Carnegie Mellon University
Research Showcase @ CMU

Department of Philosophy | Dietrich College of Humanities and Social Sciences

2004

Epistemological Foundations for the Representation of Discourse Context

Horacio Arlo-Costa
Carnegie Mellon University, hcosta@andrew.cmu.edu

Follow this and additional works at: http://repository.cmu.edu/philosophy

Part of the Philosophy Commons

Published In

.

This Conference Proceeding is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Philosophy by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Epistemological foundations for the representation of discourse context

Horacio Arló Costa

February 17, 2004

The first four sections of the paper focus on characterizing the elements that enter into the characterization of the notion of discourse context. One way of doing so is by identifying this notion with the set of commonly presupposed items. I propose a multi-agent account of context where it is essential to represent what each agent takes as being commonly presupposed, aside from what is commonly presupposed. The account requires adding dynamic features to context, in terms of the capacities of supposing that something is the case, given a current context; or in terms of the capacity of updating a context with new information. These dynamic features figure prominently in the proposed characterization.

The final sections of the paper focus on the inferential role played by the doxastic commitments induced by discourse context. I argue that these commitments do play a crucial role in understanding how agents reason defeasibly from the point of view of a given context. I discuss also some of the existing accounts, in terms of autoepistemic operators. I argue that they cannot provide a good encoding for conversational implicatures of the type Grice studied. The article offers instead an alternative account of autoepistemic inference based on an insight presented by Paul Grice in a recent addendum to his seminal article 'Logic and Conversation'.

1 Introduction

Utterances are evaluated in context. Suppose someone utters 'The current president lost the popular vote by almost half a million votes'. It is clear that the evaluation of such utterance depends on who is talking, as well as on the time and place of the utterance. If the speech act occurs during the year 2001 in the Unites States of America, the speaker does express a true proposition in uttering this sentence. For other places or times of utterance it is even dubious that the speaker manages to express a proposition in uttering this sentence.

The context of utterance contains also background information related to the utterance. For example, in this case it might contain: 'George Bush lost the popular vote by almost half a million votes'. Examples abound. A classic one is

'Nixon is guilty too'. An appropriate element of the context for this utterance might be 'Haldeman is guilty', and so on.

Of course, the notion of evaluation can be variously articulated. As it is clearly pointed out in [15], there are various conceptions of the way in which evaluations of utterances depend on context. Evaluation can be understood in terms of the identification of the proposition expressed (if any) by the speaker. But evaluation can also be articulated more liberally in terms of *assertibility* conditions of the sentences used by the speaker in his utterance. And assertibility can, in turn, be a matter of probability, acceptability, justifiability or truth. We can remain neutral here regarding this issue. What concerns us is the problem of how to understand what a context of utterance is.

There is a fair amount of consensus regarding the fact that utterances are evaluated in context. There is less consensus as to what are the propositional (or sentential) components of context. Some authors propose *context sets* , which basically are non-empty sets of possible worlds satisfying certain rationality conditions [47], [50]. This view is sometimes accompanied (but not always) by an epistemic construal of context, according to which the context set encodes the set of shared assumptions of a relevant set of agents participating in an information exchange - which, in turn, can be a simple conversation or a more sophisticated process guided by a protocol.

Other authors strongly disagree. Hans Kamp, for example, points out in [21] that 'theories of belief that identify belief with sets of possible worlds cannot differentiate them finely enough to do justice to our common understanding and use of the notion. As it is most often put: Possible worlds accounts entail that belief sentences - i.e. sentences of the form 'A believes that s' - are truth-invariant under substitution of necessarily equivalents for the embedded clause s, and this does not appear to be the way in which belief sentences are actually understood.'[1] Most of the criticism comes from cognitive theories of discourse processing. Some of these theories - like *dynamics semantics*[8] , or Discourse Representation Theory (DRS) [20] - focus on describing the form-meaning relation as an idealized account of the process whereby the recipient of an utterance comes to grasp the thoughts that the utterance contains. Therefore these theories privilege the study of the process by which cognitive subjects construct internal representations from syntactically specified inputs. This process is seen as an operation which always applies to a pair of structures - on the one hand the syntactic description of the sentence under consideration and on the other the internal representation, which is also syntactically specified. This internal representation functions as the *context* in which the new sentence is being interpreted. This tension - between syntactic and propositional accounts - is just one among the many stumbling blocks impeding the construction of a unified account of context.

Most of the standard theories of the common ground postulate that the right account of context in discourse should focus on an abstract representation of

[1] Defenders of the possible worlds account have tried to articulate a response by producing a two-dimensional construction of the so-called notion of *narrow belief*. Nevertheless this response is seldom incorporated in the standard description of context sets.

the presuppositions shared in common by a set of interacting agents. Moreover, the theory proposes to include in the representation not only what the agents commonly presuppose but also a representation of what they *are committed* to commonly presuppose (including all logical consequences of actual presuppositions).

I shall argue in the last section of this article that this normative account is of interest in order to capture some inferential aspects of context, namely defeasible inferences from commonly held attitudes, or defeasible inferences from the agent's take on the contents of commonly held attitudes. I shall use some insights first proposed by Paul Grice in order to differentiate these inferences from other inferential phenomena, like implicature. Nevertheless, the first part of the article will focus not on what agents are committed to presuppose, or on the agent's takes on those commitments, but on the representation of the *cognitive performances* of agents.

This presupposes an underlying distinction between normative and descriptive ideals. One possible strategy is to introduce the notion of context as a representation of the *doxastic commitments* of the speaker and hearer while a discourse unfolds. Another is to introduce contexts as a description of the *actual performance* of the agents involved in an information exchange.

I shall argue in favor of an epistemic account of discourse context. The received epistemic view proposes that contexts are constituted by the shared assumptions (presuppositions) of interacting agents. The body of shared assumptions plays, without doubt, a crucial role in understanding how utterances are evaluated in context; but an analysis of Scott Soames's characterization of presupposition [45] will suggest a different view, where shared assumptions are just one epiphenomenal result of the interplay of various other relevant epistemic parameters for the interacting agents. The proposed account is conceptually related to some multi-agent constructions of context common in Artificial Intelligence [11].

This alternative view crucially requires the development of a more sophisticated conception of how contexts are updated. Most of the existing accounts (syntactic or propositional) tacitly assume a very limited view of context change. Perrault [41] has recently characterized this view in terms of what he calls the *persistence theory of belief*. We will argue that such account of context change is inadequate for many applications, and we will offer an alternative based on a proposal first presented in [52]. The consideration of various examples will suggest that an adequate account needs to be able to explain *how* to effect realistic transitions among contexts in the presence of given inputs. More importantly, we will also argue that an adequate theory of context requires an account of *what* inputs are acceptable when various stimuli are admissible. As a result we propose to incorporate the *value of information* [29] as one of the crucial components of epistemic context . The view of context that thus arises offers a parametrical account, of which both descriptive and normative accounts of context can be obtained as special cases.

2 The common ground as a context set

In one of his most recent articles on the representation of discourse context in [50] Robert Stalnaker proposes the following model:

> I propose to identify a context (at a particular point in a discourse) with the body of information that is presumed, at that point, to be common to the participants in the discourse.
>
> [...]
>
> We can represent the information that defines the context in which a speech act takes place with a set of possible situations or possible worlds - the situations that are compatible with the information.

Stalnaker calls these sets of possible worlds *context sets* . Two additional constraints are imposed on them. Context sets are supposed to be non-empty sets of possible worlds, and the actual world need not be among them. We elaborate below on the meaning of these constraints.

Representing *doxastic (or epistemic) context* as a set of possible worlds is a usual strategy in formal epistemology. So, if W is the set of primitive points in the representation[2] a non-empty set $K_X \subseteq W$ can be seen as representing the strongest proposition to which some agent X (or a group of agents) bear some attitude A. If the attitude in question is, for example, belief, one can say that a proposition p is believed by agent X as long as $K_X \subseteq p$.[3]

The issue of what is the attitude that should be utilized in the characterization of discourse context is quite important. As is indicated in [50] the attitude in question seems to be *presupposition.* In fact, in [50] Stalnaker defends a variant of the usual theory according to which presuppositions are propositions presumed to be common ground. This seems to indicate that Stalnaker proposes the characterization of discourse context in terms of the propositions that are commonly presupposed.

Notice that this proposal need not be interpreted as saying that a speaker's presupposition is just what he takes as common belief in a group.[4] This interpretation has the unpalatable consequence of entailing that a speaker must believe whatever he presupposes. This can be blocked in various ways. One of them is by insisting that a speaker's presuppositions are determined in terms of what he or she *presumes* to be common belief. This is, after all, what is literally said by Stalnaker characterization. This strategy, nevertheless, requires

[2]The set W can be understood as a set of possible worlds or situations, or a set of points in some structure, like a sigma-field or an algebra, depending on the underlying assumptions about the nature and theoretical role of W. We will not go deeper into this issue here, although the theoretical decisions as to the nature and theoretical use of W are far from being idle. For the moment we will assume that the points in W as unstructured primitives, and we will not assign to them any intended interpretation - ontological or epistemic.

[3]From now on upper case letters D, E, ..., will stand for sentences and the corresponding d, e, ..., will stand for the propositions expressed by those sentences.

[4]Stalnaker's analysis begins with this simplifying assumption by identifying the common ground with common belief and characterizing common belief a la Schiffer [42].

the postulation of a third attitude (presumption) for which one does not have a separate account. Another strategy is the one utilized in [50] namely to invoke a third attitude that is better understood in epistemological terms, namely *acceptance*. According to Stalnaker acceptance is:

> a category if propositional attitude and methodological stances towards a proposition, a category that includes belief, but also some attitudes (presumption, assumption, acceptance for the purposes of argument or inquiry) that contrast with belief and with each other. To accept a proposition is to treat it as true for some reason.

So the idea now is to understand common ground as common belief about what is accepted. Or one can adopt a more radical stance by characterizing speaker presuppositions in terms of what the speaker accepts as being commonly accepted by the group for some communicative purposes. Any of these articulations of common ground will not entail that a speaker must believe whatever he presupposes.[5]

In addition any of these permissible articulations will identify discourse context with a structure that keeps track of the common presuppositions of the participants in the discourse. Moreover, the information that defines context is represented via a context set and relatively strong rationality constraints are imposed on the representation. For example, context sets cannot be empty, which imposes by fiat the assumption that the common ground *must* be coherent (i.e. logically consistent). Moreover, sccording to Stalnaker : 'agents know what they are presupposing, so they presuppose that they are presupposing p if they are, and that they are not if they are not.'

These rationality constraints might make sense if we are engaged in representing what the participants in a discourse ought to presuppose, but they are less evident as descriptive constraints on representing context. Even the consistency condition (context sets cannot be empty) has been challenged by recent work in semantics and computational linguistics where the idea is to represent how sentences are evaluated with respect to a common ground capable of tolerating local inconsistencies (which do not trivialize the representation) [26].

As we remarked above, the use of context sets seems to borrow heavily form previous normative work in theories of rationality. One of the main goals of these theories is to construct representations of the rational commitments of agents at each instant. Much of the current work in computational linguistics (and related fields in cognitive science and artificial intelligence) seems to focus instead on the study of the cognitive representations which agents construct in response to verbal inputs. The main idea of some of these accounts (DRT, for example [20]) is to describe the form-meaning relation as an idealized account of the process whereby the recipient of an utterance comes to grasp the thoughts that utterances contain.

[5] Of course, all these alternative strategies can still use variants of Schiffer's characterization of common belief – common acceptance, for example, can be defined by utilizing the same Schiffer-style construction that is usually run for belief.

