Scorekeeping in an Uncertain Language Game

David DeVault[1]
[1]Department of Computer Science
Rutgers University
Piscataway, NJ 08845-8019
David.DeVault@rutgers.edu

Matthew Stone[1,2]
[2]Human Communication Research Centre
University of Edinburgh
Edinburgh EH8 9LW, UK
Matthew.Stone@rutgers.edu

Abstract

Received views of utterance context in pragmatic theory characterize the occurrent subjective states of interlocutors using notions like common knowledge or mutual belief. We argue that these views are not compatible with the uncertainty and robustness of context-dependence in human–human dialogue. We present an alternative characterization of utterance context as objective and normative. This view reconciles the need for uncertainty with received intuitions about coordination and meaning in context, and can directly inform computational approaches to dialogue.

1 Introduction

The question we address in this paper is how utterance context should by represented in implemented conversational systems. Strong intuitions about coordination in conversation (Clark and Marshall, 1981) have led many researchers, e.g. (Traum, 1994; Poesio and Traum, 1997; Rich et al., 2001; Blaylock, 2005), to aim to represent the common ground beliefs that seem to guarantee principled coordination between speaker and hearer on each new utterance (Lewis, 1969; Stalnaker, 1974). Other researchers, in pursuit of robust implementations for real-world dialogue, have opted to represent narrower aspects of interlocutor and conversational state using models that afford a straightforward treatment of uncertainty (Roy et al., 2000; Horvitz and Paek, 2001; Gruenstein et al., 2004).

These differences might seem to be a matter of emphasis rather than substance. In fact, however, the notion of uncertainty about the context is profoundly at odds with received views of context both in theories of presupposition in formal pragmatics (Stalnaker, 1974; Poesio and Traum, 1997) and theories of coordinated activity in AI (Cohen and Levesque, 1991; Grosz and Kraus, 1996; Blaylock, 2005). As we will argue, this tension originates in the central role these theories grant to various nested or higher-order beliefs that interacting agents may have about each other. In Section 2, we review both the rationale for defining utterance context in terms of these beliefs and the challenges that doing so poses to system builders. The contribution of this paper is a new characterization of utterance context which overcomes these challenges by narrowly circumscribing the aspects of interlocutor mental states that are necessary for utterance interpretation. We present this new characterization in Section 3. The discussion in Section 4 shows how this view of context reconciles the practical need for uncertainty with received intuitions about coordination and meaning in context.

2 Context and mutual belief

We use the term *utterance context* to label the body of information used in utterance interpretation, including both grammatical conditions required for the utterance to be meaningful and situational factors required to analyze the utterance as a contribution to interlocutors' ongoing joint activity. It is commonly assumed that this information must be mutually believed[1]; see, e.g., Stalnaker (1998). One of the first and most widely known definitions of mutual belief is due to Schiffer (1972). The definition records an infinite, hierarchical interrelation between the private beliefs of a speaker *S*

[1]or some analogous status of mutual knowledge, mutual supposition, etc.

and a hearer H about some proposition p:

$$\begin{array}{rlll} & \mathtt{B}_S\ p & (a) & \\ \wedge & \mathtt{B}_H\ p & (b) & \\ \wedge & \mathtt{B}_S\mathtt{B}_H\ p & (c) & \\ \mathtt{MB}_{S,H}\ p =_{\text{def}} \wedge & \mathtt{B}_H\mathtt{B}_S\ p & (d) & (1) \\ \wedge & \mathtt{B}_S\mathtt{B}_H\mathtt{B}_S\ p & (e) & \\ \wedge & \mathtt{B}_H\mathtt{B}_S\mathtt{B}_H\ p & (f) & \\ \ldots & & & \end{array}$$

The modal operators $\mathtt{B}_S$ and $\mathtt{B}_H$ represent the beliefs of S and H, respectively.

