An implementable semantics for utterance context

Abstract

Rich representations of context increasingly drive dialogue systems, yet it remains difficult to assess whether those context models are good or bad, right or wrong, or used correctly within a system. We think the problem lies in the received views of context in pragmatic theory, which characterize the occurrent subjective states of interlocutors. These views are both wrong and unhelpful for system-building. We present a new characterization of utterance context as both objective and normative. This view is intuitively clearer, more directly informs the design of practical representation and reasoning, and introduces criteria that will support substantive evaluation of implemented models.

1 Introduction

Dialogue systems depend on their representations of context. Systems can use them to link utterances to what the user might be trying to do (Allen et al., 2001; Rich et al., 2001), to pursue courses of action that meet their obligations to users (Traum and Allen, 1994), and to support clarification and other flexible communicative strategies (Purver, 2004; Ginzburg and Cooper, 2004).

Managing such sophisticated resources is a methodological challenge. Teams need principled ways to circumscribe their context representations and link modeling decisions to domain requirements. Yet models of context and interpretation remain very difficult to evaluate quantitatively, especially during the design and implementation of a new system (Thomason and Jordan, 1995). Consequently, many researchers view rich context as simply incompatible with effective system-building.

We think the problem lies with received views of context in pragmatic theory. In particular, we believe it has been a mistake to adopt a semantics for context representations in terms of the occurrent subjective states of interlocutors, as captured by technical notions like *mutual belief* (Schiffer, 1972) or even *mutual supposition* (Thomason, 1990). This semantics cannot be realized in systems, because we can't determine the "ground truth" about mutual belief (and interlocutors' attitudes generally) for human-human or human-computer dialogues. This makes it impossible to build models of context based on empirical observations, or to evaluate models of context based on their compatibility with a corpus.

In this paper, we present a new characterization of utterance context as both objective and normative. While systems might aim to *achieve* mutual supposition to avoid misunderstanding (Clark and Marshall, 1981), their context representations, we will argue, should not *mean* that propositions are mutually supposed. Instead, context should be seen as an objective *product of action* taken by the interlocutors; this objective product is what system representations should target. In Section 3, we use a clear analogy with the game of chess to illustrate what we have in mind, and then discuss an implementation of our approach in a dialogue system, COREF, that collaboratively identifies visual objects with human users. Examples of misunderstanding and uncertainty highlight the differences and advantages of our approach.

On our semantics, an external observer can determine how context is to be represented by examining the utterances that interlocutors use as they perform real-world tasks, without access to their mental states. This yields much clearer guidance about what context representations a practical system should have; it enables simpler models of contextual inference, including straightforward ways to accommodate probabilistic reasoning about the current context; and it introduces criteria that promise to allow us to assess objectively whether context models are good or bad, right or wrong, or used correctly within a system. These contributions make the new account a significant result for computational linguistics.

2 Context and mutual belief

On theoretical grounds, we know that agents must *have* mutual belief[1] to solve coordination problems (Lewis, 1969), to interpret definite references in conversation (Clark and Marshall, 1981), or to participate in multi-agent collaboration (Cohen and Levesque, 1991; Grosz and Sidner, 1990; Grosz and Kraus, 1996), at least in certain cases. At an intuitive level, when a belief is mutual between S and H, not only do S and H privately believe it, but they each believe the other believes it, and so on, so they can safely take it for granted. One of the first and most widely known definitions of mutual belief is due to Schiffer (1972):

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

The definition records an infinite, hierarchical interrelation between a speaker and hearer's private beliefs, which are represented by the modal operators $\mathtt{B}_S$ and $\mathtt{B}_H$, respectively, about some proposition p.

If agents sometimes need to *have* mutual attitudes, must their representations therefore *describe* mutual attitudes? Of course not: there's a huge gap between the *conditions* rational system behavior depends on and the *meaning* of the underlying representations (Dennett, 1989). Nevertheless, it's common to assume that dialogue system context representations *do* describe mutual attitudes—see, for example, (Traum, 1994; Poesio and Traum, 1997; Rich et al., 2001; Blaylock, 2005). On this approach, the developer intends the agent's representation of context fundamentally to characterize the mutual beliefs of the agent and its human user. It is against this "common ground" of mutual belief that utterance interpretation occurs, in the hope of avoiding potential misunderstandings (Clark, 1996).

Assuming a mutual belief semantics for context poses a serious methodological problem: for many propositions p, we have no way to tell during system development whether (1) will hold. In fact, linking context to attitudes at all is problematic for system-building. Once we move away from private representations of the world to higher-order beliefs like (1c,d) and (1e,f) there is insufficient evidence for a developer to make principled decisions *herself* about whether such beliefs will obtain, much less automate these decisions. What will the user believe about a new agent's beliefs? What will the user believe the agent believes the user believes?[2]

The absence of any corpus of "ground truth" about higher-order attitudes means, in practice, that researchers leave the meaning of the context representations of implemented dialogue systems relatively unspecified; see, e.g. (Allen et al., 2001; Purver, 2004). We cannot try to represent what's true, so we simply include x in the context whenever doing so leads to some desired system behavior y. The fact that we can no longer judge whether a given representation of context is *correct* is a real drawback to the "semantics-free" approach. In effect, we can no longer analyze success or errors in agent behavior except in terms of the agent's entire decision-making process and all the complex and

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

[2] (Perrault, 1990) does offer one formal model, using default logic, for how mutual belief about prior asserted content can be inferred in conversation, but the model depends on such strong assumptions and elaborate proofs that it would be difficult to exploit it for practical implementation.

heterogenous representations involved. This pressures toward a "global optimization approach" in development, where the meaning of the context representation effectively changes with each new interaction in the code that uses it. While a global approach can lead to rapid coverage in the early stages of development, we have found it to be unmanageable as a strategy for scaling up system capabilities in the long term, and a significant impediment to a persuasive evaluation.