Even when some of the current cognitive theories appeal to idealizations, their main task seems to be primarily descriptive. Representation of context in these theories is done via finite syntactic representations. For example, in the theory presented in [20] a message carried by an incoming sentence triggers the construction of a discourse representation. This representation is sensible to syntactic considerations. Therefore two sentences carrying identical content, but distinguishable under a syntactic point of view, might produce different posterior representations (even when the prior representation coincide). For example, consider the following two sentences:

P. Exactly one of the ten balls is not in the bag.

Q. Exactly nine of the ten balls are in the bag.

P and Q carry identical content (they are true in the same set of possible worlds), but they differ syntactically. Call C the prior context and C_P and C_Q the posterior contexts after updating them with P and Q respectively. In a theory like the one presented in [20] C_P does not coincide with C_Q.[6] Hans Kamp has recently used this example (due originally to Barbara Partee) in order to criticize the idea of representing contexts via sets of possible worlds.

The argument runs as follows. If assertion of a sentence is understood as 'a proposal to alter the context by adding the information that is the content of the assertion to the body of information that defines the context set' [50], then C_P and C_Q should coincide. At least this is so if one accepts a principle stating that a prior context set modified by identical propositions yield a unique posterior context set. This is a reasonable principle, even when the proposition corresponding to P (Q) has an empty intersection with the prior context set C - and therefore the update in question requires a 'genuine' revision of C. But now it seems that we arrived to an unpalatable conclusion, given that the context set induced by the ulterior utterance of the following sentence:

S. It is under the sofa

seems to lead to posterior contexts $C_{P,S}$ and $C_{Q,S}$ which cannot receive the same interpretation.[7] Nevertheless, if C_P and C_Q coincide, $C_{P,S}$ and $C_{Q,S}$ should coincide also.

Stalnaker has responded to this argument in [50]. The gist of the rebuttal is to propose that incoming sentences update the context as a result of a two-step process. The first part of this process invokes Gricean priciples of conversation. The assertion of a sentence is, of course, a speech act, and Stalnaker argues that the mere occurrence of such act changes the context, before the agent engages

[6] Basically the context for C_P has a reference marker for an individual ball while C_Q lacks it.

[7] The DRT representation articulates this asymmetry as follows: since the context for C_P has a reference marker for an individual ball, while C_Q lacks it, the construction algorithm implemented by DRT recommends to take *it* as referring back to the bag when processing Q and to an individual ball when processing P.

in the epistemic process of changing view. Commenting on Kamp's argument against representing contexts via context sets , Stalnaker argues as follows:

> Since it is a manifestly observable fact that, in each case, a certain sentence was uttered, this fact, together with any additional information that follows from that fact, conjoined with standing background information about linguistic and speech conventions, is available to distinguish ... [C_P and C_Q] ...

What Stalnaker offers here is a two-layered theory of update. Given a prior context C and an incoming proposition expressed by a sentence S, C_S is not just the update of C by S, which can be straightforwardly calculated when $C \cap s \neq \emptyset$, by eliminating from C all worlds where the proposition s (correposnding to S) is false.[8] The mere utterance of S might alter C yielding a modified prior context C'. Any posterior is then calculated with respect to this modified prior.

Stalnaker's idea is to abstract away from any theory of English or from any theory of the practice of speaking English, which nevertheless can be part of the background information true in all possible worlds of the context set. Whatever the nature of the facts invoked by cognitive theories in order the explain how agents construct data structures in the presence of verbal inputs, these facts - Stalnaker argues - will be taken into account in order to distinguish the possible worlds that define the universe of possibilia from which contexts are made. Stalnaker seems to be right in pointing out that proceeding in this way does not commit him to add nothing new to the possible worlds that define context - according to his characterization in terms of context sets. When an assertion is made the set of possible worlds compatible with the relevant linguistic information might be updated by the mere fact that the speech act occurred. And this updated context set can be further updated by adding to it the content of the speech act. It should be pointed out, nevertheless, that as long as the representation of context lacks the capacity of representing pragmatic changes of the type invoked by Stalnaker, one needs an account of context change where *updates with logically equivalent sentences might not produce the same output*. We will elaborate on this dynamic aspect of context below.

Even if one assumes that the right characterization of context should proceed epistemically by specifying the body of shared assumptions of speaker and hearer, further distinctions are needed in order to narrow down the set of possible theories compatible with that view. The set of shared assumptions could just be the set of shared assumptions that speaker and hearer consciously share at same stage of a dialogue, or it could be the set of assumptions that they are committed to share (even when they are not aware of some of these commitments). The notion of commitment can be here loosely constructed in terms of the epistemic obligations of reflective and rational agents with access to the information exchanged in a dialogue. The following example can help to clarify

[8] More about the universality of this Bayesian proposal for update below.

the distinction just made. You call the cab company and ask for the number of the cab that will pick you up in the next five minutes. The dispatcher tells you:

D. The number of the cab is the number 1729.

The posterior context set C'_D encodes both the pragmatic and epistemic consequences of the utterance. First the prior context might shift to a modified context C', as a consequence of the mere fact that the speech act itself takes place. Second C' is updated by d. It is reasonable to suppose that in this simple case the posterior context set C'_D entails the proposition expressed by d itself. Notice then that an agent whose presuppositions are described by the context set C'_D should be described *also* as being committed to presuppose the information carried by the following sentences :

E. The number of the cab is the smallest number expressible as the sum of two cubes, in two different manners.

F. The number of the cab is the number obtained by subtracting one to 1730.

G. The number of the cab is 1729 or my uncle lives in Boston.

Of course, the agent might eventually fail to realize that he is committed to presuppose that e – when he actually presupposes that d. And even when g is a logical consequence of d, the information carried by g might be deemed irrelevant.

If the theoretical goal pursued by representing context is to describe the assumptions shared by speaker and hearer, and this includes only the assumptions that they are aware of, then context sets are an inappropriate representation of context. [9] This does not mean that the explicit representation of commitments is useless. I shall argue below that representing contextual commitments

[9]The problem under consideration here is the so-called problem of *logical omniscience*. Once certain standard of rationality is fixed, establishing the commitments of a rational agent is unproblematic. If the agent is supposed to be logically consistent and as well as logically proficient, his view should be represented via a non-empty context set - or (syntactically) by a logically closed set of sentences strictly included in the underlying language. If the agent is supposed to be probabilistically coherent, then his degrees of belief should obey the laws of standard probability, and so on and so forth.

It is less clear how to represent the explicit beliefs of *bounded agents*. Trimming down commitments in order to capture inferences of sufficient complexity will not due. Complexity considerations are obviously agent-dependent and difficult to characterize exactly. For example, that d and e are co-extensional was perfectly obvious for the Indian mathematician Srinivasa Ramanujan - perhaps as obvious as the fact that d and f are co-extensional. When visited by a friend in the hospital, his friend told him that the number of the cab in which he traveled to visit him was 1729, a number that he considered insipid. Ramanujan responded immediately that this was not the case, mentioning the property used in e.

One of the possible manners of dealing with this is to derive explicit beliefs in a principled manner from other theoretical considerations, which could be probabilistic [24] or algorithmic [39]. For example, the first strategy would suggest representing explicit beliefs in terms of high probability. No principled criteria of this sort will be invoked here, but we will make room for representations of this sort by weakening the context set approach.

serves an important normative purpose in representing the inferential capacities of agents. For the moment, nevertheless, I shall pause to consider some dynamic aspects of context representation.

3 The conversational record and context change

It should be evident by now that when context sets are invoked in order to characterize speech acts, the dynamics of context is of foremost importance . In the previous section we considered the case of assertion . Following ideas first presented in a seminal paper published in 1978 [47] Stalnaker understands assertion as 'a proposal to alter the context by adding the information that is the content of the assertion to the body of information that defines the context set'.[10] In other words, *the goal of asserting* a proposition is to update the current context. But as Stalnaker clearly explains assertion as a speech act cannot be *identified* with the update operator:

> I should emphasize that I do not propose this a DEFINITION of assertion, but only as a claim about one effect which assertions have, and are intended to have - an effect that should be a component, or a consequence, of an adequate definition. There are several reasons why one cannot define assertion in terms of this effect alone. One reason is that other speech acts, like making suppositions, have and are intended to have the same effect. A second reason is that there may be various indirect, even nonlinguistic, means of accomplishing the same effect which I would not want to call assertions. A third reason is that the proposed essential effect makes reference to another speech act - the rejection of an assertion, which presumably cannot be explained independently of assertion.

In fact, there are a variety of non-verbal acts capable of effecting transitions of currently held attitudes. Thomason mentions the well-known example of Herod when he caused John the Baptist's head to be brought in a platter. Of course, Herod did not assert anything, but his action caused an update in the body of presumptions. On the other hand, epistemology commonly studies *mental acts*, like acceptance, whose (constitutive) point is also effecting transitions of epistemic states. Moreover, even if the point of asserting P is to produce a change of view in the current context, the change in question might not be straightforwardly construed as a function of the content of P and the body of information that constitutes the context set . We just saw the current context C might change as a result of merely asserting P.

But even if speech acts, like assertion, cannot be fully characterized in terms of epistemic transitions, it is undeniable that one can gain interesting insights on those acts by studying the epistemic transitions they purport to effect. Even if this is so, most of the current theories of context refrain from adding specific

[10] The quote is from [50].

dynamic parameters to context sets. Two underlying assumptions can perhaps be invoked as an explanation of this fact. The first assumption is based on the idea that all needed changes can be construed as incremental additions to the current context set. In other words, all represented transitions are construed as changes where the set of possible worlds constituting the current context C has non-empty intersection with the input proposition p. Let's call such changes *additions*. The second assumption suggests how to perform additions. The idea is to add p to a context C by just eliminating from C all possible worlds where P is false. We intend to argue here that both assumptions are based either on misconceptions or on simplifications, which are unduly restrictive. In a recent article Perrault [41] clearly articulated the first assumption, which he calls the *persistence theory of belief*.

> Speech acts reveal certain aspects of the speaker's mental state and cause changes in the state of the hearer(s) that are based on their perception of the state of the speaker. An agent's beliefs after an utterance, for example, will in general depend on his beliefs before it, as well as on its content. Ideally, one would like to have a theory in which it is possible for ones agent's beliefs, say, to change according to how strongly he believed something before the utterance, as well as on how much he believes what the speaker says. I cannot give such an account in detail, so I will rely on something simpler. I assume what might be called a *persistence theory of belief*: that all beliefs persists and that new ones are adopted as a result of observing external facts, provided that they do not conflict with new ones. In particular, I assume that an agent will adopt the beliefs he believes another agent has, as long as those do not contradict his existing beliefs.

Perrault is aware of the fact that adopting persistence is a gross oversimplification. There are various cases that can only be analyzed in terms of belief-contravening changes of view. The contextual evaluation of conditional sentences is an obvious case, but examples abound. John and Terry might be looking for Paul. They might take for granted that Paul is somewhere on campus (in the city of Pittsburgh). Nevertheless, if they meet Al and he tells them that Paul is in NYC (or if Al tells them something that presupposes so), they might face a non-persistent change of presumptions - as long as they consider Al as a reliable oracle. In order to input consistently the information provided by Al some information has to be *contracted* from the prior context.[11]

[11] Stalnaker distinguishes between DEFECTIVE and NONDEFECTIVE contexts in [47]. A nondefective context is one in which the presuppositions of the various participants in the conversation are all the same. Stalnaker thinks that defective contexts have a kind of intrinsic instability, and that therefore will tend to adjust to the equilibrium position of a nondefective context. Of course, non-persistent context changes presuppose defective contexts. Stalnaker assumes that in the normal case contexts will be close enough to being nondefective. The assumption seems too strong. Perhaps Stalnaker uses the assumption because he thinks that defective contexts can lead to communication failure. This could happen in extreme cases but

The foundational and mathematical aspects of a theory of contraction (and revision) of belief has been worked out by epistemologists, computer scientists and decision theorists during the last 20 to 30 years - see, for example, [14], [29], [46] for general background on recent work. Parallel work has been done in analyzing the theory of *supposition* that goes into the evaluation of conditionals. Much of this theory can be immediately applied in order to analyze context change in the domain we are studying. Nevertheless the application calls for a careful use of some of the tools already developed, and in some cases it seems that suitable modifications and extensions are needed. Some preliminary observations concerning these issues will be made below. Nevertheless, our main point in this piece is to rethink what goes into the theoretical notion of context. And the relevance of the issue at hand for our main theme is that, once the need for representing non trivial additions is recognized, the notion of context set is too poor as a general foundational device for representing context. Additional dynamic parameters need to be added.