The rationale for defining utterance context as the set of mutually believed propositions is theoretical. For example, we know that an agent that *has* mutual belief with its partner can avoid certain errors in solving coordination problems (Lewis, 1969), in interpreting definite references in conversation (Clark and Marshall, 1981), and in participating in multi-agent collaborations (Cohen and Levesque, 1991; Grosz and Kraus, 1996). In conversation more generally, it is argued that an agent that interprets utterances against the "common ground" of mutual belief can better avoid potential misunderstandings (Clark, 1996).

Yet the mutual belief view of context poses several serious challenges for system builders. First, it is unclear what implications arguments about the role of mutual belief in coordination have, if any, for *representation*. If agents sometimes need to *have* mutual attitudes, must their context representations therefore *describe* mutual attitudes? Of course not: agents might not be coordinating using only their context representations. Even if they are, there's still a huge gap between the *conditions* rational system behavior depends on (e.g., that a successful agent acts in accord with what is mutually believed) and the *meaning* of the underlying representations (e.g., that an agent's representations directly track what is mutually believed). See Dennett (1989). Nevertheless, it's common to assume that dialogue context representations *should* track the mutual beliefs of agents and their interlocutors—see, for example, (Traum, 1994; Poesio and Traum, 1997; Rich et al., 2001; Blaylock, 2005).

This leads immediately to the second problem, the methodological challenge of correctly identifying what is mutually believed, so that utterance context can be implemented correctly. It is relatively straightforward to implement a theoretically sound formalism for mutual belief in dialogue. However, it is another matter to enable such an implementation to model its conversations accurately. In the absence of any available "ground truth"—such as an utterance-by-utterance trace, for each conversation in a corpus, of empirically observed higher-order attitudes—we have no way to tell whether (1) does or should hold.

For this and other practical reasons, some researchers endorse a weaker notion than mutual belief for context representations in dialogue. For example, Taylor, Carletta and Mellish (1996) argue that we can eschew the indefinite nesting suggested by (1) in favor of a bounded depth of three. More commonly, researchers suggest that context representations should be relativized to a specific perspective (Ginzburg, 1996), so that they track nested information such as $\mathtt{B}_S\ \mathtt{MB}_{S,H}\ p$ or $\mathtt{B}_H\ \mathtt{MB}_{S,H}\ p$ or both. Note that such changes undermine one of the key virtues of the mutual belief view: its explanation of why reasoning about context helps interlocutors solve coordination problems. As long as there is any asymmetry across interlocutors, we can apply constructions like Clark and Marshall's "Roxy" scenario (1981) to show that coordination failure is possible.[2] In any case, even weakened versions of mutual belief still require higher-order beliefs like (1c,d) and (1e,f). And there is insufficient evidence for an analyst to make principled decisions *herself* about whether such beliefs obtain, much less automate these decisions.

A third challenge for treating context as mutual belief lies in cases where utterance interpretation felicitously exploits information one interlocutor lacks. For example, Kaplan (1989) argues that, regardless of interlocutors' information states, use of the noun phrase *that* refers to whatever the speaker designates with the accompanying demonstration. The correct interpretation, therefore, reflects what was actually designated, even when this differs from what the speaker believes was designated. Similarly, Gauker (1998) presents a hearer-independent explanation for the "informative presuppositions" of factive verbs like *regret*. For Gauker, *We regret that tonight's show is canceled* is felicitous because it requires for its meaningfulness only the *fact* that the show is canceled, not *mutual belief* between speaker and hearer that the show is canceled.

[2] Of course, coordination failure does sometimes occur in human dialogue, so this certainly does not rule out notions of context that differ from mutual belief.

A final challenge, which we will particularly emphasize in Section 4, comes in characterizing problematic communication on the mutual belief model. It has been common for discrepancies between the contexts believed to obtain by two interlocutors to be marginalized.[3] Yet in computational models of interpretation, some degree of uncertainty about what an utterance means is the *norm*, so discrepancies are unavoidable. When discrepancies do arise, interlocutors often seem to know they *lack* mutual belief, but manage to communicate with context-dependent language anyway. In the next two sections, we present a view of context that explains this capacity in terms of the interlocutors' uncertainty about a true context, and that answers the other challenges as well, while at the same time maintaining the intuitions about coordination that have historically made a higher-order attitude model of utterance context attractive.

