CHAPTER 2

A Theory of Common Ground

Abstract. A new theory of common ground is proposed.

2.1. Introduction

In the 1970s Robert Stalnaker (and, independently, Lauri Karttunen) described the notion of the *common ground* of a conversation: a body of information shared among participants in the conversation (Stalnaker 1974, 1978, Karttunen 1974).[1] In the first instance, Stalnaker proposed that these were the propositions participants "mutually assume that they take for granted". But in later work (especially 1998, 2002, 2014), he has elaborated this idea, arguing that the attitudes which determine the common ground have the "iterative" structure of common belief, where a group commonly believes a proposition just in case all believe it, all believe that all believe it, and so on.

At the outset, the common ground was envisioned as an enrichment of formal approaches to semantics, allowing for an equally formal explanation of certain pragmatic phenomena. Stalnaker's theory of assertion, for example, uses properties of the common ground to explain divergences between the semantic content of a sentence in context and what is said by an assertion of that sentence in that context. Stalnaker himself has continued to emphasize, broadly following Grice, that the laws of pragmatics—including, for example, the theory of assertion—should be derivable from general laws of rational interaction. Since the common ground is a primitive postulate of Stalnaker's pragmatic theory, he accordingly holds that the common ground itself is a mandatory feature of anything which is to count as rational communicative activity.

Some aspects of the theory of the common ground are hotly contested today, even among those who use the common ground in their own linguistic theories. For example, "dynamic" semanticists argue that the common ground plays an even more central role in communication than Stalnaker believes it does. These theorists have developed the hypothesis that the value of a sentence is a function from common grounds to common

[1]Thanks to Frank Arntzenius, Emil Möller, Daniel Rothschild and Timothy Williamson for detailed comments on this material. Bob Stalnaker generously gave helpful comments on a much earlier draft. Parts of this paper are deeply indebted to numerous conversations with Jeremy Goodman.

grounds (often in this setting called "contexts").[2] Whereas Stalnaker's original theory adhered to a more traditional notion of semantic content, these theories make the much more radical step of holding that the meaning of some if not all expressions cannot be understood in isolation from their effect on the common ground. In a second, equally prominent example, discourse representation theorists argue that the common ground must have a highly articulated logical structure to account for phenomena such as tense anaphora.[3] Whereas the common ground was initially understood as a body of information with only Boolean structure, discourse representation theorists have endowed it with a much more complex structure, which also requires new laws governing the dynamics of update.

But one aspect of Stalnaker's theory has enjoyed almost unanimous support: that if there is a common ground, the attitudes which determine what is common ground have the iterative structure of common belief.[4] Stalnaker and his followers have proposed various different hypotheses about the precise nature of the attitudes which determine the common ground, but in each case they have agreed that the attitudes have this infinitely iterated structure.[5] In other traditions, too, the point is often simply taken for granted. To return to the earlier examples, dynamic semanticists and discourse representation theorists are not often given over to detailed discussion of the epistemology of the common ground (their "contexts"). But the minimal discussion one does find refers favorably to Stalnaker's views on the common ground.[6]

But is this unanimity justified? There are at least two reasons for thinking that Stalnaker's theory of the epistemology or psychology of the common ground is mistaken. First, the "iterated" theory of common ground makes intuitively irrelevant and even unnecessary features of agents' psychology essential to the possibility of communication. To see this, suppose scientists discover that an isolated linguistic community—whom I will call "Luxembourgers"—do not possess a part of the brain which is active whenever non-Luxembourgers think about what others believe they believe others believe they believe (and similarly, for other propositional attitudes relevant to conversation). Moreover, if you ask a Luxembourger about people's higher-order beliefs beyond this level, she will shrug and say that she has no idea about attitudes of this kind. Worse still: except for

[2]For a good survey of this development, see van Eijck and Visser 2012. Standard references include Heim 1983; Groenendijk and Stokhof 1991 and Veltman 1996.
[3]For an excellent overview, see Kamp et al. 2011.
[4]An important complicating example is Veltman 1996.
[5]See, e.g. Clark and Marshall 1981, Clark 1996, Yalcin 2007: 1007 for examples.
[6]See, for a "dynamic" example, Portner 2007: 357; for an (admittedly less clear) example from discourse representation theory, see van Eijck and Kamp 2011: 191.

the consistent shrug, Luxembourgers' dispositions to act regarding higher-order beliefs beyond the specified limit are so inconsistent that by far the best hypothesis concerning their behavior is that they have no such beliefs.

If the common ground has the infinitely iterated structure of common belief, this scientific discovery about Luxembourgers would amount to the discovery that Luxembourgers do not have a practice of assertion. In fact, given Stalnaker's views about the importance of the common ground to communication, some of his remarks suggest that we should think of this discovery as a discovery that Luxembourgers' practice of communication is fundamentally defective. But both of these judgments are implausible. It seems far more likely—at least on a first pass—that the members of the hypothesized community might be very normal, even with respect to their communicative practices. And this point holds regardless of whether we think people in fact generally *do* commonly believe propositions which are relevant to their conversations. The point is simply that the common ground should not be defined so that it requires an infinite hierarchy of attitudes.

A second example of the same form may help to clarify the point. Suppose it turns out that building computers which can "think" about others' thoughts is considerably more resource intensive than building them with the ability to think about others' actions and concrete physical objects. We may suppose we are in a situation where it would require a whole new set of programming tools, as well as new hardware, to endow computers with the capacity to describe others' thoughts. Moreover, to make the point as clear as possible, let us suppose that with each embedding of attitudes (thinking about others' thoughts about others' thoughts...), the programs and hardware required to implement the possibility of machines engaging in this sort of "cognitive" activity become harder and harder to design. And, finally, suppose that we know that, even if we did successfully write these programs, and build this hardware, the computing power utilized by the new processes would swamp all other processes.[7]

Here again, it seems that we would not need to overcome all of these design hurdles in order to create a computer which could mimic various basic communicative activities.

[7]According to some philosophers—and to a large body of work in psychology—the hypothesized Luxembourgers and these computers are in fact much like real people: using theory of mind in reasoning requires a great deal of effort, and humans' dispositions regarding embedded attitude reports are highly unstable. For the first point, see Lin et al. 2010; for the second see especially Kinderman et al. 1998 (cf. Stiller and Dunbar 2007). In fact, one might even go so far as to suggest that the computational difficulties described in the main text correspond to evolutionary pressures away from devoting cognitive resources to unimportant questions about iterated attitudes (unless the capacity for them was a consequence of some other, more fundamental capacity).

The computers might be capable of performing a wide range of normal conversational behavior, even if they do not have an infinite capacity for thoughts about thoughts about thoughts. Depending on the precise limitations of these machines, they would be more or less obviously different from us—for example, in their ability to parse attitude reports or make inferences about others' beliefs. But these differences in degree do not seem to be ones which would prevent them from engaging in many of our ordinary linguistic practices.

The second reason for doubting Stalnaker's theory of the epistemology or psychology of the common ground is that the theory predicts a sharp divergence between one-on-one, face-to-face communication and other kinds of communication. Consider the following example:

> Sam is at the airport, about to leave town for a few days. He knows that his neighbor has set up a device which will destroy Sam's house either today at noon or tomorrow at noon. Sam knows that the day of destruction depends on the outcome of a coin flip the device will perform at 11:30 today. Sam knows that his wife will stop by today at 1pm, to stay at home for the night. He has left a note telling her about his neighbor's plans and the device. If she sees it, the house will not have been destroyed, but will be destroyed only the following day.

In this case, the contents of Sam's message cannot be commonly believed, since Sam does not believe that his house has not yet been destroyed, and so does not believe that his wife believes that Sam knows about the neighbor's dastardly plans. If the message gets through, Sam's wife will know that Sam knows about the plans, but she will also know that Sam does not know whether she knows that he knows about the plans.

If the common ground has the structure of common belief, this kind of message passing will be an unusual or defective form of communication. Written communication does seem to involve a variety of complexities which may not be present in face-to-face conversations, but these complexities do not seem to mean that, at the appropriate level of abstraction, note passing is a fundamentally different form of communicative activity than conversation. To mention just one consideration, agents who have a practice of assertion would thereby be able to understand taped recordings of speakers of their own language. And it seems they would be able to do this, even if they did not have the capacity to pretend to have a common ground (or to have whatever other complex

attitudinal states one might wish to invoke to preserve the iterated theory in the face of noisy message passing examples such as the one above).

More mundane examples than Sam's odd situation also generate a related problem. People who make speeches at weddings know that some listeners will be whispering to their neighbors, but their speeches are generally unaffected by the fact that they don't know who their audience is. Even more clearly, it doesn't bother speakers in such situations that their audience doesn't know who the audience is—although such knowledge *would* be required if the group were to achieve common knowledge. This example is no different from that of writers who send their books into the world to be read by an unspecified audience. As in our first example of written notes, these more mundane examples are still instances of intricate social practices which undoubtedly involve very complex psychological states and cultural backgrounds. But once again, these forms of communication do not, at the appropriate level of abstraction, strike us as forms of communication that are fundamentally different from one-on-one, face-to-face communication.

These two reasons for doubting Stalnaker's theory of the epistemology or psychology of the common ground are certainly not full-fledged arguments against the theory. I don't wish to pretend that the proponent of this "iterated theory" of common ground has no way of responding to them. But these two problems with the Stalnaker's view do motivate the search for a different theory of common ground, which delivers more intuitive verdicts about these cases. The aim of this paper is to provide just such a theory of common ground.

In developing this theory, I will follow Stalnaker in considering only propositional attitudes. This means that the theory I propose here will be limited in important ways. For example, my models of common ground cannot be used to describe some of the more complex structural features of the common ground we find in discourse representation theory. But just as with Stalnaker's theory, I intend my theory as a framework or starting point, from which other developments may begin. Much of the structure missing in my theory could easily be added. To take a simple example, the universe of objects utilized in discourse representation theory can be explicated in a first-order modal language with a distinguished predicate of "salience". We say that if it's common ground that o is salient, then o belongs to the universe of objects. Many other structural features—for example ordering sources, probability distributions, and self-locating attitudes—can be included by similarly small extensions of the theory presented here.

But still these extensions will be left for another day. Our question is: what attitudes must agents be capable of having about each other in order to have something recognizable as the common ground? Since this question can be posed using a purely propositional language, we will only need a propositional theory to answer this question.

The plan of the paper is then as follows. Section 2 introduces the notion of conversational acceptance, and canvasses various different theories of acceptance. The bulk of the paper is then dedicated to showing how the minimal theory can be used to explain three important conversational phenomena: speaker presupposition (Section 3); update of the common ground (Section 5); and the infelicity and felicity of certain patterns involving epistemic sentences (Section 6). Section 4 takes a brief digression to consider a direct argument for the iterated theory as opposed to the minimal one. Section 7 concludes. The Appendices present a variety of formal facts used in the main text. The main text can be read without referring to the appendices. The interested reader may find it most convenient to read a given appendix immediately after reading the main text which refers to it.

2.2. Defining the Common Ground

I will develop the theory of common ground using the notion of "conversational acceptance", or "acceptance for the purposes of conversation". When I use the word "accept", it will mean "conversationally accept". Using this notion, I define a sentential operator "it's common ground that", as follows:

Common Ground: It's common ground that p if and only if all participants accept that p.

But what is acceptance? Although I will use this notion extensively in the paper, I will not settle this important question here. Instead, my aim will be to understand the notion—and its role in theories of conversation better—by canvassing various theories of acceptance. Accordingly, I'll begin with a taxonomy.

2.2.1. Acceptance. Our first two theories of acceptance make use of the idea that a conversation has what we might call a *conversational tone*.[8] This conversational tone determines what attitude is appropriate to the propositions introduced in the course of the conversation. This attitude might be belief, knowledge, supposing, pretending, or something else. I will make the simplifying assumption that conversational tone determines a unique attitude, although in real conversations different attitudes may be

[8]The phrase comes from Yalcin 2007.

appropriate to different (subsets of the) propositions which are relevant to the progress of the conversation.

The three theories of acceptance I will consider are as follows:

(1-Identity) To accept a proposition for the purposes of conversation is to take the attitude determined by the conversational tone to that proposition.

(2-Necessity) To accept a proposition for the purposes of conversation it is necessary but not sufficient that one take the attitude determined by the conversational tone to that proposition.

There are three different ways in which this might be so:

(a-Attention) To accept a proposition is to attend to it, and to take the attitude determined by the conversational tone to it.

(b-Table) To accept a proposition is to believe that the proposition is "on the table", and to take the attitude determined by the conversational tone to that proposition.

(b1-Reducible) To believe a proposition is "on the table" is to believe that all others take the attitude determined by the conversational tone to this proposition.

(b2-Primitive) The property of "being on the table" is taken as a primitive property, which is understood independently of others' attitudes, perhaps explicated as a mixture of relevance, salience, importance and other features of appropriateness.

(c-Distinct) Acceptance is a distinct psychological attitude, which happens to be accompanied by the attitudes determined by the conversational tone. Whereas the foregoing views of type (1) and (2) define acceptance in terms of other attitudes, this final theory does not

take such a strong view of the relationship between acceptance and the attitude determined by the conversational tone.

(3-Primitive) Acceptance is a primitive attitude, and one need not take other attitudes to propositions one accepts. This view may deny the idea of a conversational tone altogether, or it may simply hold that conversational tone is not very important. If the idea of a conversational tone is maintained, this last view holds that one may accept a proposition without supposing it, even when the conversational tone determines supposition as attitude appropriate to that proposition.

It will be useful to begin with a few remarks about the advantages and disadvantages of these theories of acceptance, and its role in the common ground.

The first, "Identity" theory of acceptance takes conversation to be a psychologically heterogeneous phenomenon: there is no single attitude which is common to all conversations. For some, this feature of the theory may be undesirable: the phenomenon of conversation may seem sufficiently uniform that it demands a unified psychological explanation. A second, perhaps more interesting problem with the identity theory concerns examples in which belief is the attitude determined by the conversational tone, but in which one does not accept every proposition one believes. For example, even if everyone at a fancy party can see that there is toothpaste on the hostess's forehead, the polite party-goers may still not presuppose this proposition in the course of their conversations. The first theory of acceptance owes us an explanation of this phenomenon.[9]

The Necessity theories of acceptance offer various different answers to the first of these questions—about the psychological uniformity of conversation. For example, both (2a) and (2b1) offer a conservative story about the attitude which unifies conversations. Neither of these theories posit a new, distinct attitude which we employ only in conversation; instead they hold that conversation involves a combination of attitudes which are familiar from other areas. In the Attention theory, this attitude is attention, plus the attitude determined by the conversational tone; in (2b1), the story involves higher-order attitudes, although only one iteration—not the infinitely iterated attitudes mentioned in

[9]I don't take either of *these* arguments to be decisive against the identity theory. But we'll see in a moment that the Identity theory is unattractive on Stalnaker's own view of the common ground. Later, in Section 5, we'll see that the minimal theory also has trouble given this view of the nature of acceptance. So there is some reason to doubt the prospects of this theory.

the introduction. (The precise strategy for avoiding iteration in (2b1) is somewhat subtle; later we will spend more time understanding its mechanics.)

But both of these fairly conservative theories are similar to the Identity theory, as far as the second kind of example is concerned: neither offers a simple solution to the story about the hostess's toothpaste. In the example, it's plausible that all parties may not just believe that the toothpaste is there, staring them in the face, they may also be attending to it, however discreetly. Similarly, everyone may not just believe that the toothpaste is there, but believe that all others believe it is there, as well. In any case, each of these three theories may advocate the plausible position that the phenomenon should be explained in terms of general social norms, independent from the norms of conversation.