This point is not completely new in the standard literature. Richmond Thomason elaborated it in [52]. The theoretical device used by Thomason in order to represent context is different from Stalnaker's context sets in many ways, as the following paragraphs make evident:

> [...] the most important component of the conversational record is a structure $\mathbf{P}$ that determines the *presumptions*: the things that are supposed, or established, at a given stage of the conversation. I'll say that $\mathbf{P} \vdash p$ when $\mathbf{P}$ yields the conclusion that p. These presumptions are perhaps best thought of as a kind of shared memory or common database that the participants construct for the purposes of the conversation.
>
> Of course, presumptions will in general be modified in the course of a conversation. I'll suppose that for each proposition p there is an update operator Ap on the presumptions that gives the result of updating the presumption structure by adding the proposition p. $A\mathrm{p}\mathbf{P} \vdash p$ and in many cases $A\mathrm{p}\mathbf{P} \vdash q$ if $\mathbf{P} \vdash q$. However, update is a nonmonotonic operation, so that we can have $\mathbf{P} \vdash q$ but $A\mathrm{p}\mathbf{P} \nvdash q$; also we can have $A\mathrm{p}\mathbf{P} \vdash r$ but $A\mathrm{p} \wedge \mathrm{q}\mathbf{P} \nvdash r$, where '$\wedge$' is propositional conjunction. Since the conversational record is public, the update operator must be public also. I assume that the reasoning mechanisms that operate here are essentially the same as those that operate in the suppositional or conditional reasoning.

This previous account of context (aside from explicitly including dynamical elements) is more liberal than the one given via context sets.[12] The nature of

it seems also true that there is a large set of cases where the process of adjusting towards the equilibrium position of a nondefective context can be accomplished via non-persistent updates, without threatening communication.

[12]It is important to keep in mind that the P-structures just introduced focus on describing occurring or explicit attitudes, not on inference. Perhaps an agent can presuppose that X is a

the structure $\mathbf{P}$ is left unspecified. Notice that, in particular, the behavior of the structure need not be determined by what is entailed by a set of possible worlds. For example, the behavior of a permissible instance of $\mathbf{P}$ might depend on what propositions are members of a set of sets of possible worlds S. In other words, SP might just be the set of presumed propositions, in such a way that $\mathbf{P} \vdash p$ if and only if $p \in SP$. When this is the case we can say that $\mathbf{P}$ is *determined by a presupposition set* SP. This representational structure is capable of circumventing some of the problems mentioned at the end of the previous section. For example, if $\mathbf{P}$ is determined by a presupposition set SP and $\mathbf{P} \vdash p$, there is no need that $\mathbf{P} \vdash r$, when $p \subseteq r$. This filters the 'irrelevant' case mentioned at the end of the previous section. In this case we had $\mathbf{P} \vdash d$ and the fact that $d \subseteq g$ entailed the unintuitive conclusion $\mathbf{P} \vdash g$. If we use context sets in order to determine the behavior of $\mathbf{P}$-structures this consequence is inescapable. In order to see this it is useful to notice that context sets are a limit case of structures of type $\mathbf{P}$. Let $\mathbf{C}$ be a set of possible worlds. Then if $\mathbf{P} \vdash q$ if and only if $\mathbf{C} \subseteq q$ we have a limit case where all presumed propositions are entailed by their infinite intersection $\mathbf{C}$. In this limit case we can say that $\mathbf{P}$ is *determined by a context set* $\mathbf{C}$. Now, of course, when a $\mathbf{P}$-structure is determined by a context set, it is easy to see that if $\mathbf{P} \vdash d$ and $d \subseteq g$, $\mathbf{P} \vdash g$.[13]

So, *some* of the problematic cases considered in the last section can be circumvented by using $\mathbf{P}$-structures. It is easy to see though that not all examples are so easily handled by adopting $\mathbf{P}$-structures. In fact, if d, e and f are the propositions expressed by D, E and F, then it is clear that d, e and f are *the same* proposition. So, nothing forces us to legislate that $Ad\mathbf{P} \vdash g$, but $Ad\mathbf{P} \vdash e$ is unavoidably determined by the fact that $\mathbf{P}$-structures are updated with *propositions*.

Most of the standard theories of belief change allow for finer-grained representations. In fact, most of these theories study how databases (eventually composed by sets of sentences) change when *sentences* are added, preserving consistency. We can call these theories *sentential theories of update*. Nevertheless, most axiomatizations of these sentential functions undo this freedom by adopting a postulate usually called[14] the Principle of Irrelevance of Syntax. The principle legislates that updates prompted by co-extensional sentences are identical. This principle is adopted in virtue of various reasons. One reason is mathematical simplicity. Another, more principled, reason is based on the fact that most of the existing theories of belief change are *normative theories*, describing how the doxastic commitments of rational agents are updated over time .

lawyer and that Y is also a lawyer. He might also presuppose that both X and Y are lawyers, but the P-structures do not require universal closure under conjunctions. After section five I shall consider the *epistemic commitments* contracted by finite doxastic representations of the kind we are studying here.

[13]The behavior of $\mathbf{P}$-structures is reminiscent of the so-called *neighborhood semantics* for modalities defended by Scott and Montague. See [1] for a direct application of this type of semantics to epistemic logic.

[14]Specially among computer scientists. See, for example, [23].

Some of the cognitive theories of discourse processing, like DRT , focus on syntactic representations of context. Hans Kamp, for example, points out in [21] that 'it is tempting (and, I think, up to a point legitimate) to see DRT as a providing a model of the process by which the recipient of a discourse acquires new beliefs as he takes in successive sentences.' Now, we already know that Kamp is a critic of the context set model, and this is evident here too. After proposing examples of the type we considered above, Kamp refuses to analyze belief as a relation between a believer and a set of possible worlds. Kamp adopts instead a syntactic model according to which (roughly) 'beliefs are identified with sentences of a natural language such as English, while recognizing the need to explicate the intentionality of that natural language, presumably by providing a truth- or model-theoretic semantics for it.'

A detailed comparison of P-structures and DRT is beyond the scope of this piece. Such a task requires making explicit the general principles of doxastic change tacitly used in rich accounts of discourse processing (like the one offered in [21]). It should be mentioned in passing nevertheless that the Principle of Irrelevance of Syntax is not among the general principles governing doxastic change in this theory. Given the centrality of this principle in almost all contemporary theories of belief change, it is fair to assume that the format of the theory of change arising from the use of the construction algorithm in [21] is today mostly unknown.

3.1 Success and the act of supposing

Aside from the previous considerations about update, it is important to see that the only substantial property imposed on the update operator A is the property that we can call Success: $Ap\mathbf{P} \vdash p$. Success requires that when the common ground represented by a structure $\mathbf{P}$ is updated with a proposition p, the resulting updated structure yields the conclusion that p.[15]. Some minimal considerations about the notions of supposition and belief revision (with external inputs) can help to put in perspective this and other properties of A.[16]

Notice that changes in view might occur as the result of processing inputs which might have different epistemic origins. Perrault explicitly assumed that 'new beliefs (presumptions) are adopted as a result of observing external facts'. But inputs can, of course, have a purely internal origin. This is the case of changes of view done 'for the sake of the argument', or as a result of *supposing* an item. In other words, supposition can be understood as a species of update . This is indeed the way in which supposition is usually construed in many branches of Bayesian epistemology.

Supposition is the mental act needed in order to evaluate conditionals. Consider John who might face the task of considering the acceptability of the conditional 'If Paul is in NYC, the he will return on Monday'. Even if John firmly

[15]Sentential theories of update would implement: $AP\mathbf{P} \vdash p$

[16]For the sake of terminological clarity we use the term 'update' in order to refer to any change of view, hypothetical or not. 'Belief revision' will be used in order to refer to updates with external inputs.

believes at the moment that Paul is somewhere in the city of Pittsburgh, he can evaluate the conditional by changing hypothetically his mind. He can do so by considering for the sake of the argument that Paul is in NYC. The construction of this hypothetical state will be a function of the background information John currently holds and the hypothetical input he is entertaining.

Success is an appropriate (if not a constitutive) principle that any reasonable suppositional operation should satisfy. It is less clear that updates caused by external inputs should be constrained by Success. If Al tells John that Paul is in NYC this event will trigger a process of belief revision . John will change his mind after receiving the report, but after changing his mind John need not believe that Paul is in NYC. His current belief to the extent that Paul is in Pittsburgh might be so well *entrenched* in his prior view, and his trust in Al as a reliable oracle so flimsy, that John might decide to stay put. Therefore his view about Paul's location might not change after receiving Al's report. Of course, John's posterior view will be different. Among other things John will firmly believe that he received a report from an unreliable source about Paul's whereabouts. But notice that the mental act of supposing does not seem to involve this epistemic freedom. Supposing that p necessitates the implementation of a hypothetical update entailing p. Otherwise one might say that an attempt to suppose that Paul is in NYC failed (the scenario might be unconceivable, given prior beliefs).

Thomason seems to have the notion of supposition in mind when he proposes to constraint A by Success. And we just argued that, for this theoretical application, the postulate seems well motivated. Nevertheless, we will argue in the next section that some applications might call for the addition of other update operators for each agent, which need not obey success. Most of the recent literature on belief revision (with external inputs) has adopted the Success postulate, but there is an increasing consensus in the field towards developing more sophisticated models, which are usually called *non-prioritized*.[17] We will comment on this issue in the coming section.

3.2 Presuppositions and update

In the previous section we argued in favor of a model of context of the type sketched by Thomason in [52]. The model is at the same time weaker and more expressive than models in terms of context sets. Weaker because it allows for finite models of explicit attitudes. More expressive because it explicitly contains dynamic parameters.

Of course, as we saw in the previous section, more work seems to be needed in order to complete the picture first offered by Thomason . For example, it would be nice to have a better understanding of the nature of the dynamic parameters included in the model. More work seems to be needed as well in order to understand the nature of the attitude that goes into the conversational record. In this section we will try to tackle some of these issues. Our starting

[17] See [18] for an introduction to work in this area.

point will be a characterization of presupposition offered by the philosopher Scott Soames in [45].

> (P) An utterance U presupposes p (at t) if and only of one can reasonably infer that the speaker accepts p and regards it as uncontroversial, either because:
>
> a. S thinks that it is already part of the conversational record, or because
>
> b. S thinks that the audience is prepared to add it, without objection, to the context against which U is evaluated.

Soames's model permits to have a more detailed insight into the notion of presupposition itself. Its characterization depends on having a previous account of common ground of conversational record, which can go along the lines considered in section 2 (roughly in terms of common beliefs about what is accepted). An example can help us illustrating Soames's ideas. While John is looking for Paul, he might find Mary, and she might volunteer the following piece of information.

(S) Paul is not on campus, he went to the airport in order to pick up his sister who is flying from Boston.

According to (P) the utterance of (S) presupposes the fact that Paul has a sister, as long as Mary believes so and she believes that the audience 'is prepared to add this fact, without objection, to the context against U is evaluated'. Various issues need immediate clarification.

First let's just assume, for the purposes of considering the example, that the audience is constituted by John. Secondly let's focus on the dynamic aspects of the definition. Notice that the relatively terse presentation of clause (b) hides a considerable degree of complexity.