3 Objective, normative context

The two basic principles in our characterization are that utterance context is *objective* and that it is *normative*. By *objective*, we mean that *there is a fact of the matter about what the context is at each time t in a conversation, and this context is not a function of the interlocutors' beliefs at time t*.[4] Thus, context is not a matter of one or the other interlocutor's perspective on the situation, and nor is it an interaction between their combined perspectives. Instead, the objective context is the *product of action* taken by the individual interlocutors at times $t' < t$. Agent mental states still play a role, but this role is limited to classifying actions as interlocutors intend them. Action-based characterizations of context have also been advanced on higher-order attitude views of context (Thomason, 1990; Poesio and Traum, 1997), but the presence of higher-order attitudes in these models creates all the challenges discussed in Section 2.

By *normative*, we mean that the job of interlocutors' context representations is to target the objective context as it really is. While systems might aim to *achieve* mutual belief to avoid misunderstanding, their context representations, we will argue, should not *mean* that propositions are mutually believed. They should mean simply that those propositions are *true* in the objective context.

More generally, we believe that *all* interactions where people coordinate by following conventional social rules give rise to an objective, normative context. A clear case is correspondence chess, where players send moves by email. Normally, we might expect each player to keep track of the game by moving pieces on a physical chessboard, keeping the board in sync with their moves. But actually two ambitious players could use only their emailed moves and their imaginations to play chess. In what follows, we adapt Lewis's (1979) scorekeeping metaphor to this case: we use such mental chess to develop a vocabulary for describing context as the abstract product of coordinated activity (Section 3.1), show how this vocabulary applies to dialogue (Section 3.2) with its much more complex and open-ended conventions and context, and use a case of misunderstanding to show how this vocabulary differs from models based on mutual belief (Section 3.3).

3.1 Context as a product of action

We can treat the state or *context* of a chess game as an abstract structure $c = \langle t, s_1, s_2, ..., s_{32}, h \rangle$ recording whose turn t it is to move next (one or the other of the players), the current status s_i of each of the 32 chess pieces (piece type and position—either some board position or "captured"), and limited historical information h (e.g. whether certain pieces have ever moved). Let us write c_t for the context at time t, and let the initial context c_{t_0} be the starting configuration for a game of chess.

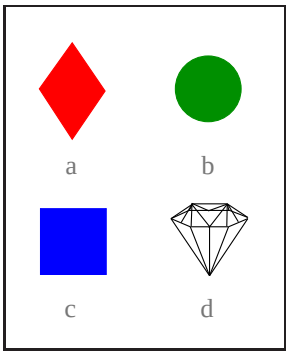
In chess there is a set $\mathcal{A}$ of possible moves or *action types*, which we might formalize parametrically as $\mathcal{A} = \{\texttt{advancePawnOneStep}(P), \texttt{moveQueen}(Q, Pos), \texttt{castle}(R), ...\}$. Each move a is $\sigma(\alpha)$ where $\alpha \in \mathcal{A}$ and σ instantiates the free parameters of α. Doing a effects a *deterministic* transformation on the current context. We can formalize this by way of an update function:

$$c_{t+1} = \texttt{update}(a, c_t) \tag{2}$$

One goal of each participant in a mental chess game, then, is to track the evolving context c_t as a stream of chess moves $\langle a_1, a_2, ... \rangle$ plays out over email messages.

[3] This trend goes all the way back to the first formal model of context, that of Stalnaker (1978). Stalnaker calls each speaker's private context model *nondefective* if it coincides with that of his interlocutor, and suggests that this be treated as the normal case.

[4] When we say context is *objective*, we don't mean to suggest that context is visible, or easily definable in the language of physics, or even that it can be defined independently of human minds and purposes. The point is just that context is not determined by what the interlocutors are currently, privately thinking.



