worlds at which the utterance receives divergent interpretations have been removed from the context set.

One might worry that our use of accommodation is unconstrained.[4] Take a case of the sort involving $w^\star$. Why would audience members not eliminate all *but* the worlds relevantly like it? There is a simple answer to this worry. The $w^\star$-like worlds are of two kinds; one kind consists of worlds at which audience members presuppose Bill is in the room, the other kind consists of worlds at which the audience members presuppose Ben is in the room. So, if accommodation left all $w^\star$-like worlds in the actual context set, then there would be worlds at which the asserted content of the relevant utterance would be the proposition that Bill is in the room, and worlds at which the asserted content would be that Ben is in the room. Hence, interpretation would not get off the ground even after audience members accommodate. But the whole point of accommodation is to allow for interpretation. Additionally, it will not do to say that accommodation eliminates all but $w^\star$-like worlds where audience members presuppose that Bill is in the room — or, alternatively, all but $w^\star$-like worlds where audience members presuppose that Ben is in the room — for the choice between the two kinds of $w^\star$-like worlds would have to be completely arbitrary. The accommodation we are suggesting is not arbitrary at all; the suggestion is that speakers eliminate exactly those worlds whose presence in the context set would block interpretation. In sum, to apply the mechanism of accommodation in a way that guarantees

[4] Some authors argue that, in general, appealing to accommodation is far too unconstrained to play an explanatory role in accounts of presupposition and utterance felicity. These concerns are too large to address here, but we take it that the issue is orthogonal to the problem of non-transparency for Stalnaker's framework.

that interpretation can get off the ground, one must eliminate w^*-like worlds.

Similar reasoning explains why accommodation would eliminate worlds relevantly like $w\dagger$, rather than all but the $w\dagger$-like worlds. If the $w\dagger$-like worlds were left in the context set post accommodation, then audience members would still not be able to interpret the speaker's utterance.

In order to block our account of how update proceeds in their cases, Hawthorne and Magidor must argue that in at least some contexts where transparency fails, accommodation will not be available to cooperative audience members. And to pose a problem for Stalnaker's framework, they must also show that even in such contexts audience members will be able to successfully interpret the speaker's utterance. But it is far from clear that audience members will be able to do so without the interpretive strategy we characterize above.

One point bears emphasis. The present defence of Stalnaker simply appeals to a mechanism that is generally available and independently motivated, namely, accommodation. An utterance like 'I am going to pick up my sister at the airport' is conversationally awkward unless the audience presupposes that the speaker has a sister, but the audience need not do so prior to the speaker's utterance. When the audience becomes aware that the utterance would be awkward without the presupposition, they treat the speaker as an authority on whether she has a sister and simply adjust the context set in the necessary way.

Our response is merely an application of the general picture of discourse that Stalnaker defends in 'Assertion' and, more recently, in 'On the Representation of Context' (1998). In the more recent paper, Stalnaker responds to an objection from Hans Kamp. Kamp's worry was that Stalnaker's formal representation of contexts as sets of possible worlds breaks down when we consider anaphoric links across sentences. Stalnaker's reply to Kamp was that he ignored an important way in which context sets evolve, specifically, through the recognition of manifest facts about the sentences speakers have uttered. One can think of Hawthorne and Magidor's objection, and our response, in similar terms: Stalnaker's formal characterization of contexts may appear to run into problems when we consider how contexts interact with failures of epistemic transparency, but one would take this appearance seriously only if one were to ignore the same mechanisms that undermine Kamp's argument. We thus find it peculiar that

Hawthorne and Magidor cite the papers we rely on, and yet ignore one of their central points.

7. Assessing the alternative model

A natural question at this point is: How will Hawthorne and Magidor explain successful interpretation in the original story about Bill and Ben, given that they do not help themselves to a mechanism like diagonalization? Stalnaker, as we have said, appeals to Uniformity, but Hawthorne and Magidor are not convinced by the so-called 'quick and dirty' argument that motivates this principle. The argument they have in mind goes something like this: suppose a speaker utters a sentence in a context where, in some worlds compatible with what the audience members mutually presuppose, its semantic content is p and, in other worlds compatible with what the audience members mutually presuppose, its semantic content is q. If the asserted content of the utterance were identical to the semantic content of the sentence, then the audience members would not know how to update the context. In response, Hawthorne and Magidor write:

> Suppose I do not 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] 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. (Hawthorne and Magidor 2009, p. 394)

It is certainly true that members of the audience know that they should accept whatever is asserted, but if what is asserted is the semantic content of the sentence, then they only know *de dicto* that they should accept the semantic content. The semantic content of the sentence is not such that they know they should accept *it*. Hawthorne and Magidor say little about what it is to accept a sentence and thereby the proposition one would assert by uttering the sentence, but one would think that to accept a proposition by accepting a sentence requires one to understand the sentence as expressing that proposition. This is not

the notion of acceptance that Hawthorne and Magidor have in mind,[5] since an auditor cannot understand 'He is on fire' to express any particular proposition in the relevant case. But they do not see this as a problem. We, however, disagree. In so far as we have an intuitive grip on their notion of accepting a sentence, it seems to us too weak to play much of a role in an account of communication, understanding, and action.

Suppose a trustworthy friend utters a sentence in Farsi; you accept the sentence, but have no idea what it means. Obviously, you do not successfully interpret the utterance in such a case. You know it conveys a truth, but you do not know the truth it conveys. Or suppose the auditor in the Bill and Ben example loves Ben but hates Bill, vowing never to save him from a fire. If accepting the sentence 'He is on fire' amounts to accepting the proposition that Ben is on fire (since Ben is in fact the man in the room), then one would expect the auditor to come to the rescue. But since the auditor does not know whether Bill or Ben is on fire, she does not know whether she should come to the rescue. However, Hawthorne and Magidor's account predicts that she would come to the rescue, since accepting 'He is on fire' is enough to accept that Ben is on fire. At least, it makes this prediction if the notion of accepting a proposition is to play the role that understanding traditionally plays in an account of action. If it does not play that role, then Hawthorne and Magidor sever the connection between communication theory and the theory of action, and they have given no answer to what is actually understood by the auditor when she interprets 'He is on fire'. Indeed, they have not given us an account of how one interprets an utterance of that sentence.[6]

8. Conclusion

We doubt that Hawthorne and Magidor's argument for the non-transparency of presupposition is sound. But if we are wrong, we think the considerations above show that principles of rational communication are such as to push the higher-order presuppositions of speakers and auditors in line with their first-order presuppositions. Well-run conversations will tend toward transparency. The reason is that well-run conversations involve cooperative audience members

[5] Thanks here to John Hawthorne (personal communication).

[6] We think Stalnaker (2009) has a worry like this one in mind when he presents his playing card example at the end of his response to Hawthorne and Magidor.

who freely and regularly accommodate in order to remove obstacles that would otherwise impede successful interpretation. This result may have interesting epistemological applications. At any rate, we see no reason to think that the failure of epistemic transparency presents special problems for Stalnaker's theory of conversation. But even if we are wrong about this, there is an additional point to keep in mind.

Suppose presupposition is non-transparent, and that the non-transparency of presupposition does indeed pose the sort of problem for Stalnaker's theory that Hawthorne and Magidor describe. A question is left open: How regular are failures of transparency in everyday conversation? For all that Hawthorne and Magidor have shown, such cases are quite rare. Perhaps they are so rare that a theorist can blamelessly ignore them. Stalnaker's model is, after all, an idealization.[7]

References

Hawthorne, J. and O. Magidor 2009: 'Assertion, Context, and Epistemic Accessibility'. *Mind*, 118, pp. 377–97.

Perry, John 1988: 'Cognitive Significance and New Theories of Reference', in his 2000, pp. 189–206. Originally published in *Noûs*, 22, pp. 1–18.

—— 2000: *The Problem of the Essential Indexical and Other Essays*. Stanford: CSLI Publications.

Stalnaker, Robert 1978: 'Assertion', in his 1999, pp. 130–49. Originally published in *Syntax and Semantics*, 9, pp. 315–32.

—— 1998: 'On the Representation of Context', in his 1999, pp. 96–113. Originally published in *Journal of Logic, Language, and Information*, 7, pp. 3–19.

—— 1999: *Context and Content*. Oxford: OUP.

—— 2002: 'Common Ground'. *Linguistics and Philosophy*, 25, pp. 701–21.

—— 2009: 'On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility'. *Mind*, 118, pp. 399–409.

[7] Thanks to Robert Stalnaker for specific comments and questions as well as general discussion about related topics.