(2b2) sits at a threshold between the more conservative theories above it, and the more radical ones which follow. On the one hand, it uses only the familiar attitude of belief, but on the other it invokes a new, primitive property. A theorist of this kind must offer us some explanations of this new property, and in particular some explanation for why it is not to be understood in terms of others' attitudes. As so often, this explanation need not be particularly informative: it may be that the logic of the theory of acceptance provides the best explication of what it is for a proposition to be "on the table", and that we cannot hope for anything more informative.[10]

The final two theories of acceptance (2c) and (3) do not suffer from the putative difficulties about the unity of conversation or the presence of toothpaste, because they may take the attitude of acceptance to have whatever properties they like. This flexibility, however, comes at a fairly severe cost: of introducing a previously undiscovered attitude. According to these two theories, acceptance is supposed to be a very important attitude, which pervades our social lives, and yet somehow we do not have a word for this attitude. One way of responding to this problem is to identify acceptance with an attitude for which we do have a name. For example, one might think that presupposition is an attitude which cannot be explained in terms of others. One might then replace "accepts" in my statements with "presupposes".[11] But in any case, the "Distinctness" theory, (2c), suffers from this problem less severely than (3), the "Primitive" theory, since (2c) can explain the absence of a word for this attitude by the fact that our acceptances are always masked by the

[10]In this paper, I won't define how propositional quantification would work officially, in a defined object language. This is mainly because I won't introduce a language or logic at all. Unofficially, I will be able to quantify over propositions since propositions are just sets of worlds. In any case, it should be clear that formalizing this theory by introducing propositional quantifiers and constants would be easily done.
[11]For the relationship of this view to Seth Yalcin's theory of the common ground, see below, n. 15.

attitude determined by the conversational tone. Once again, however, this explanation comes at a certain cost: the Distinctness theory, (2c), owes us an explanation for why these different, distinct attitudes are always found in pairs. Is this a matter of metaphysical necessity, or a contingent feature of our psychology, or some other, alternative form of connection?

The Primitive theory, (3), has a further interesting feature. Moore sentences, such as "it's raining but I don't believe it is", are widely agreed to be unassertable in all contexts. A natural explanation, available to all theorists of the first and the second kind, is that in contexts where one asserts a sentence, if one has the attitude determined by the conversational tone toward a proposition then one also believes that proposition. These sentences are then unassertable because they cannot be truly believed. (Other sentences, involving knowledge or certainty, might force stronger views of the attitude determined by the conversational tone, but assuming that true belief is required will be enough for our purposes in this paragraph.) But the theorist of our third kind—of primitive acceptance—cannot straightforwardly adopt this kind of explanation. He or she must explain the unassertability of Moore-sentences by different means—perhaps by a separate norm governing assertion: for example, that one should assert a proposition only if one believes it. This theory would separate the norms for assertion from the norm that one should converse in such a way that one's contributions have the appropriate relationship to the common ground. This in itself is not a problem for the theory, it simply marks a difference between the primitive theory and the earlier theories which place emphasis on the notion of the conversational tone.

These different theories of acceptance will be useful in later sections—especially Section 6 on "epistemic sentences"—when we come to consider different styles of explaining important conversational phenomena. As I've said, I will not attempt to decide between these theories here, but merely to understand their different properties in the context of the minimal theory of common ground.

2.2.2. Stalnaker's Theory. Stalnaker has offered a variety of different theories of the common ground, which it will be useful to have before us for the sake of comparison.

Recall that, where ϕ is an arbitrary propositional attitude, a group mutually ϕs (or: mutually ϕ^1s that p) if all ϕ that p, and mutually ϕ^ns that p if they mutually ϕ^1 that they mutually ϕ^{n-1} that p. They then commonly ϕ that p just in case, for all natural

numbers n, they mutually ϕ^n that p. Common belief, common knowledge, and common acceptance are all defined in this way, substituting "believe", "know" and "accept" for ϕ.

Stalnaker's first theory takes acceptance to be determined by conversational tone, as above in (1), and then defines

(CB-A): It's S_1-common ground that p if and only if the parties commonly believe that they accept that p. (1998, 2002)

One could also use knowledge in place of belief:

(CK-A): It's S_2-common ground that p if and only if the parties commonly know that they accept that p.

(This definition could be seen as derived from Yalcin 2007, although see below.)

More recently, Stalnaker uses acceptance, apparently as a primitive attitude, although he does not give a detailed account of its nature:

(CA): It's S_3-common ground that p if and only if the parties commonly accept that p, (2014)

where acceptance is a primitive attitude of type (3). (When I don't want to disambiguate between the S_1, S_2 and S_3 sentential operators, I'll simply write "it's S-common ground that".)

Each of the first two of these theories could be developed in tandem with any of the theories of acceptance described in the previous section. The proponent of (CA), however, cannot adopt either of the theories of acceptance (1) or (2). The reason is fairly easy to see. If I ask you to suppose that we do not exist, then it may become common ground that we do not exist. But if our supposition is to be consistent, we cannot then also be supposing that we are supposing that we do not exist, for that would be a way of supposing that, after all, we do exist. Since it cannot coherently be commonly supposed that we do not exist, one must, on this theory of acceptance, be able to accept a proposition without bearing the attitude determined by the conversational tone to it. (Rejecting closure for supposition is not a real way out of this problem, since while our suppositions certainly often fail to be logically closed, we *might* engage in a conversation where we were aware of the conflict in supposing that we do not exist and supposing that we are supposing that we do not exist.)

This is a deep problem for related theories of the common ground. To see this, consider the proposal of Seth Yalcin 2007. Yalcin takes "presuppose" as a primitive, and holds that it validates the axiom: S presupposes that p if and only if S presupposes that

the others presuppose that p.[12] Yalcin then defines common ground as the propositions that subjects commonly know that they presuppose, making his theory at least formally analogous to CK-A. He does not articulate the claim that one presupposes that p only if one bears the attitude determined by the conversational tone to p, but he does say that the conversational tone determines the attitudes one should bear to the propositions in the common ground. When we coordinate on a conversational tone, we commonly know that this is the attitude we should bear to the propositions in the common ground. But then, at least on a Stalnakerian, coarse-grained view of content, the various iteration principles Yalcin endorses lead him back into the problem just described. For suppose it's commonly known that: if p belongs to the common ground, I am supposing it. Then by definition of the common ground, it will be commonly known that, if it's commonly known that we all presuppose that p, I am supposing it. But because one presupposes that p only if one presupposes that others do, then if it's commonly known that all presuppose that p, it's also commonly known that all presuppose that all presuppose that p. So if it is common ground that p, it will be common ground that all presuppose that p. So in addition to supposing that p, I will have to suppose that all presuppose that p. But if I wish to suppose that we do not exist, then I am engaging in an incoherent supposition, for I am also supposing that all presuppose that they do not exist, and hence that they both do and do not exist.

Once again, my point is not to discard CA, Yalcin's theory or any other. I wish merely to point out that neither CA nor Yalcin's theory can endorse theories (1) or (2) about the relationship of acceptance (respectively, presupposition) to the attitude determined by the conversational tone. Later, we'll see that the minimal theory should also reject (1), although it may be some small point in favor of the minimal theory that it can accept theories of type (2), and thus is not forced to posit a new social attitude of acceptance or presupposition.

Before moving to the main argument, it will be useful to have Stalnaker's notion of the "context set" or *the* common ground. To state this idea, we use the framework of multi-agent Hintikka-Kripke models in epistemic logic, and in fact we will restrict attention to models which have a finite state space.[13] Each agent has three binary accessibility

[12]Presumably "others" includes oneself, so that one presupposes that p only if one presupposes that one presupposes that p, but what I say in the main text won't rely on this assumption.

[13]The Appendices to this chapter develop the full formal apparatus behind this paper. There, I use the more general setting of neighborhood models of belief and knowledge. The Appendices are designed so that they could be read on their own, or could be read, one at a time, as they are referred to by the main text. See now Appendix A.

relations, associated with the attitudes of acceptance, belief, and knowledge. We think entirely from the perspective of the model, without introducing a language or logic, and define belief, knowledge and acceptance by functions from propositions to propositions, that is, sets of worlds within a universe Ω. Thus for example, $B_i(E)$ takes the proposition that E to the proposition that i believes that E; a world w belongs to the proposition $B_i(E)$ just in case $R_i^B(w) \subseteq E$, where $R(w)$ is the set of worlds accessible from w.

Common knowledge, belief or acceptance is defined using the transitive closure of the union of the agents' accessibility relations, whether for acceptance, belief or knowledge. If this relation is denoted $R^*(w)$, we define (for example) the common belief function on propositions by $w \in CB(E)$ if and only if $R^*(w) \subseteq E$.

We use these definitions to define, at last, a common ground accessibility relation. For example, according to the minimal theory of common ground, the common ground accessibility relation is the union of the agents' accessibility relations for acceptance. According to **CA**, by contrast, the relevant relation is the transitive closure of the agents' accessibility relations for acceptance. We can then define:

Context Set: The context set or the common ground is the set of worlds reachable in one step by the common ground accessibility relation.

When discussing Stalnaker's theories, I will sometimes speak of the S-context set or the S-common ground to avoid confusion. It should also be emphasized that the noun phrase "the common ground" does not have the same denotation as the sentential operator "it's common ground that". In our simplified Kripke models, the two have the following relationship: a world belongs to the context set just in case it's not common ground that that world does not obtain. In more general settings, however, as in the next chapter, this relationship would not be preserved.

2.3. Presupposition and Accommodation

The next four sections will consider and respond to a series of purported arguments for Stalnaker's theory as opposed to the minimal one. In each case, I show that these arguments do not succeed. In responding to the arguments, I develop the mechanics of the minimal theory in more detail. I also show how the arguments which supposedly tell in favor of Stalnaker's theory lead to considerations which in fact tell against his theory and in favor of the minimal one.

The first argument for the iterative conception of the common ground is based on Stalnaker's theory of speaker presupposition. Stalnaker has proposed that a speaker S-presupposes that p if and only if the speaker believes (or, now, accepts) that it's S-common ground that p. If this elegant theory of speaker presupposition depends on Stalnaker's iterated theory of the common ground, then one might take this fact to support this theory of the common ground. Since the theory of speaker presupposition seems attractive in its own right, we should prefer a theory of common ground which allows us to explain speaker presupposition in this way.

My main strategy for defusing this argument will be to show that Stalnaker's theory of speaker presupposition does not in fact depend on Stalnaker's iterated theory of the common ground. But before turning to this main response, let me first make one point about Stalnaker's theory of speaker, since I think this point already diminishes the force of the proposed argument. The point is that Stalnaker's theory of *speaker* presupposition is at least not clearly a theory of the interesting phenomenon in the vicinity of presupposition. Speakers have a clear grasp of sentential presupposition, which is independent of their understanding of the attitudes of a specific speaker on a specific occasion. But the deep phenomena in the area seem to be facts about sentential presupposition—not speaker presupposition. So it is at best unclear how powerful an argument based on Stalnaker's theory of *speaker* presupposition could be.

While I think this is an important observation about Stalnaker's theory of presupposition, in the present context it is just an aside. We can respond to the proposed argument without settling the importance of speaker presuppositions or their potential relevance to the theory of sentential presupposition, because, as I will now argue, the minimal theory can exactly reproduce Stalnaker's theory of speaker presupposition.

Presupposition: A speaker S presupposes that p if and only if S accepts that p and believes that the others accept that p.

Sometimes, speakers can introduce new information by presupposing it. This phenomenon of presupposition *accommodation* is one of the main data which make it seem that the theory of speaker presupposition captures an important component of conversational behavior. So it is important to show that the minimal theory can also explain accommodation. And in fact, it can: the minimal theory can replicate Stalnaker's law governing the dynamics of presupposition accommodation.

Accommodation: If B believes that A presupposes that p, and B comes to accept that p, then B presupposes that p.[14]

Since Presupposition and Accommodation are the main principles of Stalnaker's theory of speaker presupposition, this completes the argument that the minimal theory can also deliver the main components of that theory. There are of course many further issues about presupposition. But I will not discuss them here. It suffices for our purposes to see that the core of Stalnaker's theory of presupposition can be replicated by the minimal theory. The defects of Stalnaker's theory of speaker presupposition will also be defects of the new theory, but so will its virtues. There is thus no argument based on speaker presuppositions in favor of the iterated theory of common ground as opposed to the minimal theory.

There are two principles which my theory of presupposition does not deliver, but which Stalnaker's CA theory, the theory of S_3-common ground does. Most important is a principle of "social" introspection. Discussing presupposition, Stalnaker writes: "What is most distinctive about this propositional attitude is that it is a social or public attitude: one presupposes that φ only if one presupposes that others presuppose it as well." (2002: 701) Following Stalnaker, Seth Yalcin agrees that "one presupposes that p only if one presupposes that one's interlocutors presuppose that p" (Yalcin 2007: 1007). Related to this principle of "social" introspection is the principle of "positive introspection": which Stalnaker want the logic of presupposition to obey: "if Alice presupposes that ϕ, then she presupposes that she presupposes that ϕ" (Stalnaker 2002: 708).

Now in general, where a is some attitude, if a group commonly a's that p, they commonly a that they commonly a that p. This "positive introspection" principle simply follows by definition: a group commonly a's that p if and only if, for all natural numbers n they mutually a^n that p. But then it's clear that for all natural numbers n, they mutually a^n that they mutually a^n that p. This is why Stalnaker's (2014) theory, the CA theory delivers both of these principles.[15]

[14]See, e.g. Stalnaker's 2002; the idea stretches back to his 1974, and has been developed in a number of places. This is as it should be, since one of the aims is to show that the minimal theory can account for the principles he holds are crucial to the notion of common ground.

[15]The emphasis on the "social introspection" and "positive introspection" principles in Stalnaker's (2002), however, is somewhat surprising, since the theory of presupposition in that paper invalidates both of these supposedly crucial principles. There, it is understood that belief is not sufficient for conversational acceptance: there may be settings in which the conversational tone determines attitudes other than belief or knowledge as the attitude appropriate to the conversation. Presupposition is defined as above: we say that one S_1-presupposes that p if and only if one believes that it's S_1-common ground that p. So if one S_1-presupposes that p, it follows that one *believes* that others S_1 presuppose that p, but since one need not *accept* this, nor believe that others accept it, it does not follow that one S_1 presupposes that one S_1-presupposes that p, nor that one S_1 presupposes that others S_1-presuppose that p. As we've seen above Yalcin 2007 takes presupposition as a primitive attitude, so he's not subject to this particular

It may be helpful to see why the minimal theory fails to validate this principle, in the subtlest case, the theory of acceptance (2b). Recall that according to this position, to accept a proposition is in part to believe that it is on the table, where for a proposition to be on the table is for all participants to adopt the attitude determined by the conversational tone to that proposition. Consider a conversation in which the conversational tone determines the attitude of belief. According to this view, it's common ground that p in this conversation just in case the participants mutually believe that p, and mutually believe^2 p. (Note that the proposition that all believe that p may not belong to the common ground, since even though all believe that all believe that p, it may not be the case that all believe that all believe that all believe it.)

Now suppose in this same context that a speaker s presupposes that p: that is, s believes that p, s believes that the participants mutually believe that p , and s believes that the participants mutually believe^2 that p. To presuppose that p, s need not believe that the participants mutually believe^3 that p. But if s presupposes that all presuppose that p, s would have to believe that the participants mutually believe^n that p, for $3 \leqslant n \leqslant 6$. So the question whether a speaker presupposes that p and the question whether a speaker presupposes that others presuppose that p are simply independent questions, which may receive different answers.

But if anything the failure of the minimal theory to validate this principle seems an advantage, not a disadvantage. Suppose in a context where the conversational tone determines the attitude of belief, I tell you "I can't come to the meeting, I have to pick up my sister from the airport." With all the qualifications already mentioned about the notion of speaker presupposition, it's plausible that I presuppose in the relevant sense that I have a sister. But do I presuppose that you believe that I believe that you believe that I have a sister? Utterances involving this kind of iteration are hard to process, so giving a concrete example of an *obviously* infelicitous sentence of this kind does not seem possible. (At any rate, I haven't been able to come up with an example.) But the putative presupposition of the speaker in this example is odd. This is especially clear if we consider again the relationship between speaker presupposition and sentential presupposition. The sentential presuppositions of the sentence "I have to pick up my sister from the airport" do not make reference to the attitudes of others at all. One might be attracted to the schema: if a sentence s presupposes (sententially) that p, a speaker utters s felicitously only if

problem, although, as we'll discuss in the next section, he still does not validate positive introspection for the common ground itself.

the speaker presupposes that p. But this schema cannot be used to explain the bizarre *speaker* presupposition in our example. For the sentence has no sentential presuppositions whatsoever about the attitudes of the participants in any conversation.