C_1: *the target is a diamond*

U_2: *you mean a diamond jewel?*

C_3: *yeah*

U_4: *ok*

Figure 1: User interaction with the COREF agent. The user (U) can see the four displayed objects, but not COREF's (C) private labels {a,b,c,d} for them. The target in this example is object d.

We maintain that the evolving context c_t in such a mental game of chess is objective, and that even though the current context is not physically realized (on a chess board, for example), it would be very *misleading* to define it in terms of the players' beliefs at time t. The context is objective because, at each time t, the context c_t is an abstract structure that is well-defined given the sequence of moves $\langle a_1, a_2, \ldots, a_{t-1} \rangle$ that have been exchanged by email. It would be misleading to define c_t by way of the players' beliefs about it because their individual beliefs may manifest any number of errors: one or even both players may have forgotten or misunderstood where one piece or another is, whether a knight has been captured, etc.[5] If we were to model the chess context by way of the beliefs (or mutual beliefs) of the players, our model would capture more of the players' *perspectives*, but it would obscure the objective status of the underlying game, and it would hide the normative role played by the true state as players improve their chess skills, recover from mistakes, and cope with their private uncertainties.

3.2 Utterance context and intended actions

We illustrate our approach to utterance context using COREF, an implemented dialogue system that collaboratively identifies visual objects with human users (Thomason et al., 2006). Figure 1 shows an excerpt of an interaction with COREF. COREF is designed to participate in *collaborative reference* (Clark and Wilkes-Gibbs, 1990), in which human interlocutors come to agree on a target object through an interactive, multi-utterance dialogue involving linguistic expressions of heterogeneous form and function.

We understand utterance context as an abstract, objective structure, analogous to the chess state, but now populated by the familiar attributes of dialogue state: sets of propositions contributed to the conversational record, plans that are underway, outstanding interlocutor obligations, linguistic forms of prior utterances, etc. The state depends on what interlocutors are doing. In COREF's domain, we have found that dialogue context takes the form $c^* = \langle R, P, T, C, U \rangle$, where R is a set of referents yet to be identified, P is a set of agreed propositions, T is a stack of tasks (where each task specifies what actions can occur next), C is a set of constraint networks (one for each target referent), and U is the universe of discourse (a set of properties and objects).

Such an utterance context evolves over the course of the dialogue through the domain-dependent set of action types, $\mathcal{A}$, that interlocutors take. The COREF action set $\mathcal{A}^*$ includes actions that select the referent sequence, initiate collaborative reference to a particular target referent, add a constraint C to the constraint network for a target ($\texttt{addcr}(C)$), mark a target as identified, initiate a clarification subtask, and inquire whether some action can be taken. Each action $a = \sigma(\alpha)$ for $\alpha \in \mathcal{A}$ has a *deterministic* effect on the current context, which we again capture by an update function as in (2). This way, we can implement an update mechanism (Larsson and Traum, 2000) that tracks the *objective* context—taking idealized representations of linguistic interpretation, as in (Asher and Lascarides, 2003), and using them for on-line processing, *without* analyzing updates in terms of higher-order attitudes as Poesio and Traum (1997) do.

A key feature of COREF is that the context engendered by these actions *is not mutually believed*. COREF dialogues, unlike chess, include *tacit actions* that allow one interlocutor unilaterally to update the context in ways the other does not know about. These moves allow COREF to handle grounding, clarification, task progress and problem-solving robustly with a model that applies symmetrically in understanding and generation and avoids exceptional pragmatic processes of accommodation or implicit repair. In understanding, when a speaker S utters a linguistic form l, we

[5] In case of a dispute, if the email logs were available, the actual chess state could be settled by examining the move history. This would show who was right.

view S as intending to signal both *what the current context is* and *how it should be updated*. If the last utterance resulted in context c_t, the next utterance should be interpreted in a new context c_{t+n} that incorporates the results of some tacit action sequence $A_l = \langle a_1, ..., a_n \rangle$:

$$c_{t+n} = \texttt{update}(a_n, \texttt{update}(..., \texttt{update}(a_1, c_t)...))$$