Let's first tackle this intuitively by considering a case where the act of uttering (S) does presuppose that Paul has a sister (let's call this proposition p). Notice that Mary might believe that John does not presuppose p. She might believe also that John believes that Paul does not have a sister. Moreover she might think that John thinks that it is part of the common ground that Paul does not have a sister. This will not stop Mary from presupposing p as long as she thinks that John will be willing to take her words at face value and revise his views as a result.[18]

If John considers Mary a reliable source of information, the utterance of (S) will generate an immediate change in John's view about the presupposition set. If John thinks, before the utterance, that it is part of the common ground that

[18]This assumes that Soames' idea of characterizing presupposed items in terms of what is uncontroversial for the audience should not be equated with the idea of characterizing presuppositions as uninformative items - for the audience. More about this below.

Paul does not have a sister, the mere utterance of (S) will cause a change in John's take on the presupposition set. After the utterance John cannot think any longer that it is commonly presumed that Paul does not have a sister. This will lead to a contraction of his take on the presupposition set. I.e. John will have to move to a position of doubt considering p. Of course, this contraction might be followed by the addition of p, after the intermediate state of epistemic suspense. The compound process of contracting and then (eventually) adding leads to a revision with p.

Now, notice that what counts here in order to determine whether p is presupposed by uttering s is Mary's expectation as to whether John will actually revise his view. John may very well refuse to open his mind in order to accommodate p. This might be due to different reasons. For example, John can consider Mary an unreliable source of information.

Whether the utterance carries or not the presupposition largely depends of how the expectations of agents interact. For example, Mary might take for granted that Paul has a sister and she might not know whether John has any views about this issue. In this scenario she can utter (S) and she can presuppose p while doing so. In other words, she might think that either John knows about p or that he will unproblematically accommodate p in his view in order to continue the conversation. But, of course, her expectations might miscarry and John might refuse to do so. In this case the conversation can continue as follows. John might say:

Paul's sister? I did not know he had a sister. Are you sure we are talking about the same person?

Of course, the fact that John does not respond as Mary expected, does not conflict with the fact that her utterance did carry the presupposition that p.

We will consider below similar examples as well as some counterexamples. But before doing this let's clarify the formal background presupposed in the analysis. For the sake of terminological simplicity we will keep unchanged the notation used above, when possible.

> An *interactive structure* M is a triple $\langle I, S, U\rangle$, where I is a set of n agents (which at each instant can be partitioned into a speaker s and $n-1$ members of the audience); S is a collection of structures $\{\mathbf{P}, P^1, ..., P^n\}$, and U is a collection of update operators $\{A, A^1, ..., A^n\}$. As before the propositions entailed by the structure $\mathbf{P}$ constitute the body of commonly presumed information. Each structure P^i, with i ranging over I, encodes the propositions that agent i thinks are commonly presumed. Finally A is an update operator on $\mathbf{P}$, which, as before, is successful and non-monotonic. Each A^i, on the other hand, is an update operator on the structure P^i, mapping the prior P^i and non-empty proposition into posterior structures $A^i \mathrm{p} P^i$, for arbitrary (non-empty) propositions p. None of the update operators for the n agents are supposed to obey Success, and all of them are non-monotonic, in the sense defined above. In addition we assume

the following properties (a structure is called consistent if and only if it is not the case that it entails all propositions).[19]

(Preservation) If for all p such that $P^i \vdash p$, $p \cap q \neq \emptyset$, then $A^i \mathrm{q} P^i = \{r = p \cap q\colon\ P^i \vdash p\}$

(Consistency preservation) $A^i \mathrm{p} P^i$ is consistent.

We are primarily interested in *dialogues* where I has cardinality two. The letters 's' and 'h' will be used in order to represent speaker and hearer in a dialogue. Of course, the information provided by updates of P^h, is assumed to be provided by the speaker s. Now we can present Soames' view as follows:

(IS) An utterance U presupposes p (at t) if and only if one can reasonably infer that the speaker accepts p and regards it as uncontroversial, either because:

a. S thinks that $\mathbf{P} \vdash p$, or because

b. S thinks that the audience is disposed to revise its view (successfully) with p (i.e. the speaker believes that $A^h \mathrm{p} P^h \vdash p$).

We should pause here for a moment in order to sum up some of the main features of the analysis so far. First, the reader probably noticed that although the structure $\mathbf{P}$ figures in the definition of interactive structures, $\mathbf{P}$ does not play any crucial role in (IS). Clause (a) uses the $\mathbf{P}$-structure, but (a) can be easily reformulated without it. In fact, the role of P^s is exactly to encode the propositions that the speaker thinks are entailed by $\mathbf{P}$. So, clause (a) can be rewritten as follows:

a. $P^s \vdash p$.

The structure $\mathbf{P}$ is supervenient on the attitudes of speaker and hearer. Theoretically it can be treated as an epiphenomenon created by the interaction of the agents. What is crucial in order to evaluate sentences in context (for example in order to determine speakers's presuppositions, etc) are the speaker's thoughts about the content of the common ground, rather than the common ground per se.

To be sure, a group of agents sometimes directly engages in the process of updating their presuppositions . This can happen in different manners. A simple one occurs when the group consciously engages in the task of building presuppositions step by step. Teaching, joint research or any other complex group activity might require being sure that agents are in epistemic control of the nature of their agreements. But in less regimented situations agents might

[19]Similar principles hold for sentential updates. Such updates will be used in some of the following sections.

only have beliefs about commonly held attitudes. When two agents initiate a dialogue they might in fact share a large body of beliefs, but they might not know that this is the case. As a matter of fact the point of a dialogue is, at least in part, to verify which is the nature of the agreements and disagreements among the participants in the dialogue. So, it seems that a general description of a dialogue requires postulating not only structures in order to describe the agreements among agents, but it also requires postulating structures describing the agent's beliefs about the nature of the agreements. This is the role played by the P^i structures in the model. As we explained above, the updates of these structures is supposed to model not a process of supposition, but a process of belief change with external inputs, which need not obey Success.

A detailed account of the type of revision of P^i-structures requires more elaboration. For example, an analysis of the reliability of sources of information needs to be entered into the model. And what is needed is not only an account of the reliability of the speaker, but also an account of the expectations that agents have about the reliability of the speaker.[20]

We will not go deeper into this issue here, given the foundational nature of this paper. It should be said, nevertheless, that the update process required by clause (b) of (IS) is presented by Soames as a *cognitive decision*, where the agent makes up his mind about accepting, or not, a piece of information. Most of the existing theories of belief change cannot represent this process. This is in part due to the fact that these theories have focused on studying the process of changing view, *after* it is settled that an input will be added to the current view (preserving consistency). In other words, most of the existing theories of belief change accept the postulate of Success and focus on the nature of the process *after* a decision has been made to input information.[21] More sophisticated theories, like the one presented by Isaac Levi in [29] and [30] give a more comprehensive picture. These theories understand belief change as a decision, where part of the problem is to decide *what* to accept, aside from the problem of *how* to accept information in conflict with the current view, in case that this were needed. In order to implement theories of this sort one needs to add a new important parameter in order to characterize context: *value*. In this case we need some measure of the *value of information*, aside from the standard doxastic parameters. Levi, for example, proposes in [29] an account of contraction as an optimal cognitive decision where the goal is to minimize losses of informational value. The theory also has the resources needed to characterize the process of acceptance (i.e. the resources needed to decide *what* to come to

[20]Some recent work has been done in this direction in a probabilistic tradition. An article in this vein, available to members of the context community is [7].

[21]The imposition of the postulate of Success in many contemporary theories of belief change was probably motivated by a methodological decision made in order to divide theoretical labor. The idea in [14], for example, is that the study of how to accommodate information inconsistent with a view, while preserving consistency, is complex enough in order to study it separately. These theories therefore focused in studying the process that follows the decision of imputing belief-contravening pieces of information, leaving aside the complementary process of deciding *what* to accept. Theories of acceptance focus on this second aspect of belief change. It seems that the correct articulation of (IS) requires attending to *both* processes.

believe). Once a theory of this sort is implemented we have a notion of context including a full-fledged epistemological theory as one of its components. The route that led us here is simple. Dynamic parameters are needed in order to specify contexts. And those parameters require the use of a theory of acceptance, i.e. a theory capable of implementing cognitive decisions. We will not go further here concerning this issue, but ultimately the correct specification of context seems to require, at least, this level of complexity.

The evaluation of clause (b) of (IS) also involves suppositional aspects that is worth analyzing in some detail. What is required is a process by which the nature of an update (made by the hearer) is estimated from the point of view of the speaker . We have not considered so far how exactly this process takes place. Perhaps the speaker *simulates* the hearer's body of information (as well as an update of it) in order to form his belief about the prospects of success of the operation in question. This might involve an elaborate process of supposing. The speaker supposes that he is a different agent and he also supposes that an epistemic input is provided to this hypothetical body of belief. Then he can estimate what he would do in case of being in this epistemic situation. Ultimately what transpires in clause (b) is the evaluation of complex conditionals of the form 'If I were the hearer, I would do this and that'. More precisely: 'If I were the hearer, and I were facing a process of belief revision with external input p, then I would decide to accept the input and perform the revision (or I would decide not to accept the input, etc)'. The previous paragraph focused on the complexities associated with this decision. Now, we will pay closer attention to the role of supposition on the part of the speaker.

One simple way of tackling this problem is to postulate a further structure $P^{s,h}$ representing the view that the speaker has about hearer's point of view and an update operator $A^{s,h}$ that the speaker thinks the hearer has. Update operators on this structure are supposed to obey identical properties than the A^i operators - in particular they are not constrained by Success. In addition we added an element previously considered, namely the update operator is sentential, mapping pairs of sentences and propositions to propositions. And, as we discussed above, the operator in question need not obey the principle of Irrelevance of Syntax. So, we will change slightly our notation. $A^{s,h}$ [P,$P^{s,h}$] yields a new P-structure $P'^{s,h}$, which is the result of updating $P^{s,h}$] with the sentence P. Now we can reformulate clause (b) as follows:

(b) S thinks that the audience is disposed to revise its view (successfully) with p (i.e. $A^{s,h}$ [P, $P^{s,h}$] $\vdash p$).

The nature of the kind of suppositional reasoning involved in the speaker's performance is still slightly masked by this representation. More can be done in order to make the suppositional aspect of the reasoning required by (b) transparent. In order to do so we need to represent update operators in the object language. Computational theories of update do this routinely. The idea is to work with structures that are finitely axiomatizable, in the sense that given P^i,

there is a sentence X, such that the proposition x determines P^i. Then we have that $x \subseteq p$ if and only if $P^i \vdash p$.[22]

Now the formula $X_h * P$ can be used in order to represent the hearer's update operator in the object language. $X_h * P \rightarrow P$ indicates that the operation is successful, where $\rightarrow$ is material implication. $|X_h * P|$ represents the proposition corresponding to $X_h * P$. Let now S^s be a suppositional operator for the speaker *obeying Success*. And, of course, '*' is not constrained by Success. We have now all the elements on order to represent the suppositional elements of clause (b) of (IS):

(Suppositional representation of b) S thinks that the audience is disposed to revise its view (successfully) with P (i.e. $S^s(X_h * P) \vdash |X_h * P \rightarrow P|$).

The idea of the representation is that if the speaker supposes that the hearer considers an update with P, then he expects this update to be successful. Thomason's idea that the main ingredients of the reasoning required to characterize context are of suppositional nature is basically correct. Nevertheless, the exact nature of the required reasoning is a bit elusive. The speaker needs to engage in an involved type of suppositional reasoning, entertaining hypotheses about whether the agents in the audience will or not decide to accept certain pieces of information.