In fact, there is some reason to suspect that Stalnaker and Yalcin's endorsement of the iteration principle for presupposition stems from a misunderstanding of a much more obviously desirable principle. Consider the following "maxim" presented as an imperative, following Grice:

Aim-P: Presuppose that p if and only if your interlocutors presuppose that p![16]

An important goal of many conversations is to coordinate on a body of information. We do this in part by attempting to bring our presuppositions in line with each other. Presuppositions about others' presuppositions are not important to achieving this aim of coordination. What is important is that our presuppositions in fact agree. One very good way to ensure this kind of agreement is to know what others presuppose, and bring one's presuppositions into accord with what one knows they presuppose. But even in this good case, it is knowledge of others' presuppositions, not presuppositions about them, which play the role of ensuring coordination on the relevant body of information. In any case, one might reasonably hold that the activity of coordinating on this body of information is rarely if ever as sophisticated as even the knowledge-based picture suggests. Much more usually, we adjust our presuppositions, one step at a time without thinking about it. We try a variety of alternatives, perhaps in parallel, until we find a hypothesis about others which seems to work in processing the conversation we are having. Sometimes, these adjustments involve beliefs about the attitudes of others, but perhaps just as often they involve adjustments in our beliefs about the world. In this latter case, we do not even need beliefs or knowledge about others' presuppositions: we coordinate on a body of information by presupposing only what is true.

For some of our theories of acceptance, the maxim Aim-P can be explained by a second, semi-Gricean maxim, governing acceptance:

Aim-A: Accept that p if and only if your interlocutors accept that p!

Typically, when we move around the world, we update our beliefs directly by acquiring new information. According to the minimal theory, if Aim-A is satisfied in a given context, we can follow a similar practice in conversation. Since in this case, for each speaker s, it's common ground that p if and only if s accepts that p, s doesn't need to think about

[16]This is related to Stalnaker's notions of "non-defective" and "equilibrium" contexts. See the appendix to the next chapter for more detailed discussion of these notions.

the common ground to decide what has been said or to update his own beliefs and acceptances. Instead, he must simply refer to what he himself accepts.

This "Gricean" maxim for acceptance is not available on the "Identity" theory of acceptance, (1). Aim-A will sometimes conflict with the aims of other attitudes, for example, belief. But the other theories of acceptance can endorse this maxim, perhaps claiming it as an aim of the special attitude of acceptance. This approach is perhaps most natural for the theories which take acceptance as primitive or as defined in terms of a new primitive, that is, theories (2b2), (2c) and (3). According to (2b2), for example, one should believe that a proposition is on the table if and only if the others do, too. For each of these theories, Aim-A could state a special aim of acceptance, just as truth or knowledge is the aim of belief.

2.4. Informativeness and Positive Introspection

My main aim is to show how the minimal theory accounts for some basic conversational phenomena, as a way of responding to various objections to this minimal theory. Sections 5 and 6 will continue to pursue this project. But in this section, we take a brief digression to consider some differences between Stalnaker's "iterated" theory and the minimal one I have been proposing. These differences have been thought to provide the basis for an argument against the minimal theory, so although this discussion will be a digression from the project of explaining conversational behavior, we are still dealing with a topic which is central to the goals of this paper.

In Stalnaker 2014, we find a number of places where Stalnaker offers the beginnings of an argument for the claim that the common ground has the structure of common belief:

> When one speaks, one presupposes (takes it to be common ground) that one is speaking, and this means that the conversation is taking place, not only in the actual situation, but in each of the possible situations in the context set. This fact is reflected in the formal representation of the common ground by the iterative structure.

This passage, from Chapter 7, comes closer to a direct argument:

> When we are engaged in conversation, it is common ground that we are engaged in that particular conversation. This is crucial for the role of common ground in constraining the means used to communicate, and is built into the iterative structure of common ground. (The iterative

> structure implies that information presupposed includes information about what is being presupposed in the conversation in question.)

It is important to distinguish the claim Stalnaker is making here from something that sounds similar, but which he cannot be saying. When people are speaking they tend to believe that they are. Presumably all of the theories of acceptance on offer above would agree that people do not just believe that they are speaking, they accept this, too. But if all parties accept that Bob is speaking, for example, then on the minimal conception of the common ground, too, it will be common ground that Bob is speaking. "In each of the possibilities in the context set", this act will be taking place. Similarly, since we do tend to accept propositions about others' beliefs in the context of a conversation (and about what others accept), our presuppositions will typically include "information about what is being presupposed in the conversation in question." The fact that "we are engaged in that particular conversation" is typically part of the common ground in the minimal sense of common ground, just as much as it is in Stalnaker's.

Instead Stalnaker is, I think, adverting to the fact that his theory S_3 delivers the principle of presupposition discussed in the previous section: that if one presupposes that p, one presupposes that all others also presuppose that p. In fact, as we've discussed, S_3 also validates the principle that if it's common ground that p, it's common ground that it's common ground that p. The minimal theory does not validate this principle. (For reasons mentioned earlier, the theories S_1 and S_2 do not deliver the relevant principle for presupposition *or* for common ground. Even on Yalcin's version of S_2, where presupposition is taken as a primitive, and assumed to be "socially" introspective, we are not guaranteed to satisfy the principle that if it's common ground that p, it's common ground that it's common ground that p, since participants may commonly know that p but nevertheless fail to presuppose that they commonly know that p.)

From a formal perspective, the fact that Stalnaker's latest theory validates this "positive introspection" axiom for common ground has what might seem to be important consequences. If we think again in terms of our Hintikka-Kripke models of belief, knowledge and acceptance, the positive introspection axiom has the consequence that from any point within the context set, the context set from that point will be a subset of the actual one.[17] In this sense, as Stalnaker says, positive introspection for the common ground ensures that the common ground "constrain[s] the means used to communicate".

[17]The point is subtler in the neighborhood frames used in the Appendices, since the definition of the context set does not have such a simple relationship to the operator "it's common ground that".

But Stalnaker buys this particular formal constraint at the cost of preventing the common ground from constraining communication in much more important ways. On his view, the common ground is constrained with respect to what it "thinks" the common ground is. But it is much less constrained with respect to many other matters of importance, which do not concern the common ground. To see this, consider the following example.

Louis comes from Luxembourg. I overheard him saying he's from Luxembourg the other day, so I know he's Luxembourgish. Louis slyly saw me listening to his conversation when Luxembourg came up; he knows I know he's from Luxembourg. But I saw him looking at me only later, when he was talking about the Netherlands, so I think he thinks I think that, in spite of his name, he's Dutch. Louis saw me see him see me watching him only when the conversation turned to France, so he thinks I think he thinks I think he's from France.

Now if Louis and I find ourselves in a conversation where the attitude determined by the conversational tone is belief, the minimal theory predicts a much more informative common ground than Stalnaker's theory does. Here, it's S-common ground that Louis is from Luxembourg or the Netherlands or France, and that's it. (The example could of course be extended so that we *commonly* believe nothing about Louis' origins. It would thus only be S-common ground that he's from some country.) But according to the minimal theory, it's common ground that Louis is from Luxembourg. In fact, if we take the simplest version of the minimal theory, where to accept a proposition is to take the attitude determined by the conversational tone to that proposition, then it will also be common ground that it's common ground that Louis is from Luxembourg. (It won't, however, be common ground that it's common ground that it's common ground he's from Luxembourg.) So while S-common ground does ensure that facts about the common ground will belong to the common ground, it does so at the cost of the common ground being much less informative than it might be.

The point is not specific to this example: it's easy to see that in *any* situation, if it's S-common ground that p, it's common ground (in my sense) that p, while the converse doesn't hold in general. If facts about the world *are* commonly believed, then they're also believed, so if the conversational tone determines the attitude of belief, they will be common ground in my sense, just as much as in Stalnaker's sense. It's just that the minimal theory of common ground says that much more is common ground, since we may mutually believe many propositions we fail to commonly believe.

In fact, this kind of example allows us to add a new motivation for seeking an alternative to Stalnaker's theory of common ground. This new example will also help us to develop a more concrete sense for the difference between the two theories.

Suppose that in the serious context described above, where one should accept a proposition only if one believes it, I say, "Wow, Louis, you must be very glad you're going home soon!" He replies "Yes, my country is very beautiful."

What does Louis say? In a context like this one, it's S-common ground that p if and only if it's common belief that p (or common knowledge that we believe that p). The most informative proposition of this kind, at least as far as Louis' origins are concerned, is that he's from Luxembourg, France or Holland.

Our example will now concern Stalnaker's theory of assertion. Recall that the diagonal proposition expressed by a sentence contains exactly those worlds w such that the semantic content of the sentence *as uttered at* w contains w itself. According to Stalnaker when phrases such as "your country" are uttered in a context where it's not common ground which country that is, the semantic content of the sentence cannot be asserted. Instead, we take the asserted content to be the diagonal proposition. Thus in the situation just described, Stalnaker would hold that Louis asserts the diagonal proposition, the proposition that if he's from Luxembourg, it's beautiful and if he's from the Netherlands, it's beautiful and if he's from France it's beautiful. Louis doesn't say very much.

On my view, by contrast, it is common ground that Louis comes from Luxembourg. So we don't need to consider the diagonal proposition: Louis asserts the semantic content of the sentence. He says that that Luxembourg is beautiful. Since the context is much more informative, the content of what is said is much more informative, too.

As with our first two examples considered in the introduction, this example is hardly an argument for my view as opposed to Stalnaker's. It's hard to disentangle our judgements about the semantic value of Louis's assertion from what we take the content of his assertion to be. And even if one agrees with my judgement about what Louis says in this context, one might respond to the example by rejecting the use of diagonalization, while holding fast to Stalnaker's theory of common ground. But still, the example adds to the growing body of strange predictions made by the "iterative" conception of the common ground. My theory of common ground, by contrast, guarantees what seems the right result: Louis successfully communicates the proposition that Luxembourg is beautiful, and not the far weaker claim that, if he's from a country, it's beautiful.

2.5. Update

We now end the digression and return to the project of demonstrating how the minimal theory of common ground can explain important conversational phenomena. In this section I consider perhaps the most important desideratum for a theory of common ground: that it be able to explain how the common ground alters in the course of the conversation. In the next section, I will then turn to a final question: how to explain the felicity and infelicity of certain patterns involving epistemic vocabulary.

An influential cluster of arguments for the "iterative" conception of the common ground concerns update, and, in particular, the "transparency" of update. In this section, I will consider two such arguments, and show that they are not really arguments for the iterative conception of the common ground at all. I will do this by demonstrating that these demands can be met within the minimal theory just as easily as within the iterative one.

The first argument takes the perspective of the speaker. It begins from the observation that linguistic communication involves intentional action. It is then claimed that a necessary condition for intending to perform an action is that one believes one can perform it, if one chooses to do so. In the context of communication, it is urged, if speakers are to be capable of the intentional action of, say, assertion, there must be some effect on context such that they believe their utterances will have that effect.

A second line of thought takes the perspective of the listener, and is more direct. Instead of beginning with abstract claims about intentional action, it begins from the claim that in ordinary circumstances we are not in fact at a loss as to how to update context on hearing others' utterances. We do not generally find ourselves in a state of confusion when someone makes an assertion or asks a question. If we do find ourselves confused, we count the utterances "defective", and our theory should aim to predict their defectiveness.

These demands become stronger if they are paired with specific commitments about how the "effect" of an utterance is calculated from the semantic value of a sentence together with features of the common ground itself. Stalnaker holds that it's a norm of rational communication that one perform an utterance u in a context c only if, for all w in c, what is said by u in c is the same as what is said by u in c_w. The only way to guarantee this identity within Stalnaker's theory by ensuring that, for all w in c, $c = c_w$. I will be arguing that *this* demand for the relationship of *contexts* to their "candidate

contexts" is much stronger than what is required to make sense of the possibility of updating contexts.[18]

2.5.1. Intentional Action? Our first task is to think more carefully about intentional action.

Consider the following game. There is a row of equally marvelous and indestructible prizes arrayed on a shelf. One wins a prize just in case one manages to knock it off the shelf. To do so, one must operate a mysterious machine, by pressing one of two buttons, the green or the red. The mysterious machine has a number of states, corresponding to the marvelous prizes. If the machine is in state 1, and one presses the green button, then the machine will knock off the first prize. If it is in state 2, it knocks off the second prize, and so on for all of the prizes. Players of the game are given no information about the state of this mysterious machine. But they know that the red button, regardless of what the state of the machine is, knocks off no prizes, whereas the green, regardless of the state of the machine, will knock off a prize.

Players of this game may be unable to coherently form certain intentions. For example, because of their limited information about the state of the machine, it seems plausible that there is no prize such that they can coherently intend to knock off that prize. They may want to knock off one rather than another, but they are uncertain of the outcome of their actions, and so cannot have intentions about which one they will knock off. On the other hand, they can have other intentions: they can intend, in pressing the green button, to knock off some prize.

I propose that conversation is not so different from this mysterious machine. We can have the appropriate kind of intentions about what we are saying, even if we do not always know enough to know what precise effect our utterances will have on the common ground.

---

[18]The fact that Stalnaker's theory of assertion requires negative introspection for the common ground was observed by Hawthorne and Magidor 2009. Appendix 2.B gives more detailed discussion, correcting one important formal error in their argument. It may be worth remarking here on one feature of the discussion inspired by their paper (Stalnaker 2009, Almotahari and Glick 2010, Hawthorne and Magidor 2011). A casual reader of that debate would be forgiven for thinking that the central issue between those authors concerns introspection principles for individual attitudes such as knowledge and belief. But the problem about access to asserted content is independent of this debate (as Hawthorne and Magidor recognized (2009: 387)). Even granting that belief and acceptance satisfy positive and negative introspection, or that knowledge satisfies positive introspection, *common belief* and *common knowledge* still won't satisfy negative introspection. Stalnaker needs further assumptions to achieve that result. But also, even if one denies that individual attitudes satisfy positive introspection, common belief and common knowledge *will* satisfy positive introspection, as a matter of logic. So the real issue concerns negative introspection for the common ground, and this issue is discussed in Appendices 2.C and 2.D.

The idea will be that there is a set or "cloud" of contexts in play at any moment of a conversation.[19] We may say different things in each of these contexts. But the typical situation is that the differences in each context are not great enough or are not of the right kind to impede our choice of the "gist" of our utterances. (When the differences are great or important, an utterance may be infelicitous, as I'll discuss below.) Our capacity for intentional action in communication should be assessed against our capacity to choose this "gist" of our utterances, not our capacity to choose the precise effect of what we say. Communication is more like the pursuit of what Rawls called an "overlapping consensus" as opposed to a pursuit of precise agreement. We attempt to select our utterances, and interpret others as selecting utterances, so that the differences in what our utterances express according to the different contexts won't matter to the main points we are making.

But where does this cloud come from? It's time to make these vague ideas a bit more precise.

2.5.2. Minimal Transparency. Part of being a competent speaker is knowing how to compute a new context set from two inputs: the old context set and an utterance. The following principle makes official an idealized version of this idea:

Transparency: For any competent speaker i, context c, and utterance u, if c updated on u is c', then i knows that c updated on u is c'.

Transparency is only an idealized principle because it's implausible that we need to have settled dispositions about *every* utterance and context to be competent speakers. Moreover, for many utterances, there can be room for reasonable disagreement about the effects of expressions on different contexts; the meaning of expressions is not as transparent as Transparency suggests. But this kind of semantic uncertainty affects all theories which claim that a successful assertion requires that one know what one is saying. For example, on Stalnaker's theory of assertion, too, if one does not know how to compute the semantic value of an utterance at a world, one will be unable to diagonalize. If we assume Transparency, we can abstract from this kind of semantic uncertainty, to focus on issues related to the common ground itself.