The utterance then signals a_{n+1} and creates context c_{t+n+1}. For example, in interpreting the user's utterance *you mean a diamond jewel?*, U_2 in Figure 1, COREF interprets the user as signaling the following sequence:

1. initiate a clarification subtask,
2. start collaborative reference targeting COREF's intended property `P`, (3)
3. inquire whether to take action `addcr(equals(P,diamondJewel))`

Only the last of these is directly associated with the semantics of the utterance. After interpretation, COREF updates its model of the context to reflect these actions.

3.3 Representing the ground truth context

It is easiest to highlight where our characterization of context differs from mutual belief with a case of misunderstanding. Consider the COREF dialogue excerpt D_1–M_2–D_3 presented in Figure 2. The figure tracks the evolution of the context, under both objective and mutual belief characterizations, in a case of misunderstanding. D begins with the red rhombus, i.e. object `a` at the top left of Figure 1, as the value of a target variable `t`. Within this domain, *diamond* can mean either `rhombus` (as in card games) or `diamondJewel` (as in jewelry stores). D utters D_1, *the target is a diamond.* While D intends action `addcr(rhombus(t))`, as it happens, M interprets D as doing `addcr(diamondJewel(t))`.

What happens, we argue, is that after D_1, the intended action `addcr(rhombus(t))` takes its objective effect. D knows what his intended action was, so D updates his model of the context *correctly*. M however comes to believe *erroneously* that `diamondJewel(t)` is in the context. By contrast, if context is mutual belief (or any higher-order attitude), the misunderstanding keeps both `rhombus(t)` and `diamondJewel(t)` out of the context. So *both* D and M are mistaken: D believes it mutually believed that `rhombus(t)`, as D intended, while M believes it mutually believed that `diamondJewel(t)`, as M interpreted.

These then are the basic facts about ground truth and the agents' representations thereof on the two views of utterance context. In the next section, we will use this example to assess the merits of the objective view for system building.

4 Discussion

In designing a representation of context, system builders should be able to explicate their agents' representations and inference in terms of the events in the dialogue on the one hand and the meanings of the representations on the other. Section 2 posed four challenges that make this difficult when context is construed as mutual belief. Our characterization of context allows system builders to meet each of them. For ease of presentation, we take them up in reverse order.

4.1 Miscommunication and uncertainty

The challenge of representing context in the presence of miscommunication and uncertainty is well-illustrated by the example of Figure 2. When D says D_3, *the diamond is red*, M will detect a problem, because while the context appears to M to describe the target as a `red diamondJewel`, there is no such object. Upon detecting the problem, M can reinterpret D_1 and thus correct his private model of the objective context: M had at first thought the context was $[c_2]$ `diamondJewel(t)` whereas M now recognizes that the true, objective context was $[c_2]$ `rhombus(t)`. This allows D_3 to be interpreted as meaning that the target `rhombus` is `red`, as intended. Because context is normative, utterances can be seen as contextually acceptable iff they are interpretable in the true context. This allows interlocutors, like M here, to reason "backwards" from a presumably acceptable intended interpretation to what the true context must be.

Compare the mutual belief model, where the true context before D_3 does not include `rhombus(t)`, because that isn't mutually believed before D_3. On this model, although M did have an erroneous representation of the context before D_3, fixing *that* error does not help to interpret D's utterance. When M discovers what is mutually believed, it's that *nothing* is mutually believed. This correction neither remedies the misunderstanding of D_1 nor makes D_3 interpretable. Thus the ground truth about mutual belief cannot play the simple