4 What is a context of utterance?

For Thomason the body of shared presumptions is the most important component of the conversational record. This is not so in the theory sketched above. Ultimately it is the body of speaker's presuppositions the part of the record that plays a crucial role in understanding the way in which a conversation unfolds. And, following our reconstruction of Soames, in order to determine what goes into the body of the speaker's presuppositions we only need to appeal to: (1) the speaker's beliefs about what is entailed by the shared assumptions, (2) the suppositions of the speaker about how other agents would change their views as a response to verbal stimuli.

Soames' theory has a dual aspect. On the one hand, it can be seen as a characterization of the speaker's presuppositions. On the other hand it can be seen as distilling the presuppositions of utterances or the presuppositions carried by sentences. In the former case we do not need an observer making inferences about the epistemic states of the speaker. Remember that the gist of Soames' idea is that 'an utterance U presupposes p if and only if one can reasonably infer from U that the speaker accepts p and regards is as uncontroversial'. But, we can easily distill a notion of speaker's presupposition from (P), by simply getting rid of this external observer:

[22] Weaker assumptions can be made here when the structures are finite. In those cases we can just take a conjunction of sentences representing each entailed proposition as the syntactic representative of the structure.

(Ps) The speaker S presupposes p (at t) by uttering U if and only if S accepts p (at t) and he regards p as uncontroversial, either because:

a. S thinks that it is already part of the conversational record, or because

b. S thinks that the hearer is prepared to add it, without objection, to the context against which U is evaluated.

First we can define the set of propositions that the speaker thinks belong to the common ground.

$PSs = \{p\colon\ P^s \vdash p\}$.

In addition we have an important set of sentences ACs collecting the propositions that the speaker thinks the hearer will be willing to accommodate in his view. This can be defined by using our doxastic representation of clause (b) in Soames' definition .

$ACs = \{p\colon\ A^{s,h}[\mathrm{P},\ P^{s,h}] \vdash p).\}$.

The presuppositions of the speaker are therefore defined as $Prs = \{$s believes p and $p \in PSs \cup ACs\}$. This is, of course, just a set of propositions, but there are a fair amount of dynamic and static parameters that go into its definition.

Is there a well-defined notion of what is presupposed by an *utterance type* or even by a *sentence type*, over and above what particular agents can presuppose while participating in a dialogue? Soames' definitional strategy clearly shows how one can proceed towards introducing such notions. The first step is to postulate a 'neutral' or 'prototypical' observer making 'reasonable' inferences. The second step is to postulate 'normal conditions of utterance' and define: 'A sentence S presupposes p if and only if normal utterances of S presuppose p'. Now what is presupposed by a sentence depends on what is presupposed *by an utterance*, not by what is presupposed by the agent uttering the corresponding sentence. And what is presupposed by an utterance depends on what a prototypical 'reasonable' observer concludes from evidence provided by verbal behavior. Various levels of normality are compressed in this definition. So, one should not be surprised by its methodological fragility.

There is a important tradition in pragmatics recommending rigorous terminology when it comes to consider the roles of agents in information exchanges. For example, one should resist the temptation of simplifying and saying that sentences express propositions. As Searle points out: 'I do not know how sentences could perform acts of this kind' [43]. By the same token, strictly speaking, sentences do not carry expectations, or, for that matter, they do not presuppose anything. The same applies to utterances. Agents presuppose or suppose or believe. Sentences or acts of uttering don't. One can, for sure, say that an act presupposes the satisfaction of some condition, but this is not the epistemological use of the term presupposition . Presupposing is here understood (and

we argue it should be understood) as an epistemic attitude. And only entities for which agenthood can be appropriately predicated are capable of having attitudes towards propositions.

To be sure, if certain types of presuppositions are sufficiently robust, in the sense that agents tend to presuppose them every time that certain form of words is uttered in relevant conditions, then one can legitimately attach the content of the attitude to the utterance as a default. The methodological robustness of this move will be, of course, dependent of the robustness of the association between utterances and epistemic states. Engaging in the specification of these default rules might be an exceedingly complex task. But such task presupposes a previous epistemological analysis of the various mental acts of presupposing occurring while a dialogue unfolds. And this requires a methodological stance studying the various viewpoints of agents, plus conditions of compatibility among them [11]. This article is mainly concerned with the latter issue.

Now, in spite of the previous defense and further articulation of Soames' theory presented above, it should be said that the theory has been recently criticized and that, therefore, some of these criticisms can easily be used in order to criticize the view of context outlined in the previous two sections. We will consider some of these arguments in the following paragraphs. Let's start with and example originally due to Karttunen [22] :

(K) We regret that children cannot accompany their parents to the commencement exercises.

In a recent paper [15] Gauker objected as follows: 'Soames' own examples show that he does not think 'uncontroversial' means *uninformative*, but is there any other sense in which the presupposition of (K) has to be uncontroversial? One can well imagine a parental revolt, in which the parents insist that the children must be admitted and bring them into the auditorium whether the authorities permit it or not. In that case it will be hard to find any sense in which the presupposition of K is uncontroversial.'

There are two possible scenarios we can entertain here. Either the speaker is aware of the existence of the parental revolt at the moment of utterance or not. If he is not, what counts is what he thinks and supposes about the situation. And, since by assumption he knows nothing about the parental discontent, he might assume by default that nothing anomalous happens. Therefore he is entitled to suppose that the audience will be willing to modify their views with k. As we explained above, ultimately what the speaker evaluates is the outcome of a decision the audience faces. And this decision might involve weighting values of different sort. Here the values are not only cognitive in nature. The audience might weight different types of costs and loses: how costly might be to challenge authority, how much is lost by not bringing their children, how great is the reward of participating in the ceremony peacefully, etc. All this can perfectly be included in the deliberation on whether to accept k or not. As long as the speaker does not have evidence to the contrary, and in the presence of previous

experience, he is entitled to conclude that the audience will be willing to accept k. And if this is so, it seems perfectly correct to say that he presupposed k in uttering (K). Even if as a matter of fact the parental revolt is in progress, there is a perfectly defined sense in which the speaker might think that the audience will find k uncontroversial. It only happens that in doing so he is misjudging the values and doxastic dispositions of this year's audience.

Of course this kind of example is a hard problem for theories that identify discourse context with the common beliefs or assumptions of the participants. For in this case it is unclear how to characterize the speaker presuppositions - especially in the situation where there is a revolt and the speaker is unaware of this. Stalnaker is aware of this problem and in recent writings he has endorsed a view of *accommodation* that solves the problem along the lines sketched above. Mandy Simons in an exegetic paper on the evolution of Stalnaker's views about context and presupposition [44] attributes to Stalnaker the following analysis: 'A speaker presupposes p in uttering U only if she believes that p will be common ground following her utterance' ([44], 267). So, Simons clocludes: 'to say that a sentence S presupposes p is not to say that its appropriate utterance requires the speaker to presuppose p *at the time of utterance*, but to presuppose p *at the idealized post-utterance point* ([44], 267). We have avoided here the postulation of post-utterance time points by appealing to the act of supposing. We prefer also to make explicit the dynamic aspects of the model.

In the other relevant scenario the speaker has reasons to believe that a parental revolt is in the making. But then in uttering (K) it seems reasonable to say that he is not taking k for granted. He might utter (K) without taking k for granted as the last attempt to prevent a conflict. If uttering (K) felicitously requires the speaker presupposing k, then this will be a situation where this does not happen. (K) in this scenario can be seen as an indirect speech act destined to transmit (as politely as possible) a command. And whether the speaker has serious doubts about whether the command will be obeyed or not. Of course Soames' theory can handle this scenario as well, due to the fact that in this case the speaker has enough evidence to falsify clause (b) in (IS).

What is crucial in Soames' definition is not whether the information carried by k is or not uncontroversial (in some objective sense), or the *facts* as to whether there is a revolt in the making. What is crucial is whether the speaker *thinks* (is justified in thinking) that k is uncontroversial.

From the consideration of cases like the one depicted by the first scenario presented above, Soames concludes that (in those cases) 'a speaker's utterance presupposes a proposition, even thought the speaker himself does not presuppose it in the sense' defined by assuming-to-be-part-of-shared-assumptions notion.' In other words, the structure Prs rather than the structure $\mathbf{P}$ seems to be the right representation of the context of utterance of (K).

We do agree with critics of the common-ground theory that Karttunen's example seems to show that 'the distinguishing feature of the members of the context cannot be that they are already shared assumptions at the time of utterance' [[15], page 162]. Nevertheless, this criticism does not seem to grant a fundamental shift in the methodological analysis of context. Some authors, like

Gauker in [15], for example, have concluded that such a shift is indeed needed. The alternative is the postulation of an ontological or objective view of context, according to which the context of utterance is made up of a set of propositions that are objectively given as those propositions that are relevant to the current conversation.

Most of the arguments offered by Gauker in favor of the objectivity of context seem to be based upon observations about asymmetries between the body of speaker's presuppositions and the body of hearer presuppositions. For example, Gauker makes the following observations concerning Karttunen's example :

> ' in light of the phenomenon of information presuppositions we ought to recognize that the propositional context of utterance need not consist of shared assumptions at all. The speaker's presuppositions may be informative to the hearer; so in cases where the presuppositions are propositional elements of the context, those elements need not belong to the hearer's assumptions at all. This conclusion is another step in the direction of objective propositional contexts. [...] We may find, for instance, that a speaker's presuppositions are merely the speaker's own *take* on the context of the propositional context and the case of presuppositions informative to the hearer is that in which the hearer had not taken the objective prepositional context to contain what, as it turn out, the speaker takes it to contain.'

Our own analysis of Karttunen example can recognize all of the observations made by Gauker, without granting that those observations undermine the epistemological methodology at the root of the common-ground analysis, which we consider basically correct. Let's analyze an (slightly modified) example proposed by Gauker in order to make this more explicit.

> Suppose I say to Alice, 'I will meet you in front of the department store at 6 pm'. If things go smoothly, then at 6pm Alice and I will both be standing near the entrance of Nordstrom's. In that case, we may say that Alice and I ended up at Nordstrom's because the content contained 'We will meet in front of Nordstrom's'. [...] But suppose things do not go smoothly, so that at 6pm I am standing in front of Nordstrom's while Alice is standing in front of Bloomingdale's. In that case, we may explain that there was a disparity between what the context really did contain and what one or the other of us took it to contain. In other words, one or the other of us had mistaken presuppositions. In such episodes of misunderstanding, the fault may lie with the hearer or, quite possibly, it may lie with the speaker.

The view of context in terms of shared assumptions says that the context of utterance is constituted by the shared presuppositions of speaker and audience. Gauker challenges this received view. For him contexts of utterance have an objective nature and presuppositions are not defined as elements of these objective

contexts . An interlocutor's presuppositions are defined in terms of what (s)he supposes belong to the propositional context proper. So, according to Gauker's analysis, the utterance:

(U) 'I will meet you in front of the department store at noon'

is supposed to be associated with a propositional context which can either contain 'The department store is Nordstrom's' or 'The department store is Bloomingdale's'. But there are numerous conceivable situations where there might not be any fact capable of objectively deciding which one of the sentences goes into the so-called propositional context. Nevertheless in these situations it is perfectly possible to envisage agents presupposing any of these sentences. Gauker's theory can still explain those cases in terms of a discrepancy between what the context did contain and what one or the other agents took it to contain. Nevertheless the theory seems unable to explain the nature of the agent's takes on objective context. The theory sketched above focuses on explaining this epistemic aspect of the problem - an aspect that need explanation even if one adopts a realistic stance. And the theory elaborates these takes in terms of what is taken to be commonly presupposed, rather than as takes on some objectively constituted aspects of context. Say that Christopher and Alice have planned to meet after work, have dinner and then go to the cinema. Say that Nordstrom's is the department store closest to the agent's work places, that Bloomingdale's is the department store closest to the cinema they want to go, and that there is a 20 minutes commute between both stores. Say, in addition, that the better restaurants are around Nordstrom's, while around Bloomingdale's there are few restaurants of lower quality.