Now, given this principle, we want to understand the origins of the cloud of contexts. The following definition introduces two natural notions of "candidate" contexts:

[19] I'm borrowing the word "cloud" from Kai von Fintel and Anthony Gillies (see, e.g. 2011 118-119). They borrow it, in turn, from Kratzer. I suppose it ultimately dates back to very natural picture-thinking about Lewis's view in Lewis 1975, or to early discussions of supervaluationism.

Candidate Context: A proposition c_w is an *i-candidate context* or a *candidate context for i* just in case i does not believe a proposition which entails that c_w is not the context set.

A proposition c_w is a *candidate context* if it's not common ground that p, for any p which entails that c_w is not the context set.[20]

It is worth emphasizing that the definition of an i-candidate context uses *belief*, not acceptance. What's important is not that i *accept* that the context has a certain form, but what i *in propria persona* thinks the context might be. In general, when we consider the agent's individual view of the situation, it is no longer appropriate to consider what the agent *accepts* about what the common ground is, but rather what the agent *believes* it is.

Transparency has the consequence that, for any *set* of contexts, the agent "knows how to update" each of its members. In particular, the agent will know how to update all of the candidate contexts, and so will correctly map the cloud of contexts to a new cloud, in which each context has been updated in the appropriate way.

But in fact, once we have the idea of candidate contexts, we can relax the idealization of Transparency. Instead, we can define a class of circumstances in which i will understand what is said:

Minimal Transparency: An utterance u is *transparent* for an agent i if, for any i-candidate context c, if c updated on u is c', then i knows that c updated on u is c'.

u is *ultra-transparent* for i if, for any candidate context c, if c updated on u is c', then i knows that c updated on u is c'.

In other words, i does not need to know this for every utterance whatsoever, but when utterances succeed (or succeed as far as i is concerned), it is because i knows what to do in the relevant "candidate" contexts.

[20]The use of the "entails" clause is needed for them to be usable in the neighborhood frames in the Appendices. In Kripke frames, we could have just written: "A proposition c_w is an *i-candidate context* or a *candidate context for i* just in case i does not believe that c_w is not the context set," and made an analogous change to the definition of a candidate context. Note also that I could have written "c_w is the context", and made no reference to the specific world-based implementation of the idea of the common ground. Thus a context here could be a very different object, with much more structure than just a set of possible worlds.

With these definitions out of the way, we can come to our real task: to make the earlier analogy to the game with marvelous prizes more precise. Say that an agent i *i-globally asserts* that p if, for every i-candidate context c, if c were the context, she would assert that p. In other words, one globally asserts that p if, when we "supervaluate" over candidate contexts, the contexts agree that one asserts that p. My proposal is then that one may satisfy the requirements of intentional action by choosing an utterance which ensures that one *globally* asserts that p, even if one is uncertain which context set is the actual one. The content of the utterance may differ as regards other propositions in the different candidate contexts. There may be some i-candidate contexts in which one asserts p and q but not r, and others where one asserts p and r, but does not assert q. This kind of uncertainty affects one's ability to have coherent intentions regarding an assertion that q, but it doesn't affect the possibility of asserting that p, and intending to do so. Just as in the game described in the previous subsection, we may intend to perform one action even if our epistemic position is too weak for us to coherently have other, more specific intentions.

So even if one is uncertain about the context, it may be coherent to have the intention to assert that p. This of course does not mean that for any set of contexts C, it will be possible to assert that p, if one knows only that the context is some member of C. But it does demonstrate that the motivation from intentional action does not force us to the position that speakers always know, of some context c, that it is the context.

We still have one further duty to discharge: to show that the "clouds of contexts" model can offer a reasonable picture of update. I'll turn to that task first, and then close by considering some lingering objections.

2.5.3. Updating Sets of Contexts. In the clouds of contexts model, there are two forms of update. The first is to alter which contexts are candidates, whether simply i-candidates or candidates; the second is to update each i-candidate or candidate context according to what is said in that context.

Some updates can occur only if the common ground satisfies some preconditions. For example, on one theory of epistemic modals, "it might be that p" is felicitous only if there's a world in the context set where p. In the sets of contexts model, this translates to the requirement that there be *some* i-candidate or simply candidate context which satisfies the precondition. Update now breaks into two cases. If no candidate context satisfies the precondition, the utterance is infelicitous, and the conversation breaks down. If some

candidate context satisfies the precondition, we rule out those i-candidate or candidate contexts which fail to satisfy it, and update those which do.

In the second kind of update, the situation is equally simple. We update each candidate context by what would be said in that context were it actual. We map the old set of contexts into a new set of contexts according to the usual rules. We tend to identify what has been said by i with what has been said i-globally: what is said in every i-candidate context. This reflects the fact that we often identify what i said with what i meant to say: i's false beliefs about what the context is are sometimes more important in determining what i said, than what the context in fact was.

But identifying what has been said is of course more delicate and flexible than these rigid rules suggest. If, for some listener j, what is said in two different j-candidate contexts is *very* different, especially concerning questions of great import to the conversation, it forces j to choose between two ways of proceeding. If one of the two j-candidate contexts yields a much more plausible view of what has been said, then j may simply update her beliefs by ruling out the candidate which yielded the implausible verdict. This may be so, even if the context in question did not yield a strictly speaking infelicitous utterance. On other occasions, however, what is said in different candidate contexts may be equally plausible, so that j is forced to ask for clarification of the ambiguous utterance. (Or she may proceed even though she recognizes that there are two possible disambiguations of what i just said; she trusts that the issue will be resolved in due course.)

This discussion brings us, however, to a lingering problem. What I've said might seem plausible when we restrict to small numbers of candidate contexts. But won't the number of candidate contexts typically be very large? In that case, how can we possibly represent all the candidate contexts we need to represent? I'll close by considering this issue, and also a related issue about learning.

2.5.4. Representation and Learning. The minimal theory appears to impose huge representational demands on speakers and hearers, since they have to compute what would be said in a vast array of candidate contexts. Do we really have to think about all of these i-candidate contexts to understand what has been said in a conversation?

To dramatize the problem, suppose we've both read *Our Mutual Friend* and remember what happens at the end. But suppose that I don't know that you've read it, and you don't know that I have. Now suppose we're discussing the World Cup, so that the conversational tone determines (as it should) the very serious attitude of belief or knowledge.

Now according to the first theory of acceptance (Identity), we're uncertain about what the context is. This problem doesn't just arise from Victorian literature: anything you might believe will be relevant to determining what the context in fact is. In such serious contexts, for example, there will be variability regarding my grandmother's birthday, the color of my favorite copy of my favorite book, and so on.

The minimal theorist can offer two different responses to this problem. First, he might note that the uncertainty here clearly does not impede our ability to communicate. We don't need to know *everything* about the context to update it. For example, suppose you respond to my comment about the German performance, by saying "They played very well last night". It would be absurd to hold that the value of this sentence depends on whether you've read *Our Mutual Friend* or not. So I can add the appropriate proposition to the common ground even though I'm uncertain what propositions exactly belong to the common ground. Since most utterances are sensitive only to a select range of features of the common ground, listeners don't have to know what the common ground is exactly in order to update it according to these utterances. In this sense, our beliefs about context are similar to our beliefs about many things: we can get by just fine even if we represent contexts only partially. We don't have to think about different possible contexts at a conscious or representational level in order to update them. We simply update the context which we represent partially, by adding new facts to our partial representation.

A second way of responding is to hold that in the example above, it's not in fact common ground that we've read *Our Mutual Friend*. This may be so either because we're not attending to this proposition (2a), or it's not on the table (2b), or we're not accepting it for some other reason (2c) and (3). Each of these theories could lead to a view on which it's not common ground that we've read *Our Mutual Friend*, and so, to a theory according to which we're not in fact uncertain about the context in the situation above.

Both of these responses are, I think, both powerful and plausible. But the discussion brings us to a second objection, about learning. Both Transparency and Minimal Transparency make use of speakers' and listeners' ability to compute updates on contexts "pointwise". But how did we acquire this ability? If we are generally uncertain about what the context is, then we will learn to update sets of contexts, not individual contexts. So it seems that the minimal theory of update undermines itself.

My response to this problem will depend on a preliminary observation. Throughout the preceding discussion I have been slowly moving the reader to thinking in terms of

i-candidate contexts as opposed to candidate contexts more generally. The reason for this shift should be clear, but it is nonetheless an important point. If i thinks there is only one possible context, and j's utterance is felicitous in that context, then even if (owing to some of k's beliefs, for example) j's utterance was infelicitous, i won't have any trouble updating on it. To think in terms of candidate contexts is to take an objectionably "context-eye" view of the conversation. The conversation is not about what the context "thinks", it is about what the participants think.

This observation is important because it shows that what is required to meet the demands of learning is not that the *context* know what the context is, but that each agent know what the context is. Stalnaker holds that it's crucial that the same thing be asserted at every point in the context set. But it should now be clear that this demand is too strong. It's much more important that, for each individual i, the same thing be asserted in every context i thinks might be the actual one.

This suggests the answer to our problem. All we need to give a theory for how agents learn how to update individual contexts is a set of cases in which individual agents know what the context is. If a child encounters situations in which she knows, of some c, that c is the context, and if later behavior confirms to her that she acted appropriately when she updated it as she did, then she will learn how to update individual contexts. It needn't *always* be the case that speakers and hearers know, of some set c that it is the context, for them to learn how to update contexts appropriately.

This response has a different character depending on our theory of acceptance. As noted before, the theories (2) and (3) of acceptance may hold that contexts are not very complex objects. There may be comparatively few propositions in a given context, since we may only accept relatively few propositions. So it can be fairly easy to know, on these theories, what the context is.

The "Identity" theorist has a harder time explaining the prevalence of contexts in which we "know what the context is". The most plausible version of this view is to hold that one only needs to know select facts about a context in order to know how to update it on a given utterance. Learning how to update "pointwise" does not then require that we know exactly what the context is in a very strong sense; we need only know some features of the context. A child who knows that it was common ground that the Germans played a soccer game may learn how to update any context which satisfies this condition (even if he or she didn't know which context was the actual one). (This version of the Identity theory, however, is in a sense a notational variant of some version of theory (2) of

acceptance. For why should we not, in the situation above, simply say that the common ground contains only the *relevant* propositions, and that the child did in fact know the exact context?)

I want to close now by noting one structural relationship between the simple-minded version of the identity theory and the iterative theory. Suppose we define a class of situations where it's common ground that p if and only if all agents know that it's common ground that p. These are the situations in which we "know what the context is". If we adopt the "Identity" theory (1) of acceptance, plus this simple-minded view of what it takes to know what the common ground is, then if these situations can be ones in which the conversational tone determines the attitude of knowledge, it will be common ground that p if and only if it's common knowledge that p. In these situations, in other words, the minimal theory will collapse into the iterative theory.

This is an interesting formal property, but, I think not an important conceptual fact. As I've argued, theorists of type (1) should hold that the contexts which are crucial for learning how to update are not ones where it's common ground that p if and only if participants know that it is. In these special contexts, they should hold that it's common ground that p *and* relevant that p if and only if participants know that it's common ground that p. As I've noted above there's a way in which this is a notational variant of a view such as (2b) where "on the table" is explicated by a primitive property of relevance or something like relevance. But notational variant or not, it seems to be the only hope for a minimal theory of this kind.

In any case, the formal collapse only occurs within the Identity theory (and given the simple-minded and implausible view of what it takes to know what the context is according to that theory). If in contexts where the conversational tone determines the attitude of knowledge, knowing a proposition is still not sufficient for accepting that proposition, then individual agents can know the common ground without the common ground having the iterative structure of common belief. More carefully: even in situations where the conversational tone determines knowledge, and it's common ground that p if and only if all agents know that it's common ground that p, it may still fail to be the case that if it's common ground that p, it's common ground that it is common ground that p. Those who accept one of the theories (2) or (3) for acceptance may thus define a set of contexts where *all agents* know what the context is. These special contexts will still not have the structure of common belief.[21] Minimal theorists of this kind may hold that

[21]The formal argument for this point is given in Appendix 2.E.

we learn how to update the common ground pointwise by engaging in conversational situations where we know what the common ground is. And they may do so without invoking the existence of any contexts where the relevant attitudes have the structure of common belief.

I've shown how the minimal theory can explain some important demands on the common ground. The minimal theory can still deliver a sense in which communication is a form of intentional action, since one may know the important effects one's utterance will have on the context without knowing what the context is. Moreover, listeners may know how to update the context even if they don't know which context is the actual one. I closed by responding to some putative problems about representation and learning. I showed that, appearances to the contrary, the minimal theory does not raise special problems about the difficulty of representing contexts or about learning how to update individual contexts. Finally, we saw how once we take an agent's-eye view of the conversation, there need be no collapse of the special contexts where we know what the common ground is into cases where the attitudes which determine what's common ground have the structure of common belief.

2.6. Epistemic Sentences

We now turn to our final topic: sentences which include epistemic vocabulary. I'll consider three example sentences, and how they are to be explained on each of the theories of common ground and update. My strategy here will be somewhat different from the strategy in previous sections. Instead of opting for a specific explanation of the phenomena, as I did with presupposition and update, I will aim to map the territory around these sentences. Our primary interest in these sentences or patterns is whether they can form the basis of an argument against the minimal theory of common ground. The short answer is that they cannot. There are two very plausible ways of explaining the sentences which force no alteration to the theory offered thus far. At the end of the responses, I'll note the ways in which the responses affect the KK principle, since some recent authors have attempted to use these epistemic sentences to argue for this principle.

The three sentences—or, properly, schemas—are:

(I) "p and I don't believe that p"; "p and I don't know p".

(II) "p and I might not believe that p."

(III) A pattern: The question "how ϕ?" presupposes that ϕ. But in most settings it's acceptable to challenge someone's assertion that p by asking: "how do you know that p?"

In general, the various views of (I) will be much the same, regardless of one's theory of the common ground. The other sentences are subtler, and this section will mainly be devoted to explaining these sentences.

2.6.1. Replies Based on the Logic of Acceptance. (I) We saw already that theory (3) of acceptance—that acceptance is a primitive attitude—can't explain Moore sentences via the logic of acceptance. For this subsection and the next, then, I won't consider theories of type (3).

On the other theories of acceptance, where speakers accept a proposition only if they also bear the attitude determined by the conversational tone to the proposition, the logic of acceptance itself explains (I). According to Stalnaker's theory of common ground, as well as to my own, one of the aims of assertion is to add a proposition to the common ground. If the conversational tone of a conversation where one makes assertions is knowledge, then both sentences are explained immediately. Since one accepts a proposition only if one knows it, then one can accept the proposition that p in an assertional context only if one knows that p, and so it cannot be the case that in that context one also doesn't know that p.

If (II) and (III) are to be explained in terms of acceptance, however, we need some elaboration of the logic of acceptance. The obvious principle is:

KA: An agent accepts that p only if she knows that she accepts that p.

(II) Let's first see how this principle explains (II). Suppose that the conversational tone for a given conversation is knowledge, so that one accepts that p only if one knows it, according to theories (1) and (2) of acceptance. Next, by KA, one accepts a proposition only if one knows that one accepts it. Now, we assume that agents also know that if they accept that p in such a context, they know that p. If they draw the inference on the basis of what they know, it follows from **KA** that they accept that p in such a context only if they know that they know it. We are now very close to the conclusion. To assert that p, one must accept it in a way appropriate to assertion. So then by **KA**, one must know that one knows it. So one cannot both assert that p and "I might not know that p", since the latter entails that one does not know that one knows that p.

(III) Does this principle also explain the felicity of sentences such as (III)? The question "how do you know?" stops a conversation in its tracks. The usual way of understanding the phenomenology of this "arrest" is to take it to be the same as that of accommodation. Let me defend this claim briefly.[22]

One reason for thinking that accommodation is involved in this utterance derives from the "hey wait a minute" test. Suppose that after I assert that p, you ask me: "how do you know?" Now I can fairly easily say, "I didn't say that I know that p, I just said that p". What usually happens in this circumstance is that we reach an impasse. You have indicated that you do not accept my utterance, and I have not satisfactorily answered your challenge. But in making this concession, I have not retracted my assertion, and this is the crucial point. My assertion has been rendered inefficacious because of your attitude to the proposition. I will be forced to find other ways of making my point. But if I am stubborn, I don't have to concede that I shouldn't have made this assertion in the first place. So in making my initial assertion I did not presuppose that I knew p; your question introduced this presupposition and asked me to accommodate to include it in my own presuppositions. (Of course, there may be circumstances where I admit that I shouldn't have made the assertion. But this is usually where I concede that I don't know that p itself, not ones where I merely concede that I don't know whether I know p.)