	D_1: *the target is a diamond*	
D intended:	`addcr(rhombus(t))`	
M interpreted:	`addcr(diamondJewel(t))`	
	objective context	**mutual belief**
update:	$c_2 = \texttt{update}(\texttt{addcr(rhombus(t))}, c_1)$	(mental events)
ground truth:	$[c_2]$ `rhombus(t)`	$\mathsf{MB}_{D,M}$ (nothing about `t`)
D private:	$\mathsf{B}_D[c_2]$ `rhombus(t)`	$\mathsf{B}_D\mathsf{MB}_{D,M}$ `rhombus(t)`
M private:	$\mathsf{B}_M[c_2]$ `diamondJewel(t)`	$\mathsf{B}_M\mathsf{MB}_{D,M}$ `diamondJewel(t)`
	M_2: *ok*	
	(no change from c_2 to c_3)	(no change)
	D_3: *the diamond is red*	
	objective context	**mutual belief**
update:	$c_4 = \texttt{update}(\texttt{addcr(red(t))}, c_3)$	(mental events)
ground truth:	$[c_4]$ `rhombus(t)` $\wedge$ `red(t)`	$\mathsf{MB}_{D,M}$ `red(t)`
D private:	$\mathsf{B}_D[c_4]$ `rhombus(t)` $\wedge$ `red(t)`	$\mathsf{B}_D\mathsf{MB}_{D,M}$ `rhombus(t)`$\wedge$`red(t)`
M private(?):	$\mathsf{B}_M[c_4]$ `diamondJewel(t)` $\wedge$ `red(t)`	$\mathsf{B}_M\mathsf{MB}_{D,M}$ `diamondJewel(t)` $\wedge$ `red(t)`

Figure 2: A misunderstanding in COREF's domain. D is the director (the initiator of reference) and M is the matcher. The visual display is as in Figure 1. We write $[c_t]$ p to mean p is part of context c_t.

D_1:	*the target is a diamond*
D intended:	`addcr(rhombus(t))`
M interpreted:	
$p = 0.6$	`addcr(diamondJewel(t))`
$p = 0.4$	`addcr(rhombus(t))`
M_2 :	*ok*

Figure 3: A probabilistic misunderstanding.

normative role that the objective context does.[6]

The normative role of context also allows an agent to employ straightforward statistical reasoning to cope with its uncertainty. Concretely, suppose M assigns probabilities to alternative interpretations, as illustrated in Figure 3. In this example, M is sufficiently certain of D's intention to proceed with *ok* in M_2. On our model, there is no impediment to treating M's private model of the context after M_2 as:

$$\begin{aligned} P([c]\ \texttt{diamondJewel(t)}) &= 0.6 \\ P([c]\ \texttt{rhombus(t)}) &= 0.4 \end{aligned} \tag{4}$$

The probabilities capture M's uncertainty about how D's intended action in D_1 changed the context. And M can exploit evidence across multiple utterances, such as when M realizes his misunderstanding after D_3, to reduce uncertainty about the true context. It's just Bayesian inference.

On the mutual belief approach, however, there seems to be no practical route to a useful internal model of context analogous to (4). Let's assume, for exposition, that beliefs and higher-order beliefs are all bivalent. Then upon hearing D_1 in Figure 3, M must choose what to believe. Suppose M chooses to believe `diamondJewel(t)`, and further to believe `diamondJewel(t)` is mutually believed. Then M must assign $P(\mathsf{MB}_{D,M}\ \texttt{rhombus(t)}) = 0$: *M's own lack of belief rules it out! M* then ends up with this private model:

$$\begin{aligned} P(\mathsf{MB}_{D,M}\ \texttt{diamondJewel(t)}) &= 0.6 \\ P(\mathsf{MB}_{D,M}\ \texttt{rhombus(t)}) &= 0.0 \\ P(\mathsf{MB}_{D,M}\ \texttt{(nothing about t)}) &= 0.4 \end{aligned}$$

The model frustrates future Bayesian inference: D's intended context is ruled out, while the seemingly irrelevant "no mutual belief" scenario remains. In fact, no matter how we take uncertainty into account,[7] M's uncertainty after D_1 is not well summarized as uncertainty about occurrent mutual beliefs with D; M's problem, as Figure 3 suggests, lies in M's *own* belief state—to which M has complete introspective access. Reducing uncertainty about mutual beliefs does not solve this problem; reducing uncertainty about objective context does.