So, when Alice receives the information conveyed by (U) she might think that, although there is a small risk of ending up in a bad restaurant, Christopher might have presupposed that they will meet in front of Bloomingdale's. This inference involves assuming that Christopher is sufficiently risk averse (with respect to arriving late to the theatre) to prefer otherwise. But, of course, she might be wrong. Christopher might not be very risk averse in this case. Sufficiently so in order to presuppose that they will meet in front of Nordstrom's, widening the choice of good restaurants, while increasing the risk of arriving late to the theatre. And, of course, Christopher might underestimate the risk aversion of Alice. It is quite clear that in this case they do not coordinate.

So, in this case the speaker will associate (U) with the presupposition or 'The department store is Nordstrom's'. The hearer associates (U) with the presupposition 'The department store is Bloomingdale's'. Neither presupposition is a shared assumption because there are no shared assumptions about this matter. And it does not seem clear in virtue of what fact there should be some objective and agent-independent context containing any of those presuppositions. What is clear is that the agents fail to coordinate because they have false beliefs. But what they get wrong is not some presupposition that is supposed to be objectively associated with (U) in this exchange. What they get wrong (in this

case) are the preferences of their interlocutors. One agent overestimates the risk aversion of its partner, while the other underestimates it.

The version of Soames' account presented above can nevertheless be used in order to model the previous scenario. Here the speaker is Christopher and Alice is the hearer. So, the proposition corresponding to 'The department store is Nordstrom's' is in Prs because the speaker believes so and he thinks that this is uncontroversial. And he thinks so because he thinks that this proposition should be commonly presupposed. And, finally, he thinks so because he thinks that Alice should presuppose one of those two items, and he also thinks that when facing the decision of what to accept, she will choose Nordstrom's over Bloomingdale's. A similar analysis applies to Alice. Her presuppositions are determined in terms of her take on the content of the presuppositions of the speaker.[23] We spare the reader the details of this analysis, since it proceeds a similar fashion, but with the opposite outcome.[24]

It should be said in passing that the body of shared assumptions plays an interesting hybrid role in conversational exchanges. I might presuppose at the moment that it is raining in Sag Harbor, and I might think that you take that for granted too, but I might be wrong about both facts. The body of our shared assumptions is something about which I can have false beliefs, as I can have false beliefs about the state of the weather in some location. There is, of course, an asymmetry between brute physical facts and bodies of common attitudes. My mental acts contribute to changes of those bodies of attitudes, while my mental acts have no causal relevance for changes in the weather. But the changes introduced in bodies of shared assumptions are *very* different from deliberate changes that I might decide to implement in my view. While we have some control over the latter changes, we often do not have complete control over the former. Only changes of shared attitudes implemented via highly regimented communication protocols are of the latter type, and those are infrequent in common conversation.

So, in common situations the participants in a conversation are uncertain about the contents of the body of shared assumptions and they can at most have beliefs about it. That this is the case seems to have been explicitly recognized by Beaver in a recent series of papers (see, for example, [6]). In this model what is updated by agents (following assertions) are beliefs about presuppositions shared in common . In terms of our model Beaver uses plausibility orderings in order to implement updates of P^i structures. Various common ground candidates are ordered in terms of how plausible it is that a particular common ground candidate is in fact the common ground assumed or intended by the speaker. Updates of this plausibility ordering with a proposition p select the most plausible p candidates. This is a particular mechanism for update first proposed by [46] and later on adopted by various computer scientists. Here

[23] In other words the hearer presupposes that p if she thinks that the speaker considers p as part of the record of in case the hearer thinks that the speaker thinks that the audience is prepared to add it, without objection, to the current context.

[24] Notice that the above considerations offer an account of *why* both speaker and hearer think that their respective presuppositions belong to the common ground.

we preferred not to adopt any particular mechanism for update, and not to say much about whether a qualitative procedure of the sort used by Beaver is sufficient to implement a variant of our model. Nevertheless, the reader can safely conclude form previous observations that a purely qualitative account of update will suffice only if it implements some suitable version of a qualitative theory of decision - incorporating not only a representation of belief, but also a qualitative version of preference or value.

Our proposal of *interactive structures* coupled by the adoption of Soames's analysis of presupposition yields a multi-agent analysis of context. The P-structures used in the model are a step in the direction of cognitive realism and descriptive accuracy (which seem to motivate much of the recent work in computational linguistics). And finally our models rely essentially in making explicit the dynamic structures used in order to update or to suppose. The proposed model intends to inherit some of the basic insights of Thomason's account of context and extend them to a multi-agent analysis compatible with Soames's ideas about presupposition. Finally our proposal is still of epistemic nature. The internal states of agents are the crucial building blocks of a reasonable construal of discourse context.

5 Implicature and the inferential encoding of doxastic commitments.

In the previous sections we adopted a skeptical stance towards the descriptive power of accounts of doxastic context closed under strong rationality conditions. The Stalnakerian account described at the beginning of this article requires the postulation of equilibrium states, which are supposed to be both logically closed and in introspective equilibrium. One way of defining a presumption operator A satisfying these epistemic ideals for a non-empty context set K encoding the common ground is the following (considered also by Stalnaker in previous writings [48]).

(S1) $K_X \subseteq p$ iff $K_X \subseteq A(p)$

(S2) $K_X \not\subseteq p$ iff $K_X \subseteq \overline{A(p)}$

We can call *stable* any set closed under $S1-2$. In addition for any sentence A, we will say that a minimal context for A is any minimal stable set including A – where minimality should be characterized in terms of set inclusion (see below for details).

We argued that equilibrium sets of this sort can be employed in order to present a theory of what is normatively required of rational agents. In other words, stable sets of this sort are better conceived as representing the *doxastic commitments* of interactive agents. Rational agents should strive to reach equilibrium states of this kind, even when these states are rarely (if ever) instantiated in actual exchanges. So, we will propose that the epistemic commitments of stable agents can be articulated in terms of an inference relation as follows:

(AC) B is an *autoepistemic consequence* of A if and only if B holds in every stable context for A.

In this section we will discuss an alternative interpretation of this notion of consequence offered by Robert Stalnaker in [49]. The gist of the proposal is:

(P) what is autoepistemically entailed by a sentence α is what is meant or implicated, but not explicitly said by uttering α.

Stalnaker suggested that a defeasible notion of consequence (explicated via AC) could be used to formally encode Grice's notion of *conversational implicature* - and to understand its context-dependent behavior.

This final section makes three main contributions concerning this proposal. First, we will focus on the tenability of (P). Paul Grice considered in [17] some minimal constraints on implicature needed in order to handle G.E. Moore's paradox of 'saying and disbelieving'. We will show that (P) is incompatible with those constraints. Secondly we will offer an alternative account of AE consequence based on several remarks made by Grice in [17]. According to this account the AE-consequences of a sentence α encode the body of full beliefs to which someone is *committed* after uttering α. In oher words Grice's insight can be represented via our proposal (AC). Thirdly we will offer a preliminary account of the formal properties of this new notion of consequence.

6 Autoepistemic logic: some background

First some clarifications about the underlying language. Let L_0 be a Boolean language and let L be the language formed inductively from L_0 by adding the formation rule :

$$\text{If A} \in L\text{, then L(A)} \in L^{25}$$

L's intended interpretation will be left open for the moment. The notation is reminiscent of the one used in modal logic for the necessity operator. This is basically the idea behind the operator although the emphasis in autoepistemic logic is, of course, epistemic rather than ontologically related.

The following notation will be useful later on (Γ here could be either a theory defined on L_0, L, or languages of intermediate complexity):

$$\mathrm{L}\Gamma = \{\mathrm{L(A)} : \mathrm{A} \in \Gamma\}$$
$$\neg\mathrm{L}\Gamma = \{\neg\mathrm{L(A)}\ ;\ \mathrm{A} \in L\text{, and A} \notin \Gamma\}$$

An AE theory is any set of sentences in L. A crucial concept is the notion of *stable* set:

[25] In order to avoid misunderstandings between propositions and sentences, in the following sections we will use upper case letters, A, B, ..., etc, for sentences of the underlying language.

DEFINITION 1: A stable set Γ satisfies the following properties: (1) Γ is closed under logical consequence, (2) If $A \in \Gamma$, then $L(A) \in \Gamma$, (3) If $A \notin \Gamma$, then $\neg L(A) \in \Gamma$.

Stable sets are sometimes called AE theories. Robert Moore proposed (see [38]) the idea of a *stable expansion* of a premise set Γ. The gist of his proposal is to represent an agent whose epistemic state is both stable and *grounded* in Γ. The basic idea is to characterize a set T containing at most the consequences of $\Gamma \cup LT \cup \neg LT$.

DEFINITION 2: A set T_Γ is a stable expansion of the premise set Γ if and only if it satisfies:

$$T_\Gamma = \{A : \Gamma \cup LT_\Gamma \cup \neg LT_\Gamma \vdash A\}$$

A set Γ_0 of non-modal sentences has exactly one AE extension, but modal sets might have various or no extension (for example, $\{L(A)\}$ lacks extensions and $\{L(A) \rightarrow A\}$ has two).

Now we have enough background to define a non-monotonic notion of AE consequence:

DEFINITION 3: $\Gamma \mid\!\sim A$ means A is contained in every stable expansion of Γ and there is at least one such stable expansion.

Of course, $\Gamma_0 \mid\!\sim A$ means that A is contained in *the* AE extension corresponding to Γ_0. For modal Γ one needs to consider all possible epistemic contexts induced by Γ, i.e. all the AE extensions of Γ.

The previously defined notion of consequence does not enforce the inference from L(A) to A. In other words the pattern:

$$\text{(P2) } L(A) \mid\!\sim A$$ [26]

does not hold. Nevertheless the pattern:

$$\text{(P1) } A \mid\!\sim L(A)$$

is indeed enforced, because every autoepistemic extension of $\{A\}$ contains L(A). The rationale behind this asymmetry is that most autoepistemic theories intend to model a notion of 'weak belief', rather than a notion of certainty or 'full belief'. [27] Rational agents should be self-aware of the facts they (firmly)

[26] If the existence of at least one relevant expansion is not required in DEFINITION 3, (P2) is actually enforced. Although the proviso is not commonly required in the literature we added it here. This seems to be required by a charitable reading of the literature.

[27] Autoepistemic logicians have not paid much attention to similarly motivated notions of consequence where the L-operator is informally interpreted as a notion of certainty, full belief or holding true. This is an unfortunate situation taking into account the centrality of this notion in decision theory and Bayesian epistemology [53] . Preliminary considerations concerning the properties of the autoepistemic closure of sets of full beliefs are presented below. See also [19], [31] and page 12 of [28].

hold as true. Nevertheless agents might believe facts that they are not willing to hold as true.

Stronger notions of consequence have been also considered in the literature. For example Kurt Konolige proposed the following alternative:

DEFINITION 4: $\alpha \Rightarrow \beta$ iff whenever α is in an AE extension, so is β.

This notion is certainly stronger. As a matter of fact it is monotonic. For assume that A $\Rightarrow$ B. Then assume by contradiction that A $\wedge$ C $\not\Rightarrow$ B. Then there is an AE extension E such that A $\wedge$ C $\in$ E and B $\notin$ E. But if this is so A $\not\Rightarrow$ B. For there is an extension (E) containing A that does not contain B.

7 Implicature and autoepistemic inference

The inference from A to L(A) is a very robust pattern enforced by all the notions of consequence considered above (and some alternatives considered below). In fact,

$$\text{(P1) A} \mid\!\sim \text{L(A) and (P1') A} \Rightarrow \text{L(A)}$$

The intended epistemological interpretation of these patterns was discussed at the end of the previous section (the idea being that agents should be self-aware of the facts they (firmly) hold as true).