Given this picture of accommodation, the principle **KA** also helps to explain **(III)**. After the assertion that p, all agents accept that p, and all know that they do. So while p is presupposed by all, it's not necessarily the case that it's also presupposed that the asserter knows that p. But since (according to **KA**), all presuppose that p, it's fairly easy to accommodate, and start presupposing that the asserter knows that p.[23]

---

[22]In fact, in asking questions of the form "how ϕ?" *speakers* presuppose that the answerer knows how ϕ, and thus knows that ϕ. Is this a problem for the explanation in the main text. I don't think it is. First, this speaker presupposition is surely not the kind involved in semantic presupposition. Second, this presupposition need not be accepted by the answerer. I say "It's going to rain." You say: "How do you know?" I shrug, "I dunno, look at the sky." I've clearly rejected your presupposition that I know how I know, and rejected the appropriateness of the *presupposition* that I know that I know (I need not be denying the truth of this claim; I just don't think it should be added to context). So while an assertion provides a basis on which a reasonable conversant can and often does subsequently presuppose that I know that I know, the assertion doesn't make that fact uncontroversially belong to the common ground. In a sense "how do you know?" makes a similarly contentious *presupposition* that I know (once again, the contentiousness of the presupposition needn't derive from a concern about the truth or falsity of the claim). But in this case, the situation is different: I can't reject the presupposition that I know without retracting the assertion. And this is why it seems crucial to have a theory which accounts for this latter datum.

[23]A different principle from **KA** is **AA**. But as I've noted before, this can't be true for the theories (1) and (2) of acceptance. If in a conversation where the conversational tone determines supposition, I accept that we do not exist, then I will be supposing that we do not exist. But then I cannot also be supposing that we are supposing that we do not exist, on pain of engaging in an incoherent supposition.

2.6.2. Replies Based on the Logic of the Common Ground. It's possible to explain these problematic sentences within the minimal theory of common ground, by expanding the logic of acceptance to include the **KA** principle. An alternative form of explanation is to expand the logic of the common ground, without expanding the logic of acceptance. Doing this requires giving up on the most basic form of the minimal theory, but it also does not force us to go all the way to the iterative conception of the common ground. (Recall that in this subsection, as in the previous, I won't be thinking about theories of acceptance of type (3).)

Instead of the the **KA** principle, we can use:

2-Common Ground: It's 2-common ground that p if all accept that p and all know that all accept that p.

Given that the conversational tone of an assertional context is knowledge, then the explanation goes as in the previous section. One asserts p appropriately only if one knows that one knows that p, and then the explanation is as before. The difference between this theory and ones based on **KA** is that 2-Common Ground tells us nothing about the psychology of acceptance. But this revised theory of common ground should be unattractive to those who endorse theories (1) or (2) of acceptance, for a reason I've noted often before. If we suppose that we don't exist, so that it becomes common ground that we do not exist, then on pain of having incoherent suppositions, we cannot also be supposing that we're supposing that we do not exist.

2.6.3. Replies Based on Speech Acts. A third response to these problem sentences is perhaps the most powerful. According to this response, we do not need to alter the logic of either acceptance or common ground. Instead, the response invokes general facts about speech acts, deriving from yet more general principles governing certain actions.

For certain actions, one must have a particular standing in order to perform the action. If one does not have this standing, performing the action is inadmissible or even impossible. Timothy Williamson (2013: 82-3) considers the example of a commanding officer. To issue a command to his soldiers, an officer must have certain standing as an officer. He cannot issue this command, while (publicly) questioning his own standing as a lieutenant. (Of course, he can have private doubts about his ability or right to command, but he cannot in one breath command, and also question his standing to do so, out loud to his underlings.)

In Williamson's view, assertion is no different. To assert that p one must know that p. Just as with the lieutenant's command, then, it is incoherent to assert that p and (publicly) question one's own standing to assert that p. But this is precisely the performance one engages in if one asserts that p and then says "I might not know that p"; one questions one's own standing to make the assertion. Williamson suggests that this is a very general feature of communication. In support of this view, he adds a further example, of someone who says: "How are you feeling today?—and I have no interest in your answer."

This response to our sentences does not impinge at all on the minimal theory of common ground, or on the theories of acceptance which we have considered. Instead, it invokes general laws about speech acts, and considers how they apply to the speech act of assertion. (The explanation of the question "how do you know?" proceeds as discussed in 2.6.1. An assertion licenses the inference that the asserter has the standing to perform the assertion. So it's easy to accommodate a respondent's new presupposition that the speaker has that standing.)

It's worth noting again that Stalnaker's own theory of the common ground must invoke an explanation of this form for Moore sentences in general. Since, as we've seen, Stalnaker can't require that one accepts that p only if one bears the attitude determined by the conversational tone to p, he cannot claim that it follows from the logic of acceptance that the serious context of assertion demands the attitude of belief to the components of the common ground. So Stalnaker must offer an explanation of even standard Moore-sentences by appealing to facts about speech acts in general, and assertion in particular. Similarly, when we come to the extended Moore-sentences, Stalnaker must also think in terms of general norms of speech acts, as Williamson suggests we should.

This is not a problem for either theory of these patterns. But it does mean that Moore-sentences can't be used to argue in favor of Stalnaker's most recent theory of the common ground as opposed to the minimal theory. Since Stalnaker himself must adopt a theory of Moore-sentences which explains them as infelicitous because of features of certain speech acts (independent of the common ground), his theory and a minimal theory which adopts this response will be on a par. There is thus no argument based on these sentences against the minimal theory.

2.6.4. The KK Principle: Summarizing the Discussion. Before concluding the whole discussion, I want to discuss briefly the relationship between these responses

and the so-called KK principle. This is of interest because the data sentences in this section have recently been argued to pose a problem for a package of views endorsed by Williamson.[24] Williamson (2000) argues that knowledge is the norm of assertion. His arguments are directed at showing that *knowledge* that p, as opposed to (say) belief that p governs the practice of assertion. Williamson has also argued forcefully against the so-called "KK" principle, that if one knows a proposition one knows that one knows it.

The first point to note is that the explanation in 2.6.1 can be given within the minimal theory of common ground, and without invoking KK. Once again, this is a very important point which I think has been overlooked in the literature.

The explanation based on the KA principle is subtler. If we adopt the Identity theory of acceptance, then in a serious context where to accept a proposition is to know it, the KA principle will entail that if one accepts a proposition, one knows it, knows that one knows it, and so on. Now in general even this phenomenon does not lead to the return of the KK principle, which is a very general principle saying that *whenever* one knows a proposition, one knows that one does. But still the phenomenon would not be amenable to the spirit of Williamson's views since such an infinitely iterated hierarchy of knowledge would amount to a non-trivial "luminous" state.

But we've seen a number of reasons to doubt the Identity theory in any case. And if we adopt either theory (2) or (3) of acceptance, the KA principle will not imply the existence of such an infinitely iterated hierarchy. For since in a serious context knowing that p will be necessary for accepting that p (but not sufficient), then one may know that one accepts that p without accepting that one accepts it. Moreover, since the KA principle should be thought of as a norm, not as a fact about the psychological state of acceptance, there need be no requirement that acceptance itself is luminous.

Many will find this discussion of the possibility that one might fail to know one's own mind perverse. But for those who do believe that there are some failures to know one's own mind, it will be natural to think that certain propositions enjoy a special status, distinguished by the fact that one not merely has a given attitude toward them, but knows that one has this attitude. And it is clear that this status, while in general less important than the first order attitude itself, is an important one in the context of communication, where one's own standing with respect to the contents of one's acceptances may be called into question, and where others will use one's own utterance as strong evidence for facts about one's own acceptances. If I myself do not know that I believe a proposition, for

[24]Cohen and Comesana 2013, Greco 2014

example, it would be bizarre for me to convey the contents of this belief to others, since I would then be giving *them* evidence about the state of my own mind which I myself do not possess.

2.6.5. Conclusion. The epistemic sentences discussed in this section don't provide the basis of an argument against the minimal theory of common ground. We may preserve the minimal theory by expanding the logic of acceptance or by invoking general laws about speech acts. On the most promising theories of acceptance, the expansion of the logic of acceptance does not entail even a restricted KK principle.

2.7. Conclusion

I have argued that the main reasons people have favored the iterative conception of common ground are uncompelling. The minimal theory of common ground can explain speaker presupposition just as well as Stalnaker's theory; it yields a theory of common ground according to which common ground is guaranteed to be at least as informative as it would be on the iterative theory; it can account for the basic phenomena of update just as well as the iterative theory; and it can be used to explain extended Moore-like sentences.

Recall now the challenge with which we began. In certain cases of communication it is not possible for the attitudes of the participants to become a matter of common knowledge or even belief among the participants. In successful communication, the attitudes of the participants toward a given subject matter change. I tell you that Mary is coming to town, and you learn something about Mary. But according to the iterated theory of common ground, this change is not enough for normal or successful communication. It must also be that your change in attitude becomes public between us. Our opening examples put pressure on this idea. Communication seems normal, even when it is not possible to make the change in attitudes public between "speaker" and "hearer".

The minimal theory can be thought of as a return to the core Gricean thought which has animated Stalnaker's theory of language throughout. Communication induces complex changes in participants' attitudes. But communication is not in the first instance directed at making these changes in attitude public. This may often be a consequence of the ways in which our attitudes change throughout a conversation, but the fact that our change in view becomes public is not a core feature of communication. In my view, the common acceptance or common belief based theories of communication adulterate this pure Gricean thought by making the publicity of our changes of attitude central to the

theory of communication. Part of my aim in this paper has been to show that Stalnaker's most profound discoveries about the nature of communication do not depend on his view about the publicity of our changes of attitudes. Once Stalnaker's theory is liberated from the unnecessary encumbrance of common belief or acceptance, we can see more clearly the elegance, simplicity and robustness of his insights into the structure of conversations.

2.A. Models

In the Appendices, I use the formalism of pointed neighborhood models. I first introduce these models, and then discuss briefly why this formalism is appropriate.

2.A.1. Models. A pointed conversational model is a structure

$$\langle \Omega, a, I, (B)_{i\in I}, (K)_{i\in I}, (A)_{i\in I}\rangle,$$

where "propositions" are subsets of the set of states Ω, and $a \in \Omega$ is interpreted as the "actual state". I is then a set of agents, the participants in the conversation. The functions B_i, K_i, and A_i represent these participants' beliefs, knowledge, and acceptances: B_i maps a proposition E to the proposition that i believes E; K_i takes a proposition E to the proposition that i knows E, and A_i takes E to the proposition that i accepts E. As usual, I will use $\neg E$ to abbreviate $\Omega \setminus E$, and $E \to F$ for $(\neg E) \cup F$. Note that this last definition allows an analog of contraposition, since obviously $\neg E \cup F = \neg(\neg F) \cup \neg E$. (I will often also write propositions with lower case p's and q's, but the reader should remember that we don't have an object language in the normal sense.)

I will use $\Box$ as a variable over the operators A_i, B_i and K_i. The functions representing these attitudes can be used to define neighborhood functions, $N^{\Box} : \Omega \to \mathscr{P}(\mathscr{P}(\Omega))$, where $E \in N^{\Box}(w)$ just in case $w \in \Box E$. Neighborhood models generalize standard Kripke semantics for modal logic. Kripke models associate with each world an ultrafilter on the power set algebra of the state space. (That is, a filter on the algebra of propositions.) In our setting, these are the propositions an agent believes knows or accepts. Neighborhood models, by contrast, associate each world with an arbitrary subset of $\mathscr{P}(\Omega)$, an arbitrary set of propositions. This may be an ultrafilter on the algebra of propositions, but it needn't be.

In addition to the above, pointed conversational models satisfy three additional requirements:

Necessitation $\cap_{i\in I}\Box_i\Omega = \Omega$

Anti-Necessitation $\cup_{i\in I}\Box_i\emptyset = \emptyset$

Factivity $K_i(E) \to E = \Omega$.

The first two conditions ensure that there's always something the agent believes, and, moreover, the agent doesn't believe the absurdity. In Kripke models (to be introduced in a moment), the first condition has the consequence that the **D** axiom is satisfied ($\Box E \to \neg\Box\neg E$). The third condition is self-explanatory.

We define group operators as follows:

$$\Box_E(E) := \cap_{i\in I}\Box_i E.$$

(Here $\Box_E$ involves a slight abuse of notation; the idea is that "everyone accepts" is defined so that $A_E(E) := \cap_{i\in I}A_i E$, and similarly for "everyone knows" and "everyone believes".)

We then define the common ground:

$$CG(E) := A_E(E).$$

Finally, fixing a pointed conversational model, the event of it being commonly $\Box$'ed that E is defined recursively. Letting $\Box_E^1 E = \Box_E E$ and $\Box_E^n E = \Box_E\Box_E^{n-1}E$, we define: $C\Box = \cap_{n\in\mathbb{N}}\Box_E^n E$. This definition gives us a common knowledge operator CK, common belief, CB, and common acceptance CA.

2.A.2. Stalnaker Models. We now define propositions corresponding to the satisfaction of some standard axioms:

(K): $\cap_{i\in I}\cap_{\Box\in\{B_i,K_i,A_i\}}\cap_{E\subseteq\Omega}\cap_{F\subseteq\Omega}(\Box(E\to F)\to(\Box(E)\to\Box(F)))$

(C): $\cap_{i\in I}\cap_{\Box\in\{B_i,K_i,A_i\}}\cap_{E\subseteq\Omega}\cap_{F\subseteq\Omega}(\Box(E)\cap\Box(F))\to\Box(E\cap F)$

(4): $\cap_{i\in I}\cap_{\Box\in\{B_i,K_i,A_i\}}\cap_{E\subseteq\Omega}\Box(E)\to\Box\Box(E)$

(5): $\cap_{i\in I}\cap_{\Box\in\{B_i,K_i,A_i\}}\cap_{E\subseteq\Omega}\neg\Box(E)\to\Box\neg\Box(E)$

(BA4): $\cap_{i\in I}\cap_{E\subseteq\Omega}A_i(E)\to B_iA_i(E)$

(BA5): $\cap_{i\in I}\cap_{E\subseteq\Omega}\neg A_i(E)\to B_i\neg A_i(E)$.

A Hintikka-Kripke conversational model is a conversational model with $\mathbf{K}\cap\mathbf{C} = \Omega$. A $KD45$ conversational model is a Hintikka-Kripke conversational model with $\mathbf{4}\cap\mathbf{5} = \Omega$ (note that D is guaranteed by the semantic form of necessitation above). A Stalnaker model is a $KD45$ Hintikka-Kripke conversational model with $\mathbf{BA4}\cap\mathbf{BA5} = \Omega$. (Given these properties, it is unclear why we should not write this B as K, since one's introspective beliefs at least are always true, but we'll maintain this version of the axioms

for present purposes.) Finally, we define an a-Stalnaker model as a conversational model such that $a \in \mathbf{K} \cap \mathbf{C} \cap \mathbf{4} \cap \mathbf{5} \cap \mathbf{BA4} \cap \mathbf{BA5}$. Analogously a specific operator $\Box$ satisfies a-K if $a \in \cap_{E \subseteq \Omega} \cap_{F \subseteq \Omega} (\Box(E \to F) \to (\Box(E) \to \Box(F)))$, and similarly for the other conditions.

We can then define three notions of Stalnakerian common ground. Fixing a pointed conversational model:

(CB-A) $CG_{S_1} E = CBA_E E$

(CK-A) $CG_{S_2} E = CKA_E E$

(CA) $CG_{S_3} E = CAE$.

Note that these notions are defined for any conversational model.