[6] To respect the mutual belief semantics, we must construe M's misunderstanding recovery process at best as one of constructing counterfactual sets of mutual beliefs, sets which could have been actual if certain private mental events had occurred that did not. For example, McRoy and Hirst (1995) can make repairs only by explicitly abducing specially-axiomatized *events* of misunderstanding.

[7] E.g., even if M somehow overcame the hopeless task of assigning meaningful probabilities to *all* the beliefs in (1).

4.2 Felicitous use of non-mutual information

The interaction in Figure 2 also illustrates the ubiquity of utterances that seem perfectly acceptable, yet exploit for their interpretation facts that are not mutually believed. Utterance D_3, *the diamond is red*, is such a case here. Objective context captures such utterances directly. Utterance D_3 is contextually acceptable because its contextual requirement `rhombus(t)` is in fact satisfied in the objective context, despite *M*'s misrepresentation of that context. On the mutual belief view, however, D_3 looks like a case of *presupposition failure* given the ground truth context, and a special explanation is required for why the utterance is felicitous and how it changes its context.

4.3 Identifying the true context

The mutual belief model poses the challenge of identifying in practice what the correct set of mutual beliefs is at any given time. In comparison, our model defines the objective context directly in terms of the interlocutors' prior communicative intentions. As illustrated in (3), modeling communicative intentions within an application domain requires connecting words to desired domain entities like ongoing subtasks, intended referents, and domain actions. Of course, we need such a model anyway—for example, even to accurately characterize the potential for a misunderstanding like that in Figure 2. Fortunately, an external observer can construct such a model by examining the utterances that interlocutors use as they perform real-world tasks, without access to their higher-order attitudes. Thus, our approach to context exploits representations that are independently necessary and situates the facts about context much closer to empirical observations than are the facts about higher-order attitudes.

4.4 Coordination and context

Perhaps the hardest challenge in representing context is understanding how a representation should fit into a more abstract characterization of collaboration. While representing mutual beliefs directly seems to preclude certain errors in collaboration, there may of course be other representations that allow an agent to collaborate equally successfully, or at any rate, effectively enough. From this perspective, we can consider agents that try to represent the *objective* context in two cases: ideal communication, and cases of miscommunication and/or uncertainty. In ideal communication, every utterance is actually understood exactly as intended, and both speaker and hearer are perfectly certain that this is so. In this case, not only does each interlocutor privately track the objective context correctly, but each is certain that the other does as well, and further that the other is certain that *they* do, and so on. Provided the speaker and hearer are non-deceptive and trust each other, they will achieve mutual belief.[8]

In cases of miscommunication or uncertainty, their private representations of objective context will differ, and mutual belief will not generally obtain. However, each interlocutor will have a clearly interpretable, practical, uncertain representation of what their prior communicative intentions have been. This means the interpretations they assign to utterances in context will be defensible in terms of these prior intentions. In our view, this highlights interlocutors' ability to target the utterance context implicitly established by their prior conversational activity and to work to make contextual information mutually believed. Thus we can see mutual belief as a desirable but contingent *outcome* of the interlocutors' interaction, rather than as a precondition for it, or as the moment-to-moment target of their representations (Thomason et al., 2006).

Compare this perspective with recent work by researchers pursuing robust human–machine dialogue, who have found it practical to simply identify "context" with the *user*'s state (Roy et al., 2000; Horvitz and Paek, 2001). While this enables coherent probabilistic reasoning, it abandons the role of context as a grammatical resource linking meaning to interpretation and as a mechanism for coordinating dialogue. Our view shows how to keep intuition and implementation aligned.

5 Conclusion

The view of utterance context we have proposed yields simpler representations and reasoning than does the mutual belief model of context. At the same time, it enables straightforward statistical reasoning about context, and offers clearer guidance about what context representations a practical system should have, and how to develop them.

In the end, of course, an interlocutor's uncertainty is pervasive: it affects not only the interpretation of individual words, but also the games

[8] or mutual supposition, etc.