It is less clear how these patterns can be interpreted when we take into account Stalnaker's pragmatic reading of $\mid\!\sim$ as a notion of (generalized) implicature . One might perhaps try to accommodate (P1) by arguing that part of the implicit speaker's meaning of every utterance A is determined by the speaker's belief in A. After all, if the utterance of A is sincere (and we can restrict our attention to this case) the main *reason* for the utterance is the speaker's belief in A. Nevertheless the grounds on which a sincere utterance is performed should not be confused with what is entailed or implicated by the utterance in question. Implicatures are calculated in terms of the suppositions needed in order to maintain the assumption that the so-called Cooperative Principle is observed (see essay 2 in [17]).

At this juncture it seems appropriate to quote Grice extensively about this point. His 'further notes' on his seminal article *Logic and Conversation* are highly relevant for the issue at hand:

> When I speak of the assumptions required in order to maintain the supposition that the Cooperative Principle and maxims are being observed on a given occasion, I am thinking of assumptions that are non-trivially required; I do not intend to include, for example, an assumption to the effect that some particular maxim is being observed. This seemingly natural restriction has an interesting consequence with regard to Moore's 'paradox' . *On my account, it will not be true that when I say that p, I conversationally implicate that*

> *I believe that p*; for to suppose that I believe that p (or rather think of myself as believing that p) is just to suppose that I am observing the first maxim of Quality on this occasion. I think that this consequence is intuitively acceptable; *it is not a natural use of language to describe one who has said that p as having, for example, 'indicated', 'suggested' or 'implied' that he believes that p, the natural thing to say is that he has expressed the belief that p. He has of course committed himself, in a certain way, to its being the case that he believes that p*, it is bound up, in a special way, with saying that p.[Italics are mine, [17], pages 41-42.]

The moral of the passage seems to be that Grice strongly opposes any possible formalization of his notion of implicature capable of sanctioning (P1). Therefore an important epistemological obstacle against reading AE-inference as a notion of implicature is that such reading does induce (P1).

The problem reappears if we consider some of the features needed for an implicature to qualify as a *conversational implicature* . One of these features is *cancelabilty.*

> [...] a putative conversational implicature that *P*, is explicitly cancelable if to the form of words of the utterance of which putatively implicates that *P*, it is admissible to add, *but not P*, [...]

So, if we focus on *generalized implicature*, anyone who says *Pete is meeting a woman tonight* normally implicates that Pete is meeting someone other than Pete's spouse, sister or mother (Grice adds close Platonic friends to the list). The fact that this putative implicature does indeed qualify as a bona fide implicature is corroborated by the fact that it is perfectly proper to say:

Pete is meeting a woman tonight, but the woman is none other than his wife.

Notice that it is also perfectly possible for Pete himself to cancel the generalized implicature carried by the form of words: 'I am meeting a woman tonight'. In fact, Pete can felicitously say:

I am meeting a woman tonight, who does not happen to be anyone other than my own wife.

Nevertheless it seems that accepting (P1) as a constraint on a formalization of implicature will force us to say that the implicatures sanctioned by (P1) are cancelable by speakers only on pain of incurring G. E. Moore's paradox (see [37]). In fact, say that X says 'It is raining'. If the putative implicature is that X believes that it is raining, then it should not be felicitous for X to say:

(M) It is raining, but I do not believe it.

Which is a form of Moore's paradox (usually called a paradox of 'saying and disbelieving').

The problem seems difficult to avoid because (P1) is a very robust AE inference preserved under different variants of AE logic. In particular the inference is granted by the type of AE theory envisaged by Stalnaker in [49].[28] One possible way out could be to deny Grice's own thoughts about implicature and epistemic paradox, but this way out does not seem feasible. In fact, Grice's arguments are quite persuasive and they seem corroborated by other considerations (like the argument in terms of cancelability offered above). Another possibility could be to remark that (M) requires an indexical use of the belief operator that might not be encoded in the autoepistemic operator L. This way of circumventing the problem does not seem available either. This is so for several reasons. Perhaps the more obvious line of response can be based on the fact that several authors have suggested that AE-operators can indeed be interpreted in a multi agent context as indexical doxastic operators. We will not analyze nevertheless this issue in detail here (see [3] for a preliminary consideration on this problem).

There is also a second obvious line of response for the indexical problem. Even if one does not interpret L indexically, the argument in terms of cancelability can be run for the speaker. Perhaps it is a little bit odd for a third person to say: 'It is raining, and Pete just sincerely and accurately said so, but nevertheless he does not believe it'. Nevertheless this is perhaps not fully paradoxical. But when it comes to Pete himself, he can only cancel the putative AE-implicature by uttering (M). And this is a fully paradoxical.

The interest of Grice's insights is that they also give us an idea of how to understand autoepistemic entailment. In fact, Grice remarks that anyone uttering p also has 'committed himself, in a certain way, to its being the case that he believes that p.' In other words, it is perhaps plausible to say that what follows (classically) from the AE-extension of (a purely Boolean) premise set Γ_0 formalizes the *epistemic commitments* contracted by accepting Γ_0.[29] This is perfectly in line with our own interpretation of the usage of context sets as encoding *doxastic commitments*. Autoepistemic inference seems to fall short of formalizing the more complex notion of implicature. But it can help to clarify the theoretical role of context sets.

This interpretation of AE-inference can make sense of the problematic inference pattern (P1). We can paraphrase Grice and say that any rational agent who accepts A should 'commit himself to its being the case that he believes that A'. And this commitment is the reason stopping him from canceling a putative implicature induced by A. Since the agent is committed to L(A), denying L(A) will put him in an incoherent scenario.

[28] [49] sketches a theory according to which the epistemic context induced by A (in a non-monotonic inference from A to B) is determined by 'all that the agent knows' at certain instant. The L-operator is, in turn, interpreted as a third-person operator of (weak)-belief. Without entering into the details of the proposal, it is intuitively clear that, as long as one endorses the idea that knowledge entails belief, the inference from A to L(A) should be made valid in Stalnaker's framework.

[29] Similar ideas have been defended by other authors in different contexts. For example see [31].

7.1 Certainty and belief

In the previous section we sketched an alternative to Stalnaker's account of AE inference. The idea (inspired by Grice's remarks on commitment) is that the AE consequences of a premise set A make explicit the doxastic commitments contracted by agents holding A true (or accepting the premise set A as given). We also considered above some possible objections to this interpretation. In this section we will focus on some residual problems importantly related to open foundational problems in AE logic.

It is easy to see that:

$$\text{(A)}\ \emptyset \mid\!\sim \text{(T), where (T) is the alethic axiom } \mathrm{L(A)} \rightarrow \mathrm{A}$$

This follows from the fact that every AE extension of a set A is a stable set containing A and that every stable set is a S5 theory. Should we then say that every rational agent is committed to the 'alethic' axiom (T)? Remember that so far we are simultaneously maintaining that the L-operator encodes a notion of weak belief (instead of a notion of certainty or full belief). A negative answer to the former question is part of the philosophical folklore.

Kurt Konolige puts the problem in the following terms:

> AE logic assumes that agents are ideal introspective reasoners who do not subscribe to the principle T. [...] It is troubling, however, that the schema T is satisfied *post hoc* in any extension, but cannot be used in the reasoning to arrive at the extension. In metatheoretic terms, AE logic is not *cumulative* ([25], pp. 228-229).

A cumulative logic is one obeying the principle also called *cautious monotonicity* stating that A, B $\mid\!\sim$ C holds whenever A $\mid\!\sim$ B and A $\mid\!\sim$ C hold. But it is clear that cautious monotonicity fails when A is $\emptyset$, α is a non-tautological formula, C is $\neg$ L(α) and B is the following instance of schema T: L(α) $\rightarrow \alpha$.[30] To be sure there might be good reasons to construct inductive machines which are not cumulative, but the failure of cumulativity in AE logic seems to appear rather artificially as the result of forcing a particular doxastic interpretation on the L-operator used to build AE extensions. This interpretation is forced on L in spite of basic logical aspects of stable theories (the fact that they are S5 theories). Autoepistemic logicians have managed to implement smart moves in order to circumvent this mismatch of intuitions, but, as Konolige makes clear in his remark, the problem persists. The problem also is an epistemological obstacle against our 'commitment view' about the nature of AE inference.

In this article we will consider the following solution. We will define an alternative notion of AE inference, which obeys cumulativity and fits our 'commitment account' of AE inference. In addition we will propose a different pre-systematic understanding of 'L' as an operator of full belief or certainty. Pros

[30] To see that the conclusion of cautious monotony fails it is useful to keep in mind that {L(α) $\rightarrow \alpha$} has two extensions: one including α and another which does not contain α. Therefore it is not true that $\neg$ L(α) is in every AE extension of {L(α) $\rightarrow \alpha$}.

and cons of this approach will be discussed below. But first we need some basic definitions used in AE logic.

DEFINITION 5: A stable set S is *minimal for a premise set A* if S contains A and there is no other stable set S' containing A such that S' $\cap L_0 \subset$ S $\cap L_0$.

Minimal stable sets have been considered in the recent literature in AE logic (see for example [25] pp. 234-5). To be a minimal stable set for A (MSS for A) is a necessary condition for a set of modal sentences to qualify as the set of introspective epistemic commitments associated with A. So a natural development of the ideas considered so far will naturally lead to the construction of a notion of inference such that B follows defeasibly from A whenever B is in all the MSSs for A.

DEFINITION 6: $\Gamma \mathrel{|\!\approx}$ A means A is contained in every minimal stable set for Γ and there is at least a stable S such that $\Gamma \subseteq$ S.

This notion is obviously non-monotonic (consider the case when Γ is empty A is $\neg$ L(A), A is a propositional atom not included in ST(), and Γ is augmented with A). This notion of consequence obeys cautions monotony. To establish this fact we need some previous definitions.

Let me first consider a class of *normal form* formulae. Consider the language L_D such that non-modal (purely Boolean) formulae are in L_D, if $\alpha \in L_D$ and $\beta \in L_D$, then $\alpha \vee \beta \in L_D$, and both L($\alpha$) and $\neg$ L(α) are in L_D, when α is Boolean. We are basically focusing on formulae of the form:

$\neg$ L(α) $\vee$ L(β_1) $\vee$ L(β_2) $\vee$ ... $\vee$ L(β_n) $\vee$ w

where α, β_i and w are all non-modal sentences.

It can be established as a lemma that every consistent and stable A-theory K (where A $\in L_D$) is such that there is a minimal stable set for K. This lemma can, in turn be generalized with the help of the fact (proved in [25], page 230) establishing that every set of modal sentences (containing the L-operator) is K45-equivalent to one in normal form.

Consider now Cautious Monotony. Assume that both A $\mathrel{|\!\approx}$ B and A $\mathrel{|\!\approx}$ C - where A $\in L_D$. We have to check that C $\in$ K where K is a MSS for A $\wedge$ B. Notice that (in the presence of the assumptions) every MSS for A $\wedge$ B is also a MSS for A. For pick an arbitrary MSS for A $\wedge$ B, say K, and assume by contradiction that it is not a MSS for A. Then (since K is an A-theory) there is a stable set K' such that A $\in$ K' and K' $\cap L_0 \subset$ K $\cap L_0$. Now, in virtue of the previous lemma, either K' is itself a minimal A-set or there is K" (such that A $\in$ K" and K" $\cap L_0 \subset$ K' $\cap L_0$). Since A $\mathrel{|\!\approx}$ B, B $\in$ K' (or K"), which contradicts the assumption that K is a MSS for A $\wedge$ B. Now, since K is a MSS for A, and we assumed that A $\mathrel{|\!\approx}$ C, C $\in$ K, which suffices to complete the proof. Almost identical strategy can be applied to establish that $\mathrel{|\!\approx}$ obeys CUT and other basic properties of non-monotonic inference.[31]

[31] Cut establishes that A $\mathrel{|\!\approx}$ C follows from A $\mathrel{|\!\approx}$ B and A, B $\mathrel{|\!\approx}$ C.