2.A.3. The Context Set. It will also be useful to have the notion of a "context set". As above, in Kripke frames, we simply defined this as the set of worlds reachable by the common ground accessibility relation. In neighborhood frames, we need to do a bit more work to define this accessibility relation, since the individual agents' acceptances cannot necessarily be defined from such a relation. So we define a relation from a given function as follows, for an arbitrary function $\Box$:

$$R^{\Box}(w) = \{w' : \exists E[(\exists F \supseteq E)(F \in N^{\Box}(w)) \wedge (\nexists G \subsetneq E)(G \in N^{\Box}(w)) \wedge w' \in E]\}$$

(Note that this $\Box$ is no longer restricted to the B_i, K_i, A_i; it can be defined also for A_E, for example.)

For each operator, we can use this accessibility relation to define a new function:

$$\Box^C E = \{w : R^{\Box}(w) \subseteq E\}.$$

Since this is an operator defined from an accessibility relation, as in Kripke models, for any w, the set of propositions $\{E : w \in \Box^C(E)\}$ is a filter in the power set algebra of Ω. Moreover, if we started with a relational (Kripke) model of the operator $\Box$, it's obvious from the definition that if $w \in \Box(E)$ then $w \in \Box^C(E)$. (This will mean that if the A_i are given by an accessibility relation as in the main text, the definition of context set here will be equivalent to the one in the main text, since everything there was finite.)[25]

[25]In finite relational frames, of course, the exact converse holds, since there can be no infinite descending chains. In infinite relational frames, we have only an inexact converse: if $w \in \Box^C E$, then $(\forall F)($ if $F \subsetneq E$ then $w \in \Box(F))$.

To understand the interpretation of the defined operator, consider the case of acceptance, where we read A_i^C as "i assumes that".[26] In our general models, assumption is independent of acceptance. In this setting, propositions may be accepted but not assumed, for example, if an agent accepts only three propositions, Ω, E, F, where $E \subsetneq F$ and $F \subsetneq E$. Conversely, a proposition may be assumed but not accepted, for example, if an agent accepts only E and Ω, where $\Omega \setminus E \neq \emptyset$. One need not consider a proposition in order to assume it. "You're right: I hadn't thought about it," I might say, "I was simply assuming it" or "I was simply taking it for granted". Assumptions lie behind each of our surface, logically ill-behaved acceptances.

Now we define a context set for an arbitrary world

Context Set $c_w = \bigcup_{i \in I} R^{A_i}(w)$

(Note again, that in finite Kripke models, this definition is equivalent to the definition used in the main text.)

2.A.4. Neighborhood Models. The use of neighborhood models in this setting can be motivated in two ways. First, from a purely formal perspective, the generality of neighborhood frames gives us insight into what is important in the models. When we attempt to do logic without the law of excluded middle, familiar equivalences no longer hold. This forces us to be particularly careful about which assumptions are required in which derivations. If we only knew of models of our logic in which the law of the excluded middle held, we would have difficulty assessing which aspects of our logical systems were due to the law of the excluded middle, since we would not be able to consider models where that principle failed. The logic of belief and knowledge is no different. Making fewer background assumptions helps us to understand which assumptions are and are not needed for particular applications.

The use of these more general models can also be motivated from a psychological perspective. As is well known, standard Hintikka-Kripke models represent agents with unrealistically idealized logical competence. In Hintikka-Kripke models, if an agent believes that p, and believes that q, then she believes that $p \wedge q$. But many reject this view of belief. For example, proponents of the Lockean thesis, that belief is sufficiently high confidence, reject it because one's confidence in a conjunction may be lower than one's confidence in the conjuncts, and so fall below the threshold required for belief. In more banal examples, people may behave as if they do not believe the conjunction of two

[26]Note that assumption is sometimes used in a very different way in epistemic logic, e.g. Brandenburger and Keisler 2006. Our usage has no relationship to theirs.

propositions they believe because they have never considered the conjunction. Similarly, in Kripke models, if a subject believes that p, and p entails q, then the subject believes that q as well. But what if the subject has not acquired the conceptual resources to consider the proposition that q under any guise? In this case, her behavior also will not accord with her supposed "belief" that q.

The relaxation of the assumption in Kripke frames that every agent believes every theorem of propositional logic is in general less significant in neighborhood models. Neighborhood frames still do not allow us to represent distinct modes of presentation, so if an agent believes *any* instance of any theorem of propositional logic, the agent will be represented as believing all theorems. But it's very plausible that we each believe some instance of the law of excluded middle (for example). So we should represent agents as believing at least one propositional tautology. In my models, accordingly, I'll be assuming that every agent accepts (and believes, and knows) the propositional tautology, so in every model I consider, agents will still be omniscient with respect to propositional logic.

2.A.5. Pointed Models? In general, I will be focusing on pointed models, where the axioms are satisfied only at the actual world a. But is this the right choice? We could have treated propositional logic and the logic of belief on a par, by requiring that every world in every model satisfy the proposed axioms of common ground, acceptance, belief and knowledge. This, after all, is what happens in Hintikka-Kripke conversational models, and in Stalnaker models, too.

But this "global" method has the immediate consequence that, if an agent believes a propositional tautology, he or she will believe and know that the target axioms hold. It has a more drastic, less obvious consequence, too. Recall that agents mutually know^1 that p if all agents know that p. Agents mutually know^n that p if they mutually know^1 that they mutually know^{n-1} that p. Then agents commonly know that p if they mutually know^n that p for all natural numbers n. Now suppose, as is extremely plausible, that at every world in every model every agent knows *some* instance of a propositional tautology ("Look", I say to you "it's either raining now, or it's not"). Then the axioms will not only be known, but, more strikingly, they will be a matter of common knowledge. Knowledge is not special in this respect; we could have run the same argument with belief or acceptance in place of knowledge (or with certainty, if our models were endowed with probabilities).

The assumption that the axioms are commonly known will be most obviously uncongenial to those who understand features of the "logic" of belief as not really features

of logic at all, but contingent facts about human psychology and biology. For example, psycholinguists who conduct empirical studies of the common ground will hold that it is neither an a priori nor a necessary matter whether the beliefs of people pattern in a given way. There are of course *some* features of belief which are necessary or a priori—for example that it has contents—it's just that the interesting features of belief we aim to study are not among those special features. If one takes this view of the properties of belief which interest us here, it is not just natural, but mandatory that the agents in our models *not* be apprised of the full psychological and biological truth about belief. After all, it's precisely that contingent truth that we're trying to discover.

But pragmatic reasons also militate against this method of building models, reasons which apply even if we should suppose that the relevant laws of belief are an a priori or necessary matter, so that any creature who had beliefs at all would have beliefs which obey these laws. The method forces us to prejudge important questions about the role of iterated beliefs, precisely the target of our investigation. Since the class of models cannot draw the appropriate distinctions between situations in which the agents mutually believe[1] the axioms and one in which the agents commonly believe the axioms, they do not allow us to distinguish when a particular phenomenon is due to belief in the axioms, and when it is due to common belief. We should not choose models which are unable to distinguish importantly distinct hypotheses about our subject matter.

Moreover, granting again the hypothesis that the relevant laws of belief are a priori or necessary, we may still wish to treat these truths differently from the laws of propositional logic. It seems fairly clear that agents in the world do not behave as if they have perfect knowledge of propositional logic. It's a shortcoming of our models that they don't allow us to represent distinctions among different theorems of propositional logic: between the "obvious" ones and the "unobvious" ones. But we should be grateful that, in the case of axioms on belief, we are not compelled to adopt this unrealistic assumption. Pointed models allow us the chance to isolate a class of propositions for which the assumption of logical omniscience is not forced on us. Whatever our views of belief, we should not forego this chance.

2.B. Stalnaker's Theory of Assertion

Stalnaker's theory can be stated as involving three principles:

Uniformity: In cases of rational communication, the same proposition is asserted in every candidate context.[27]

Default: If no norm of rational conversation is violated, the content of an assertion of a sentence s is the semantic content of s in the actual world.[28]

Recall that Stalnaker holds that a standard repair strategy is to identify asserted content with *diagonal content.* The diagonal content of an utterance is the proposition that is true at a world if the semantic content of the utterance as made at that world is true, and false otherwise. Stalnaker's hope seems to be that if we identify asserted content with diagonal content, compliance with Uniformity will be restored (at least if the assertion has any chance of success). We strengthen this principle about diagonal content into a generalization:

Diagonalization: If the content of an assertion of s is not the semantic content of s, then the asserted content is the diagonal content of s.

To state these principles more formally, we need a more precise notion of utterances. For a given conversational model M, the set of utterances U (Kaplanian characters) will be the set of functions $u : \Omega \to 2^{\Omega} \setminus \{\emptyset\}$, where each utterance is understood to be a function from worlds to the semantic content of the utterance at that world (we assume, for simplicity, that contradictions are never uttered). For $u \in U$, we write $u(w)$ to denote the semantic content of u as uttered at w. We often think of this proposition as identified with its characteristic function, letting $u(w)(w') = 1$ iff $w' \in u(w)$, and 0 otherwise. The diagonal proposition is then constructed for each u as the set $u^d = \{w : w \in u(w)\}$. Note that we assume the set of utterances is very rich; every function from worlds to propositions is associated with an utterance.

Fixing a model and an utterance u, we write the asserted content of an utterance as $a_u : \Omega \to 2^{\Omega}$.

Now, fixing a pointed conversational model, the axioms Hawthorne and Magidor attribute to Stalnaker are:

Uniformity: $\forall w(\forall v, x \in c_w)(a_u(v) = a_u(x))$

HMDefault: $\forall w\ (\forall v \in c_w)a_u(v) = u(w) \Rightarrow a_u(w) = u(w)$

HMDiagonalization: $(\forall w)[((\exists v \in c_w)(a_u(v) \neq u(w)) \Rightarrow a_u(w) = u^d]$

---

[27] Hawthorne and Magidor 2009: 380; cf. Stalnaker 1978 [1999]: 88

[28] In fact, the formulation of this principle is due to Hawthorne and Magidor, but Stalnaker does not contest it in his reply (2009) to their paper.

(I use $\Rightarrow$ for the material conditional in the meta-language. It should not be mistaken for a strengthening of the set-theoretically defined $\rightarrow$.)

These laws depart from the spirit of Stalnaker's theory in giving a world w an important place in what is asserted, even if $w \notin c_w$. This is especially clear in HMDiagonalization. Even if $w \notin c_w$, $u(w)$ has an important effect on what can be asserted in that context. The following combination (together with Uniformity) seems more in line with Stalnaker's original:

Default: $\forall w\ (\forall v, x \in c_w)(a_u(v) = u(x)) \Rightarrow \forall x \in c_w(a_u(w) = u(x))$

Diagonalization: $\forall w\ (\exists v, x \in c_w)(a_u(v) \neq u(x)) \Rightarrow (a_u(w) = u^d)$

This pair of axioms yields odd consequences if there can be $v \in c_w$ such that both $v \notin c_v$ and $c_v \not\subseteq c_w$. But Stalnaker's S_3 theory of context rules out this possibility by definition, since it validates Positive CG-Introspection:

Positive CG-Introspection: $\Omega{=}CG(E) \rightarrow CGCG(E)$

We can now ask for what class of pointed conversational models M it is the case that for any $u \in U$, u satisfies these axioms. This question was first posed by Hawthorne and Magidor 2009, who answered it in a more restricted setting.

(In fact, they claim that one-step transitivity or symmetry violations are enough to lead to conflicts in these three principles, but this is not quite enough, since Uniformity, Default and Diagonalization really impose conditions on the transitive closure of the context-accessibility relation, as the following proof shows. Their argument on the basis of transitivity violations (383) is also incorrect: they claim that the presence of a world $w \in c_a$ where one diagonalizes is insufficient to force diagonal content to be asserted in c_a, but this contradicts their own formulation of Uniformity: any such world forces diagonal content to be asserted in c_a as well. This is made clear by the "sublemma" in the proof below.)

Fixing a conversational model, let R^* be the transitive closure of the union of the R^{A_i}, defined in Appendix 2.A.3. (We could equivalently define it as the transitive closure of R^{A_E}.) Then we show:

PROPOSITION 2.B.1. *Let M be a pointed conversational model.*

(A) All utterances u satisfy Uniformity, Default and Diagonalization if and only if (*) $\forall w(\forall v \in c_w)[R^*(w) = R^*(v)]$.

*(B) All utterances u satisfy Uniformity, HM-Default and HM-Diagonalization if and only if (*) $\forall w(\forall v \in c_w)[R^*(w) = R^*(v)]$ and **(T)** $(\forall w)(w \in R^*(w))$.*

PROOF. (A) *If*: If $(\forall v, x \in c_w)(u(v) = u(x))$, then for all such x, $a_u(w) = u(x)$, and Uniformity is trivially satisfied. If $a_u(w) = u^d$, then there's some $v, x \in c_w$ such that $a_u(v) \neq u(x)$. Since $R^*(v) = R^*(w)$, both v and x are in $R^*(v)$. We show that $a_u(v) = a_u(x) = u^d$ by way of a general lemma (which itself requires a sublemma).

The statement of the lemma is: If $\exists b, c \in R^*(z)(a_u(b) \neq a_u(c))$ then $a_u(z) = u^d$.

There are finite paths from d to b and d to c using R^{A_E}: we induct on the maximum of these two path lengths. For the base case: if $b, c \in R^{A_E}(z)$ then there are two cases. If both $a_u(b) = u(b)$ and $a_u(c) = u(c)$, then since $a_u(b) = u(b) \neq u(c) = a_u(c)$ $a_u(z) = u^d$. If either $a_u(b) \neq u(b)$ or $a_u(c) \neq u(c)$, then these single worlds fulfill the antecedent of Diagonalization, so again $a_u(z) = u^d$. Now we assume we have shown that if both b and c are reachable in at most n steps, then $a_u(z) = u^d$, and show the claim for $n + 1$ steps.

For this we need a sublemma: if $\exists f \in R^*(g)(a_u(f) = u^d)$ then $a_u(g) = u^d$. Again, we prove this by induction on path length. For the base case, if $u(f) \neq u^d$, then f fulfills the antecedent of Diagonalization, so $a_u(g) = u^d$. If $u(f) = u^d$, then the only way Default could be fulfilled is if all $h \in c_g$ have $u(h) = u^d$, but this would still make $a_u(g) = u^d$. Now we suppose the claim holds for worlds reachable in n steps and show it for $n + 1$. By hypothesis there's some n-reachable f' so that $f \in c_{f'}$, and $a_u(f) = u^d$. We proceed exactly as in the base case to show that $a_u(f') = u(d)$. Since f' is reachable in n steps, it follows by the induction hypothesis that $a_u(g) = u^d$.

Now we return to the induction of our lemma. We know there are b', c', reachable in at most n steps from z such that $b \in R^{A_E}(b')$ and $c \in R^{A_E}(c')$. Now there are two cases. If $a_u(b') = u(b') = a_u(b) = u(b)$ and $a_u(c') = u(c') = a_u(c) = u(c)$, then since $a_u(c) \neq a_u(b)$, and c', b' are reachable in n steps, we're done. And if for any $h \in \{b, b', c, c'\}$, $a_u(h) = u^d$, then the sublemma gives us the desired result, that $a_u(z) = u^d$.

With the lemma and the sublemma in hand, the main *If*: claim follows trivially. If $a_u(v) = u^d$ or $a_u(x) = u^d$, then we use the sublemma to show that both of them must be u^d. The supposition that both of them have $a_u(v) = u(v)$ and $a_u(x) = u(x)$ leads to contradiction using the main lemma. So Uniformity is satisfied.

Only if: Now suppose $(*)$ fails: there's a w, w' so that $w' \in c_w$ but $R^*(w) \neq R^*(w')$. This can only be because $R^*(w') \subsetneq R^*(w)$ (and so, $w \notin R^*(w')$). Then define a u so that for all $w'' \in R^*(w')$ $u(w') = u(w'')$, but for some $w''' \in R^*(w) \setminus R^*(w')$, $u(w''') \neq u(w')$. Then u^d is asserted at w, but at w', $u(w') \neq u^d$ is asserted, in contradiction of Uniformity.