(like "collaborative reference") that other interlocutors play, including the contextual actions those games contain. Fortunately, by connecting utterance interpretation to the objective effects these games and actions have on the context, a language speaker can exploit linguistic experience to reduce uncertainty about them. Interlocutors try, in concert with their other goals, to minimize uncertainty and avoid misunderstandings. When they succeed, mutual belief may be achieved. But by adopting an objective view of context, we can understand how interlocutors proceed on sound footing in any case, and can more transparently design systems that will do the same.

Acknowledgments

This work was supported by the Leverhulme Trust, Rutgers University and NSF HLC 0308121. We thank our anonymous reviewers, Alex Lascarides, Chung-chieh Shan, Mark Steedman, and Rich Thomason.

References

N. Asher and A. Lascarides. 2003. *Logics of Conversation*. Cambridge.

N. J. Blaylock. 2005. *Towards Tractable Agent-based Dialogue*. Ph.D. thesis, Rochester.

H. H. Clark and C. R. Marshall. 1981. Definite reference and mutual knowledge. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge.

H.H. Clark and D. Wilkes-Gibbs. 1990. Referring as a collaborative process. In P. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 463–493. MIT.

H. H. Clark. 1996. *Using Language*. Cambridge.

P. R. Cohen and H. J. Levesque. 1991. Teamwork. *Nous*, 25:11–24.

D. Dennett. 1989. *The Intentional Stance*. MIT.

C. Gauker. 1998. What is a context of utterance. *Philosophical Studies*, 91:149–172.

J. Ginzburg. 1996. Interrogatives: Questions, facts and dialogue. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*. Blackwell.

B. J. Grosz and S. Kraus. 1996. Collaborative plans for complex group action. *AI*, 86(2):269–357.

A. Gruenstein, L. Cavedon, J. Niekrasz, D. Widdows, and S. Peters. 2004. Managing uncertainty in dialogue information state for real time understanding of multi-human meeting dialogues. In *Proceedings of Catalog*.

E. Horvitz and T. Paek. 2001. Harnessing models of users' goals to mediate clarification dialog in spoken language systems. In *User Modeling Conference*, pages 3–13.

D. Kaplan. 1989. Demonstratives. In J. Almog, J. Perry, and H. Wettstein, editors, *Themes from Kaplan*, pages 481–563. Oxford.

S. Larsson and D. Traum. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *NL Engineering*, 6:323–340.

D. K. Lewis. 1969. *Convention: A Philosophical Study*. Harvard.

David Lewis. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359.

S. W. McRoy and G. Hirst. 1995. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21(4):435–478.

M. Poesio and D. R. Traum. 1997. Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309–347.

C. Rich, C. L. Sidner, and N. Lesh. 2001. Collagen: Applying collaborative discourse theory to human-computer interaction. *AI Magazine*, 22(4):15–25.

N. Roy, J. Pineau, and S. Thrun. 2000. Spoken dialogue management using probabilistic reasoning. In *Proc. of ACL*, pages 93–100, Hong Kong.

S. Schiffer. 1972. *Meaning*. Oxford.

R. Stalnaker. 1974. Pragmatic presuppositions. In *Context and Content*, pages 47–62. Oxford.

R. Stalnaker. 1978. Assertion. In P. Cole, editor, *Syntax and Semantics 9*. Academic Press.

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

J. Taylor, J. Carletta, and C. Mellish. 1996. Requirements for belief models in cooperative dialogue. *User Modeling and User-Adapted Interaction*, 6(1):23–68.

R. H. Thomason, M. Stone, and D. DeVault. 2006. Enlightened update: A computational architecture for presupposition and other pragmatic phenomena. To appear in Byron, D., Roberts, C., and Schwenter, S., eds, Presupposition Accommodation.

R. H. Thomason. 1990. Accommodation, meaning, and implicature. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, pages 325–363. MIT.

D. R. Traum. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. thesis, Rochester.