Of course, $\models\!\!\approx$ obeys the pattern (P1):

$$\text{(P1) } A \models\!\!\approx L(A)$$

It is also the case that $\models\!\!\approx$ obeys:

$$\text{(P2) } L(A) \models\!\!\approx A$$

This second pattern, according to our understanding of $\models\!\!\approx$, indicates that every agent who accepts L(A) is committed to hold A true. We understand here (tacitly) that the occurrence of non-modal sentences in a stable set represents the fact that the agent in question (firmly) holds (or takes as) true the sentence in question. On the other hand we are leaving open which is the intended meaning of the L-operator compatible with our characterization of $\models\!\!\approx$. Nevertheless it is not difficult to elicit this meaning taking (P2) into account. L cannot be understood as a notion of weak belief. It should be understood as the strongest doxastic attitude of which the agent is capable. Basically L in this context should be understood as a notion of certainty or full belief.

In this context the troublesome inference $\emptyset \mid\!\sim$ (T), ceases to be problematic. In fact, such a pattern only says (in this context) that every agent who fully believes A is committed to the truth of A. It is important to realize that the pattern $\emptyset \mid\!\sim$ (T), should not be interpreted as saying that AE agents are committed to a principle of 'arrogance'[32] legislating that full beliefs are true. What the schema T says in this context is that agents are committed to *accept* that fully believed items are true. In other words, form the point of view of the agent, every item of which he is certain is an item such that he should be willing to (firmly) hold as true. Of course, the item might be false (objectively), but all that counts for AE inference are the introspective obligations of the agent in question (at a certain instant). Therefore rational agents should accept schema T as a coherence principle in charge of maintaining an equilibrium between the items held as true (non-Boolean sentences in his stable theory) and his certainties (sentences prefixed by L-operators in stable sets).

Of course the inference (T) is problematic if L is not pre-systematically understood as a notion of full belief. Why a rational agent should be committed to (firmly) hold as true items which are only (weakly) believed? In contrast it seems reasonable to require that rational agents should be committed to hold as true every item of which (s)he is certain of (or which (s)he fully believes).

AE logicians seemed to have paid attention only to either a variety of notions of weak belief or to the strong notion of knowledge, somewhat neglecting the intermediary notion of full belief. Such a notion is widely used in many fields where doxastic representation matters (the theory of games and decisions is an obvious example of a field where the notion plays a crucial role). This section has been devoted to define a cumulative notion of AE inference which seems optimally understood when the intended interpretation of the L operator is as an operator of full belief. The interpretation has the independent virtue of

[32]The terminology is used in [25].

offering an unified explanation of the role of schema T. The schema is indeed satisfied in every stable theory (as a matter of fact in every AE extension) and the principle can be cumulatively used in AE reasoning.

In addition the picture fits Grice's account of pragmatic inference in terms of commitments, and it circumvents the previous criticisms to the idea of seeing non-monotonic inference as the encoding of pragmatic implicatures. Finally, our definition of $\models\!\!\approx$ satisfies several of the widely accepted properties of non-monotonic inference (like cumulativity) in a rather natural manner.

Acknowledgements: I would like to thank Andrea Bonomi for very insightful comments and Mandy Simons for various useful pointers to relevant literature in Linguistics related to context.

References

[1] Arló-Costa, H., 'First order extensions of classical systems of modal logic: the role of the Barcan schemas,' *Studia Logica* 71, 2002, 87-118.

[2] Arló-Costa, H., Belief revision conditionals: Basic iterated systems. *Annals of Pure and Applied Logic*, 96, 1999, 3-28.

[3] Arló-Costa, H., 'Qualitative and probabilistic models of full belief,' *Proceedings of Logic Colloquim'98, Lecture Notes on Logic* 13, S. Buss, P.Hajek, P. Pudlak (eds.), ASL, A. K. Peters, 1999.

[4] Arló-Costa, H., Collins, J., Levi, I., 'Desire-as-Belief Implies Opinionation or Indifference,' *Analysis*, 55.1, January 1995, 2-5.

[5] Arló-Costa, H., Shapiro, S., 'Maps between non-monotonic and conditional logics,' *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference* (eds.) Nebel,B., Rich, C., Swartout, W., 1992, 553-564.

[6] Beaver, D., *Pragmatics (to a first approximation)*, Ms. Stanford University, 1999.

[7] Bovens, L. and Hartmann, S., 'Belief Expansion, Contextual Fit and the Reliability of Information Sources,' forthcoming in the acts of CONTEXT'01, 2001.

[8] Chierchia, G., *Dynamics of Meaning: Anaphora, Presupposition and the Theory of Grammar*, University of Chicago Press, 1995

[9] Fariñas del Cerro, L., Crocco, G. and Herzig, A., (eds.), *Conditionals, from Philosophy to Computer Science* , Oxford University Press, Oxford, 1995.

[10] Fitting, M., 'Moore's nonmonotonic logic and S5,' manuscript, CUNY, November 1983.

[11] Giunchiglia, F. and Serafini, L., 'Local modal semantics or contextual reasoning = locality + compatibility,' *Proceedings of the sixth International Conference on Principles of Knowledge Representation and Reasoning* (KR'98), 1998, 282-289.

[12] Gabbay, D., 'Theoretical foundations for non-monotonical reasoning in expert systems' *Logics and Models of Concurrent Systems*, K.R. Apt., ed., 1985.

[13] Gabbay, D.M., Hogger, C.J., Robinson, J.A., (eds.) *Handbook of Logic in Artificial Intelligence and Logic programming*, Volk. 3 Nonnmonotonic reasoning and uncertain reasoning, Oxford Science Publications, Oxford, 1994.

[14] Gärdenfors, P., *Knowledge in Flux*, Bradford Book, MIT Press, Cambridge, Mass, 1988.

[15] Gauker, C., 1988, 'What is a context of utterance?,' *Philosophical Studies* **91**, 1998, 149-172.

[16] Ginsberg, M., 'Counterfactuals,' *Artificial Intelligence* 30, 1986, 35-79.

[17] Grice, P., *Studies in the way of words*, Cambridge, Mass., Harvard University Press, 1989.

[18] Hansson, S.O., Editor's Introduction: 'What's New Isn't Always Best,' Theoria, 63, 1997, 1-13.

[19] Hild, M., 'The diachronic coherence of ungraded beliefs,' *Erkenntnis* 50, 1999, 225-242.

[20] Kamp, H. and Reyle, U., *From Discourse to Logic : Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory/Part 1*, Kluwer Academic Publishers, 1993.

[21] Kamp, H., in Anderson, C. A., Owens, J., (eds.), 'Prolegomena to a Structural Account of Belief and Other Attitudes,' *Propositional Attitudes : The Role of Content in Logic, Language, and Mind* (Csli Lecture Notes, No. 20), 1991.

[22] Karttunen, L., 'Presupposition and Linguistic Context,' *Theoretical Linguistics* 1, 1991, 181-194.

[23] Katsuno, H. and Mendelzon, A., 'On the difference between updating a knowledge base and revising it', in *Belief Revision*, (ed.) Gärdenfors, P.: Cambridge University Press, Cambridge, 1991, 183-203.

[24] Kyburg, H.E.Jr., *Probability and the logic of rational belief* Wesleyan University Press. Middletown, 1961.

[25] Konolige, K., 'Autoepistemic Logic,' in *Handbook of Philosophical Logic, second edition*, Vol III Extensions of Classical Logic (eds.) D. Gabbay and F. Guenthner, Reidel, Dordrecht, 1998.

[26] Kratzer, A. (1991). Modality. In: von Stechow, A. and Wunderlich, D. (eds.) *Semantik. Ein internationales Handbuch der zeitgenossischen Forschung*, Walter de Gruyter, Berlin, 1991, 639-650.

[27] Lehmann, D., Magidor, M., 'Rational logics and their models; A study in cumulative logics,' *Technical Report TR 88-16 of the Dept. of Computer Science, Hebrew University of Jerusalem*, November 1988.

[28] Levi, I., *The Enterprise of Knowledge*, MIT Press, Cambridge, Mass, 1980.

[29] Levi, I., *The fixation of belief and its undoing*, Cambridge University Press, Cambridge, 1991.

[30] Levi, I., *For the Sake of Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning.* Cambridge University Press, Cambridge, England, 1996.

[31] Levi, I., 'The logic of full belief,' in *The Covenant of Reason: Rationality and the Commitments of Thought*, CUP, 1997, 40-69,.

[32] Lewis, D., *Counterfactuals* (1973).

[33] Lewis, D., 'Desire as belief,' *Mind*, 1988, 323-33.

[34] McCarthy, J., 'Notes of formalizing context,' *13th IJCAI*, vol 1, 1993, 555-560.

[35] Makinson, D., 'General Theory of Cumulative Inference' *Non-Monotonic Reasoning*, M. Reinfrank, J. de Kleer, M. Ginsberg, and E. Sandewall eds., 1988.

[36] Makinson, D., 'Five faces of minimality,' *Studia Logica* 52, 1993, 339-379.

[37] Moore, G.E., 'Russell's 'Theory of Descriptions,' in *Philosophical Papers* by G.E. Moore, ed. H.D. Lewis, London: Muirhead Library of Philosophy, 1959, [2nd ed. 1963].

[38] Moore, R., *Logic and Representation*, CSLI Lecture Notes No. 39, 1995.

[39] Parikh, R., 'Logical omniscience,' in *Logic and Computational Complexity*, Ed. Leivant, Springer Lecture Notes in Computer Science, no. 960, 1995, 22-29.

[40] Pearl, J., 'Probability of causation: Three counterfactual interpretations and their identification,' Technical Report R-261, UCLA, 1998.

[41] Perrault, C. R., 'An application of default logic to speech act theory,' *Intentions in Communication*, eds. P. Cohen, J. Morgan and M. Pollack, Cambridge, MA, MIT Press, 1990.

[42] Schiffer, S., *Meaning*, Oxford University Press, 1972.

[43] Searle, J. R., 'What is a speech act? reprinted in *The Philosophy of Language*, third edition, A. P. Martinich, ed., Oxford University Press, 1996, 130-140.

[44] Simons, M., 'Presupposition and Accommodation: Understanding the Stalnakerian picture,' *Philosophical Studies* 112, 2003, 251-278.

[45] Soames, S., 'How presuppositions are inherited: A solution to the projection problem,' *Linguistic Inquiry* 13, 1982, 483-545.

[46] Wolfgang Spohn., A general non-probabilistic theory of inductive inference. *Causation in Decision, Belief Change and Statistics*, eds. W. Harper and B. Skyrms, Dordrecht:Reidel, 1988, 105-134.

[47] Stalnaker, R., 'Assertion,' in *Syntax and Semantics* vol. 9: *Pragmatics* ed. P. Cole, Academic Press Inc., 1978, 315-322.

[48] Stalnaker, R., 'A note on non-monotonic modal logic,' *Artificial Intelligence* 64, 1993, 183-196.

[49] Stalnaker, R., 'What is a non-monotonic consequence relation? *Fundamenta Informaticae* 21, 1994, 7-21.

[50] Stalnaker, R., 'On the representation of context,' *Journal of Logic, Language and Information* 7, 1998, 3-19.

[51] Stalnaker, R., 'Common Ground,' *Linguistics and Philosophy* 25 (5-6), 2002, 701-721.

[52] Thomason, R., 'Accommodation, meaning and Implicature: Interdisciplinary foundations for Pragmatics,' *Intentions in Communication*, eds. P. Cohen, J. Morgan and M. Pollack, Cambridge, MA, MIT Press, 1990, 325-363.

[53] van Fraassen, B., 'Fine-grained opinion, probability, and the logic of full belief,' *Journal of Philosophical Logic* **24**, 1955, 349-377.