(B) If: If $(\forall v \in c_w)(a_u(v) = u(w))$, then $a_u(w) = u(w)$, and Uniformity is trivially satisfied. If on the other hand $(\exists v \in c_w)(a_u(v) \neq u(w))$, so that $a_u(w) = u^d$, we now show

that $(\forall v \in c_w)(a_u(v) = u^d)$ as well. We do this with a version of the earlier "sublemma", proving by induction on path length in R^{A_E} that if $(\exists x \in R^*(v))(a_u(x) = u^d)$ then $a_u(v) = u^d$ as well. For the base case, if $u(v) = u^d$ already, then we're done. If not, then since we're assuming $x \in c_w$ HMDiagonalization immediately gives that $a_u(v) = u^d$. The induction step is similar. So since $a_u(w) = u^d$ and $w \in R^*(v)$, it follows that $a_u(v) = u^d$ as well.

Only if: If (*) fails, the argument is exactly as before.

Now suppose (T) fails: $w \notin R^*(w)$. Define a u so that for all $w', w'' \in R^*(w)$, $u(w') = u(w'')$, but for w itself, $u(w) \neq u(w')$. Then u^d is asserted at w, but at w', $u(w') \neq u^d$ is asserted, in contradiction of Uniformity. □

This result gives us another reason to prefer Default and Diagonalization as Stalnaker's intended axioms. For it is often claimed that the following two principles are sufficient to ensure that Stalnaker's theory is consistent:

Positive CG-Introspection: $\Omega{=}CG(E) \rightarrow CGCG(E)$

Negative CG-Introspection: $\Omega = \neg CG(E) \rightarrow CG\neg CG(E)$;

but if we use the HM principles, these are necessary but not sufficient to ensure that the theory is consistent.

In Hintikka-Kripke frames, Positive CG-Introspection ensures that $c_w = R^*(w)$. Negative CG-Introspection is needed to ensure that $(\forall v \in c_w)(c_v = c_w)$. As I've observed often enough Stalnaker's S_3-theory (though not, I repeat, S_1 and S_2) validates positive introspection by definition, Proposition 2.B.1 demonstrates that Negative CG-Introspection is the key to the question of whether his theory of assertion is consistent with his theory of the common ground. The next appendix studies this axiom in some detail.

2.C. Negative Introspection

To guarantee negative introspection, Stalnaker needs the following condition (the name for the condition is mine):

Genuinely Non-Defective: $a \in \cap_{i \in I}(A_i CAp \rightarrow CAp)$ and $a \in CA(\cap_{i \in I}(A_i CAp \rightarrow CAp))$[29]

[29]See next appendix for discussion of Stalnaker's own definition of non-defective contexts.

In Stalnaker models, this axiom delivers negative introspection for common ground (an analogous axiom delivers it for presupposition). I'll state and prove a more general version of the theorem which shows this, since the more general statement is illuminating.

PROPOSITION 2.C.1. *Let M be a Hintikka-Kripke conversational model. Then for any $n \in \mathbb{N}$, if the model satisfies*

(i) $a \in \neg A_i p \rightarrow A_i \neg A_i p$

(ii) $a \in A_i A_E^n p \rightarrow A_E^n p$

(iii) $a \in A_i(A_E^n p \rightarrow A_i A_E^n p)$

(ia) $a \in A_E^n(\neg A_i p \rightarrow A_i \neg A_i p)$

(iia) $a \in A_E^n(A_i A_E^n p \rightarrow A_E^n p)$ *and*

$(iiia)$ $a \in A_E^n(A_i(A_E^n p \rightarrow A_i A_E^n p))$ *then*

$a \in \neg A_E^n p \rightarrow A_E^n \neg A_E^n p$.

Stalnaker's version of the proposition replaces my A^n with CA. Even when we make these replacements, (i) and (ia) hold trivially in Stalnaker-models, since $\mathbf{5} = \Omega$. In Stalnaker models, all agents are negatively introspective, *and* it's commonly accepted that they are. Moreover, both (iii) and $(iiia)$ hold by definition when we replace A_E^n with CA. Finally, Stalnaker models are of course Hintikka-Kripke conversational models.

PROOF. It's a standard fact that any operator formed by iterating and conjoining normal modal operators is itself normal: that is, the versions of **K** and **C** for A_E^n also equal the universe Ω. We first show that (i),(ii),(iii) entail $a \in \neg A_E^n \rightarrow A_E \neg A_E^n p$. Assume $a \in \neg A_E^n p$. By contraposition of (ii) we have $a \in \neg A_i A_E^n p$, by (i) we have $a \in A_i \neg A_i A_E^n p$. By contraposition of (iii) and the fact that **K** holds for A_i at a, we have $a \in A_i \neg A_E^n p$. We repeat for each $i \in I$ to derive $A_E \neg A_n^E p$. Thus by conditional proof, we have (using (i), (ii), (iii) as abbreviations for the statements) $\Omega = (i) \wedge (ii) \wedge (iii) \rightarrow [\neg A_E^n p \rightarrow A_E \neg A_E^n p]$. Now we use RN to derive $a \in A_E^n([(i) \wedge (ii) \wedge (iii)] \rightarrow [\neg A_E^n p \rightarrow A_E \neg A_E^n p])$. We use (ia), (iia), $(iiia)$ and **C** to give us that $A_E^n[(i) \wedge (ii) \wedge (iii)]$. So then by K we have $(*)$ $A_E^n(\neg A_E^n p \rightarrow A_E \neg A_E^n p)$. But now suppose that $\neg A_E^n p$. By $(i), (ii), (iii)$ we have $A_E \neg A_E^n p$. and by $(*)$ and K, we have $A_E \neg A_E p \rightarrow A_E A_E \neg A_E p$, and so on up to n. □

The minimal theory requires fewer assumptions to prove the same conclusion, that negative introspection holds for common ground:

PROPOSITION 2.C.2. *(Minimal Negative Introspection) Let M be a pointed conversational model where each A_i obeys $a - K$ $(a \in A_i(p \rightarrow q) \rightarrow A_i p \rightarrow A_i q)$. Then if*

(i) $a \in \neg A_i p \rightarrow A_i \neg A_i p$

(ii) $a \in A_i A_E p \rightarrow A_E p$

(iii) $a \in A_i(A_E p \rightarrow A_i A_E p)$ *then*

$a \in \neg A_E p \rightarrow A_E \neg A_E p$.

PROOF. If $a \in \neg A_E p$, then by (ii), $a \in \neg A_i A_E p$. By (i) $a \in A_i \neg A_i A_E p$. By contraposition of (iii) and $a - K$ for A_i, $a \in A_i \neg A_E p$. Since i was chosen arbitrarily the claim holds for all i, and so $a \in A_E \neg A_E p$. □

This proposition drops three controversial (in my view, implausible) assumptions: (1) that agents have to accept (in fact commonly accept) that acceptance obeys negative introspection; (2) that agents' acceptances are closed under conjunction in the problematic direction: $AE \cap AF \rightarrow A(E \cap F)$; (3) that the context not just *be* non-defective, but it be commonly accepted that it is non-defective.

Stalnaker does not find these assumptions problematic, so this difference won't convince him. He believes that the introspection axioms are necessary facts about belief, and so, given his coarse grained conception of content, he is committed to the claim that we do commonly know these facts about belief. Moreover, he thinks belief does satisfy this closure principle, and, third, he may simply take it as a deliverance of his theory that genuine non-defectiveness requires common acceptance that the context is non-defective.

On its own, then, negative introspection leaves us at an impasse. Both Stalnaker and I require additional assumptions to ensure this introspection axiom. The assumptions needed on the minimal theory are weaker than those needed on Stalnaker's, but Stalnaker will not find the difference in strength conceptually important.

This brings us back to Positive CG-Introspection. Positive CG-Introspection is in fact a logical watershed between Stalnaker's theory and the minimal one. The reason has already been touched on in passing. If we add Positive CG-introspection to the minimal theory, then according to the new theory, if it's common ground that p, it is also S_3-common ground that p.[30] This holds for Positive CG-Introspection in any theory where common ground is defined by iterating mutual acceptance.

In light of Proposition 2.B.1 this difference might seem critical. If Stalnaker's theory of assertion is to be consistent for all the set of all utterances, we must have *both* positive and negative introspection.

[30]I should also note that this version of collapse is a particularly implausible theory. There are clearly cases where you and I both accept something, but we fail to mutually accept that both accept it (this is a point on which Stalnaker and I agree). But the theory under consideration in the main text would (bizarrely) rule out this commonplace phenomenon.

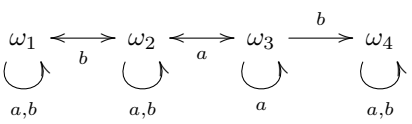
But I don't think the argument that the minimal theory fails here is a very strong one. The minimal theorist has two replies which do not require guaranteeing positive introspection in all contexts. The conservative option is to hold that Diagonalization was merely intended as a restricted principle governing some utterances in some contexts. The aim of the theory was surely not to claim that every utterance can always be interpreted! A less conservative option is to give up on Stalnaker's theory of assertion full stop. This option should not, I think, be overlooked. The theory of Diagonalization, Uniformity and Default has interesting structure, but I do not believe that its predictions are so well confirmed for its rejection to count as a clear cost to the minimal theory of common ground.

2.D. Non-Defective Contexts

Instead of "Genuine Non-Defective", Stalnaker originally defined the following notion of context:

Non-Defective: A context is non-defective if, for all i, if i S-presupposes that p, then it's S-common ground that p.

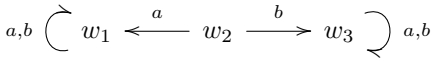
In early draft material for this paper, I provided countermodels to the following claim (Stalnaker 2009: 401) "If a speaker believes the context is nondefective, then that speaker's presuppositions will satisfy both negative and positive introspection." The construction of counterexamples to the claim is sufficiently delicate that I include an example here:



At ω_1, the context is non-defective in the defined sense, and both a and b believe it is. And yet, while it's not common ground that $\{\omega_4\}$, it's also not common ground that it's not common ground that $\{w_4\}$. (In fact, this example is simpler than my originals; thanks to Bob Stalnaker for showing me a related one, and also to Matt Mandelkern who caught some errors in an earlier draft.)

In response to these earlier counterexamples, Stalnaker now adopts a version of the condition which I called "Genuinely Non-Defective", and a variant for presupposition (it's presupposed that the context is non-defective), to ensure negative introspection for common ground and presupposition (2014: Appendix main text at n. 14). In fact, while the condition Stalnaker now provides for presupposition is correct, the one for

common ground is still not quite right: he says "if it is common ground that the context is nondefective, then the negative introspection principle will hold for the common ground." But in fact, the context must also *actually be* non-defective, as the following model shows. In the figure, at w_2 it is common ground that the context is non-defective (the common ground is $\{w_1, w_3\}$), and not common ground that $\{w_1\}$, but also not common ground that it is not common ground that $\{w_1\}$ (since at w_1, the common ground would be $\{w_1\}$).



In both cases, Stalnaker may have been led to his claims on the basis of a result by Bonanno and Nehring (2000), which shows that adding the *axiom* $B_iCBp \rightarrow CBp$ to a normal multi-modal doxastic logic where each individual's beliefs obey $KD45$ returns negative introspection for common belief. Their result depends on the fact that we can use Necessitation with respect to this axiom itself. In model-theoretic terms, it's not just that our beliefs about what we commonly believe are true *at the actual world*; they're true at every world in the model. As a result, for any world $B_iCBp \rightarrow CBp$ is both true at that world and a matter of common belief there. Proposition 2.C.1 is an easy generalization of their theorem, but the statement of it demonstrates the role Necessitation plays in the proof, by making explicit the beliefs about beliefs one must have for the proof to go through.

In any case, this point about definitions is a minor detail; the revised condition serves Stalnaker's theoretical purposes just as well as the one he adopts in the book.

2.E. Unique i-Candidate Contexts and Agent-Relative Introspection

In the main text I noted that it is important that the minimal theorist can define a class of contexts in which agents do know what the context is. This appendix fulfills that promise. We want the following two axioms.

Agent-Relative Positive CG-Introspection: $CGp \rightarrow B_iCGp$

Agent-Relative Negative CG-Introspection: $\neg CGp \rightarrow B_i\neg CGp$.[31]

[31]For reasons mentioned repeatedly above, replacing "B_i" in these conditions with A_i is unpromising, at least if one accepts that p only if one also adopts the attitude determined by the conversational tone to p. Note that in our general neighborhood frames, these axioms will not suffice, since $R^{B_i}(w)$ is defined for the version of belief which corresponds to assumption, not for the original belief operator (which corresponds to acceptance). We then have two options. We can either impose an axiom directly on B_i^C for each agent, or we can alter i-Uniformity to be defined as: $(\forall w)(\exists w')(a_u(w') \in N_i^B(w))$. It then suffices to have simply: $\forall w \exists v(\{w' : c_v = c_{w'}\} \in N_i^B(w)$. The former will be my approach in what follows in the main text.

In fact, we might as well replace the B_i with K_i. By essentially the same argument as used for Proposition 3, we then have

PROPOSITION 2.E.1. *(Contextual Omniscience) Let M be a pointed conversational model where each B_i obeys $a - K$ ($a \in B_i(p \to q) \to (B_ip \to B_iq)$). Then if the model obeys:*

(i) $a \in \neg A_ip \to B_i\neg A_ip$

(ii) $a \in B_iCGp \leftrightarrow CGp$

(iii) $a \in B_i(CGp \to A_iCGp)$ *then*

$a \in CGp \to B_iCGp \cap \neg CGp \to B_i\neg CGp$.

There is one slight difference between this proposition and Proposition 2.C.2. Condition (ii) has now been strengthened to a biconditional, delivering Agent-Relative Positive Introspection by stipulation.

I take this result to show that, as far as the intuitive demand about knowing what the context is, there is little to choose between the two theories. Given the appropriate additional stipulations, each can deliver the needed form of omniscience about context. The case of Stalnaker's theory of assertion is a subtler question. It is unsurprising that Stalnaker's theory of context is tailored to that theory of assertion. But it is unclear whether that amounts to an argument, never mind a powerful one, against the minimal theory of common ground.

Bibliography

Aarnio, Maria Lasonen. 2010. Unreasonable knowledge. *Philosophical Perspectives*, **24**(1), 1–21. 123

Aaronson, Scott. 2005. The Complexity of Agreement. *Symposium on the Theory of Computing*, Extended Abstract. Full Version at www.scottaaronson.com/papers/agree–econ.pdf. 96

Akkoyunlu, Eralp A., Ekanadham, Kattamuri, & Huber, R. V. 1975. Some Constraints and Tradeoffs in the Design of Network Communications. *SIGOPS Oper. Syst. Rev.*, **9**(5), 67–74. 13

Almotahari, Mahrad, & Glick, Ephraim. 2010. Context, content, and epistemic transparency. *Mind*, **119**(476), 1067–1086. 65

Arad, Ayala, & Rubinstein, Ariel. 2012. The 11-20 Money Request Game: A Level-k Reasoning Study. *American Economic Review*, **102**(7), 3561–73. 17

Aumann, Robert J. 1976. Agreeing to Disagree. *The Annals of Statistics*, **4**(6), 1236–1239. 18, 95, 102, 140, 141, 145

Aumann, Robert J. 1985. On the non-transferable utility value: A comment on the Roth-Shafer examples. *Econometrica: Journal of the Econometric Society*, 667–677. 27

Aumann, Robert J. 1987. Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica*, **55**(1), 1–18. 137

Aumann, Robert J. 1998. Common Priors: A Reply to Gul. *Econometrica*, **66**(4), 929–938. 115

Aumann, Robert J., & Brandenburger, Adam. 1995. Epistemic Conditions for Nash Equilibrium. *Econometrica*, **63**(5), 1161–1180. 25, 149, 154, 156, 159

Aumann, Robert J., & Hart, S. 2006 (unpublished). *Agreeing on decisions*. The Hebrew University of Jerusalem, Center for the Study of Rationality. 109

Aumann, Robert J, Hart, Sergiu, & Perry, Motty. 2005 (unpublished). *Conditioning and the sure-thing principle*. Center for the Study of Rationality. 108, 109

Bach, Christian W., & Tsakas, Elias. 2014. Pairwise epistemic conditions for Nash equilibrium. *Games and Economic Behavior*, **85**(0), 48 – 59. 25, 139, 149, 157, 160

Bacharach, Michael. 1985. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, **37**(1), 167 – 190. 103, 105, 108, 137, 145, 148

Barelli, Paolo. 2009. Consistency of Beliefs and Epistemic Conditions for Nash and Correlated Equilibria. *Games and Economic Behavior*, **67**, 363–375. 25, 113, 149, 151, 160

Barwise, Jon. 1988. Three views of common knowledge. *Pages 365–379 of: Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*. Morgan Kaufmann Publishers Inc. 42

Battigalli, Pierpaolo, Brandenburger, Adam, Friedenberg, Amanda, & Siniscalchi, Marciano. 2014. *Strategic Uncertainty: The Epistemic Approach to Game Theory*. 25, 156

Bernheim, B. Douglas. 1984. Rationalizable Strategic Behavior. *Econometrica*, **52**(4), 1007–1028. 27, 35

Binmore, Ken, & Samuelson, Larry. 2001. Coordinated action in the electronic mail game. *Games and Economic Behavior*, **35**(1), 6–30. 25

Blackburn, Patrick, de Rijke, Maarten, & Venema, Yde. 2001. *Modal Logic*. Cambridge University Press. 163

Bonanno, Giacomo, & Nehring, Klaus. 1997 (September). *Agreeing to Disagree: A Survey*. http://www.econ.ucdavis.edu/faculty/bonanno/PDF/agree.pdf. 113

Bonanno, Giacomo, & Nehring, Klaus. 1999. How to make sense of the common prior assumption under incomplete information. *International Journal of Game Theory*, **28**(3), 409–434. 113, 137, 141, 148

Bonanno, Giacomo, & Nehring, Klaus. 2000. Common Belief with the Logic of Individual Belief. *Mathematical Logic Quarterly*, **46**(1), 49–52. 93

Brandenburger, Adam, & Dekel, Eddie. 1987. Rationalizability and correlated equilibria. *Econometrica*, **55**(6), 1391–1402. 27, 35, 156

Brandenburger, Adam, & Keisler, Jerome H. 2006. An Impossibility Theorem on Beliefs in Games. *Studia Logica*, **84**, 211–240. 83

Brandenburger, Adam, Dekel, Eddie, & Geanakoplos, John. 1992. Correlated Equilibrium with Generalized Information Structures. *Games and Economic Behavior*, **4**, 182–201. 137, 139, 143, 152, 153, 154, 163

Briggs, Rachael. 2009. Distorted Reflection. *The Philosophical Review*, **118**(1), 59–85. 121

Camerer, Colin. 2003. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press. 22

Cave, Jonathan A.K. 1983. Learning to agree. *Economics Letters*, **12**(2), 147 – 152. 103, 105, 137

Chwe, Michael Suk-Young. 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton University Press. 36

Clark, Herbert H. 1996. *Using language*. Cambridge University Press. 44

Clark, HH, & Marshall, CR. 1981. Definite reference and mutual knowledge. 10–64. 44

Cohen, Stewart, & Comesaña, Juan. 2013. Williamson on Gettier Cases and Epistemic Logic. *Inquiry*, **56**(1), 15–29. 78

Collins, John. 1997. *How We Can Agree to Disagree*. http://collins.philo.columbia.edu/disagree.pdf. 121

Conee, Earl, & Feldman, Richard. 2004. *Evidentialism: Essays in Epistemology*. Oxford University Press. 103

Dekel, Eddie, & Siniscalchi, Marciano. 2014 (14 January). *Epistemic Game Theory*. 25, 156

Dekel, Eddie, Lipman, Bart L, & Rustichini, Aldo. 1998. Standard state-space models preclude unawareness. *Econometrica*, **66**(1), 159–73. 7, 128, 137

Easwaran, Kenny. 2013. Regularity and Infinitesimals. *Philosophical Review*. 97

Elga, Adam. 2007. Reflection and Disagreement. *Noûs*, **41**(3), 478–502. 121

Epstein, Larry G, & Wang, Tan. 1996. " Beliefs about Beliefs" without Probabilities. *Econometrica: Journal of the Econometric Society*, 1343–1373. 135

Fagin, Ronald, Halpern, Joseph Y., Moses, Yoram, & Vardi, Moshe Y. 1995. *Reasoning about Knowledge*. MIT Press. 13, 18, 100

Feinberg, Yossi. 2000. Characterizing Common Priors in the Form of Posteriors. *Journal of Economic Theory*, **91**(2), 127 – 179. 113, 141

Feldman, Richard. 2007. Reasonable Religious Disagreements. *Chap. 17, pages 194–214 of:* Antony, Louise (ed), *Philosophers without Gods: Meditations on Atheism and the Secular*. Oxford University Press. 97

Friedell, Morris F. 1969. On the structure of shared awareness. *Behavioral Science*, **14**(1), 28–39. 18, 102, 141

Geanakoplos, John. 1989 (May). *Game Theory Without Partitions, and Applications to Speculation and Consensus*. Cowles Foundation Discussion Papers 914. Cowles Foundation for Research in Economics, Yale University. 7, 134, 137, 139, 141, 142, 148, 161

Geanakoplos, John. 1994. Common Knowledge. *Chap. 40, pages 1437–1496 of:* Aumann, Robert, & Hart, Sergiu (eds), *Handbook of Game Theory with Economic Applications*, vol. 2. North Holland. 113

Gilbert, Margaret. 1989. *On social facts*. Princeton University Press. 41

Goodman, Jeremy. 2013. Inexact knowledge without improbable knowing. *Inquiry*, **56**(1), 30–53. 30

Gray, Jim. 1978. Notes on Data Base Operating Systems. *Pages 393–481 of: Operating Systems, An Advanced Course*. London, UK: Springer-Verlag. 13

Greco, Daniel. 2014. Iteration and Fragmentation. *Philosophy and Phenomenological Research*, n/a–n/a. 10, 78

Greco, Daniel. 2014 (forthcoming). Could KK be OK? *Journal of Philosophy*, n/a–n/a. 10, 119

Grimm, V., & Mengel, F. 2012. An experiment on learning in a multiple games environment. *Journal of Economic Theory*, **147**(6), 2220–2259. 28

Groenendijk, Jeroen, & Stokhof, Martin. 1991. Dynamic predicate logic. *Linguistics and Philosophy*, **14**(1), 39–100. 44

Hájek, Alan. 2010. Staying regular. *Unpublished manuscript*, **43**, 133. 97

Halpern, Joseph Y, & Pucella, Riccardo. 2011. Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial intelligence*, **175**(1), 220–235. 102

Hanson, Robin. 2006. Uncommon Priors Require Origin Disputes. *Theory and Decision*, **61**(4), 318–328. http://hanson.gmu.edu/prior.pdf. 113

Hawthorne, John, & Magidor, Ofra. 2009. Assertion, Context, and Epistemic Accessibility. *Mind*, **118**(470), 377–397. 65, 86, 87

Hawthorne, John, & Magidor, Ofra. 2011. Assertion and Epistemic Opacity. *Mind*, **119**(476), 1087–1105. 65

Heal, Jane. 1978. Common knowledge. *The Philosophical Quarterly*, **28**(111), 116–131. 15, 41

Heifetz, Aviad. 1996. Comment on Consensus without Common Knowledge. *Journal of Economic Theory*, **70**(1), 273–77. 113

Heim, Irene. 1983. On the projection problem for presuppositions. *Pages 114–125 of: West Coast Conference in Linguistics*, vol. 2. 44

Heinemann, Frank, Nagel, Rosemarie, & Ockenfels, Peter. 2004. The Theory of Global Games on Test: Experimental Analysis of Coordination Games with Public and Private Information. *Econometrica*, **72**(5), 1583–1599. 22

Hellman, Ziv. 2012 (June). *Deludedly Agreeing to Agree*. 137, 149, 161

Hintikka, Jaako. 1962. *Knowledge and Belief*. Cornell University Press. 100, 106

Kajii, Atsushi, & Morris, Stephen. 1997. Common p-Belief: The General Case. *Games and Economic Behavior*, **18**, 73–82. 141

Kamp, Hans, Van Genabith, Josef, & Reyle, Uwe. 2011. Discourse representation theory. *Pages 125–394 of: Handbook of Philosophical Logic*. Springer. 44

Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical linguistics*, **1**(1), 181–194. 43

Kelly, Thomas. 2013. How to be an Epistemic Permissivist. *In:* Greco, John, Steup, Matthias, & Turri, John (eds), *Contemporary Debates in Epistemology*. Blackwell Publishing Ltd. 97

Kinderman, Peter, Dunbar, Robin, & Bentall, Richard P. 1998. Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, **89**(2), 191–204. 45

Kolodny, Niko, & MacFarlane, John. 2010. Ifs and oughts. *The Journal of Philosophy*, **107**(3), 115–143. 108

Kripke, Saul A. 1979. A Puzzle about Belief. *Pages 239–283 of:* Margalit, Avishai (ed), *Meaning and Use*. Dordrecht. 42

Lewis, David. 1975. Languages and language. *Minnesota Studies in the Philosophy of Science*. 66

Lewis, David K. 1969. *Convention: A philosophical study*. Harvard University Press. 18, 102, 141

Lin, Shuhong, Keysar, Boaz, & Epley, Nicholas. 2010. Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, **46**(3), 551–556. 45

Liu, Qinming. 2010 (unpublished). *Higher-Order Beliefs and Epistemic Conditions for Nash Equilibrium*. 25, 156

McGee, Vann. 1991. We turing machines aren't expected-utility maximizers (even ideally). *Philosophical Studies*, **64**(1), 115–123. 137

Mengel, Friederike. 2012. Learning across games. *Games and Economic Behavior*, **74**(2), 601 – 619. 28

Monderer, Dov, & Samet, Dov. 1989. Approximating Common Knowledge with Common Beliefs. *Games and Economic Behavior*, **1**(2), 170–190. 18, 102, 137, 141

Morris, Stephen. 1995. The Common Prior Assumption in Economic Theory. *Economics and Philosophy*, **11**(2), 227–253. 96

Morris, Stephen. 1996. The logic of belief and belief change: A decision theoretic approach. *Journal of Economic Theory*, **69**(1), 1–23. 135

Moses, Yoram, & Nachum, Gal. 1990. Agreeing to disagree after all. *Pages 151–168 of: Proceedings of the 3rd conference on Theoretical Aspects of Reasoning and Knowledge*. 97, 109

Nagel, Rosemarie. 1995. Unraveling in Guessing Games: An Experimental Study. *American Economic Review*, **85**(5), 1313–26. 17

Parfit, Derek. 1988 (1983). *What we together do*. 108

Peacocke, Christopher. 2005. *Joint attention: Its nature, reflexivity, and relation to common knowledge*. Oxford: Oxford University Press. 41, 42

Pearce, David G. 1984. Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica*, **52**(4), 1029–1050. 27, 35

Pinker, Steven, Nowak, Martin A., & Lee, James J. 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences*, **105**(3), 833–838. 37

Polak, Ben. 1999. Epistemic Conditions for Nash Equilibrium and Common Knowledge of Rationality. *Econometrica*, **67**, 673–676. 25, 160

Portner, Paul. 2007. Imperatives and modals. *Natural Language Semantics*, **15**(4), 351–383. 44

Pryor, James. 2000. The Skeptic and the Dogmatist. *Noûs*, **34**(4), 517–549. 122

Rubinstein, Ariel. 1989. The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge". *The American Economic Review*, **79**(3), 385–391. 21, 22, 24

Rubinstein, Ariel, & Wolinsky, Asher. 1990. On the logic of 'agreeing to disagree' type results. *Journal of Economic Theory*, **51**(1), 184 – 193. 107, 137, 148

Samet, Dov. 1990. Ignoring ignorance and agreeing to disagree. *Journal of Economic Theory*, **52**(1), 190–207. 137, 145, 148

Samet, Dov. 1992. Agreeing to Disagree in Infinite Information Structures. *International Journal of Game Theory*, **21**(3), 213–218. 145

Samet, Dov. 2010. Agreeing to disagree: The non-probabilistic case. *Games and Economic Behavior*, **69**(1), 169 – 174. 108, 109

Savage, Leonard J. 1954. *The Foundations of Statistics*. John Wiley and Sons. 108, 123

Savage, Leonard J. 1967. Difficulties in the Theory of Personal Probability. *Philosophy of Science*, **34**(4), 305–10. 135

Schiffer, Stephen R. 1972. *Meaning*. Oxford: Oxford University Press. 11

Shin, Hyun Song. 1993. Logical structure of common knowledge. *Journal of Economic Theory*, **60**(1), 1–13. 137, 148

Shin, Hyun Song, & Williamson, Timothy. 1994. Representing the knowledge of turing machines. *Theory and Decision*, **37**(1), 125–146. 137

Srinivasan, Amia. 2013. Are We Luminous? *Philosophy and Phenomenological Research*, n/a–n/a. 39

Stahl, Dale, & Wilson, Paul W. 1994. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, **25**(3), 309–327. 17

Stahl, Dale, & Wilson, Paul W. 1995. On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior*, **10**(1), 218–254. 17

Stalnaker, Robert. 1998. On the representation of context. *Journal of Logic, Language and Information*, **7**(1), 3–19. 43, 53

Stalnaker, Robert. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press. 86, 102

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy*, **25**(5), 701–721. 43, 53, 57

Stalnaker, Robert. 2006. On Logics of Knowledge and Belief. *Philosophical Studies*, **128**, 169–199. 31

Stalnaker, Robert C. 1974. Pragmatic presuppositions. New York: New York University Press. 43, 57

Stalnaker, Robert C. 1978. Assertion. *Pages 315–32 of:* Cole, P. (ed), *Syntax and Semantics*, vol. 9. New York Academic Press. 43, 86

Stalnaker, Robert C. 2009. On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility. *Mind*, **118**(470), 399–409. 12, 30, 65, 86, 92, 106

Stalnaker, Robert C. 2014. *Context*. Oxford: Oxford University Press. 43, 53, 57, 60, 92

Stiller, James, & Dunbar, Robin IM. 2007. Perspective-taking and memory capacity predict social network size. *Social Networks*, **29**(1), 93–104. 45

Strzalecki, Tomasz. 2010. *Depth of reasoning and higher order beliefs.* Harvard Institute of Economic Research, Harvard University. 22, 26

Van Eijck, Jan, & Kamp, Hans. 2011. *Representing discourse in context.* Elsevier. Chap. 3, pages 181–252. 44

Van Eijck, Jan, & Visser, Albert. 2012. Dynamic Semantics. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, winter 2012 edn. 44

van Fraassen, Bas. 1984. Belief and the Will. *Journal of Philosophy*, **81**, 235–56. 121

Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic*, **25**(3), 221–261. 44

von Fintel, Kai, & Gillies, Anthony. 2011. *Might Made Right.* Oxford University Press. 66

White, Roger. 2005. Epistemic Permissiveness. *Philosophical Perspectives*, **19**. 97

White, Roger. 2013. *In:* Greco, John, Steup, Matthias, & Turri, John (eds), *Contemporary Debates in Epistemology.* Blackwell Publishing Ltd. 97

Williamson, Timothy. 1992. Inexact knowledge. *Mind*, **101**(402), 217–242. 117

Williamson, Timothy. 2000. *Knowledge and its Limits.* Oxford: Oxford University Press. 31, 38, 78, 98, 102, 103, 106, 110, 117, 119

Williamson, Timothy. 2007. How probable is an infinite sequence of heads? *Analysis*, **67**(3), 173–180. 97

Williamson, Timothy. 2013. Response to Cohen, Comesaña, Goodman, Nagel, and Weatherson on Gettier Cases in Epistemic Logic. *Inquiry*, **56**(1), 77–96. 76

Yalcin, Seth. 2007. Epistemic modals. *Mind*, **116**(464), 983–1026. 44, 48, 53, 57