3 Objective, normative context

Like many others (Thomason, 1990; Poesio and Traum, 1997), we view context as a *product of interlocutors' coordinated activity*. But we believe that this context is best understood as an objective feature of the real world, which does not directly depend on the attitudes of interlocutors. Agents' representations simply aim to capture the real context.

This probably sounds crazy. But in fact, lots of everyday activities *must* work this way. Correspondence chess is a good example. Let us suppose two players take it in turns to send each other their moves by email. (Chess moves come with a formal notation, so indeed a computer could parse the email, recognize the moves and track the resulting play.) 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 moves as they are made. But actually two ambitious players could use only their emailed moves and their imaginations to play chess. In what follows, we use 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) and use a case of misunderstanding to show how this vocabulary improves on 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 $c = \langle t, p_1, p_2, \ldots, p_{32} \rangle$ recording whose turn t it is to move next (one or the other of the players) and the current position p_i of each of the 32 chess pieces (either some board position or "captured"). 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 $\mathcal{A} = \{\texttt{advancePawnOneStep}(P), \texttt{moveQueen}(Q, Pos), \texttt{castle}(R), \ldots\}$. Each move a in the game is $\sigma(\alpha)$ where 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:

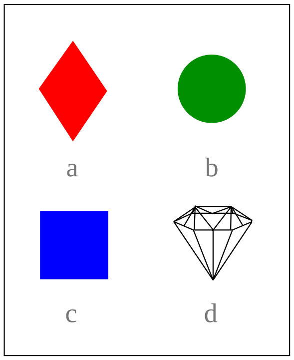
$$c_{t+1} \leftarrow \texttt{update}(a, c_t). \tag{2}$$

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 $\langle a_1, a_2, \ldots \rangle$ plays out over email messages. It is by doing so that valid and advantageous next moves can be selected.

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' mental states. The context is *objective* in the sense that, 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 does not depend on the subjective states of the players. 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. Nevertheless, the status of all the pieces is a matter of *fact* that follows from the moves that have actually been made.[3] 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.

We will argue that to define utterance context in terms of interlocutor beliefs (or mutual beliefs) is misleading in exactly the same way, with exactly the same sorts of detrimental consequences.

[3] In case of a dispute, if the email logs were available, the actual chess state could be settled by examining the move history.



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.

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. 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 dialogue context as an abstract, objective representation, analogously 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).

This context evolves over the course of the dialogue through the domain-dependent set of action types, $\mathcal{A}$, that interlocutors draw on. As in chess, each action $a = \sigma(\alpha)$ for $\alpha \in \mathcal{A}$ must have a *deterministic* effect on the current context, which we again capture by an update function as in (2). In dialogue, these actions may insert new component entities and/or remove or update existing entities from the context. By representing the objective dynamics of this `update` function, we can implement the update mechanisms of an information state approach to dialogue management (Larsson and Traum, 2000). This grounds the data structures and operations implemented systems use (Traum, 1999; Purver, 2004) *without* characterizing those structures semantically in terms of mental state, as in the current foundation for the information-state approach (Poesio and Traum, 1997).

We have been able to model the initial context in our collaborative reference scenario as empty: as far as the context is concerned, no propositions are initially agreed, no plans are initially underway, etc. (Elsewhere, the initial context could include additional information, such as an initial task, as long as we have clear standards for identifying what that information is.) Meanwhile, our 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.

When a speaker S utters a linguistic form l in context c_t, we view S as intending to signal some sequence of actions on the context by conventional means. The problem of interpretation for an utterance of l in context c_t, then, is to recover the intended action sequence $A_l = \langle a_1, \ldots, a_n \rangle$. Given these actions, the new context is:

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

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))`

Upon successfully interpreting the user's utterance, COREF updates its private model of the context to reflect these actions having occurred.

D_1:		*the target is a diamond*	
	D intended:	`addcr(rhombus(t))`	
	M interpreted:	`addcr(diamondJewel(t))`	
		objective model	**mental state model**
	ground truth:	$[c]$ `rhombus(t)`	$\mathsf{MB}_{D,M}$ (nothing about `t`)
	D private:	$\mathsf{B}_D[c]$ `rhombus(t)`	$\mathsf{B}_D\mathsf{MB}_{D,M}$ `rhombus(t)`
	M private:	$\mathsf{B}_M[c]$ `diamondJewel(t)`	$\mathsf{B}_M\mathsf{MB}_{D,M}$ `diamondJewel(t)`
M_2:		*ok*	
		(same)	(same)
D_3:		*the diamond is red*	
		objective model	**mental state model**
	ground truth:	$[c]$ `rhombus(t)` $\wedge$ `red(t)`	$\mathsf{MB}_{D,M}$ `red(t)`
	D private:	$\mathsf{B}_D[c]$ `rhombus(t)` $\wedge$ `red(t)`	$\mathsf{B}_D\mathsf{MB}_{D,M}$ `rhombus(t)` $\wedge$ `red(t)`
	M private(?):	$\mathsf{B}_M[c]$ `diamondJewel(t)` $\wedge$ `red(t)`	$\mathsf{B}_M\mathsf{MB}_{D,M}$ `diamondJewel(t)` $\wedge$ `red(t)`

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

3.3 What's the difference?

The strengths of our view of context follow from its ability to describe the dynamics of dialogue correctly and in intuitive terms. This is much harder for approaches based on mental state. Consider the COREF dialogue excerpt D_1–M_2–D_3 presented in Figure 2. The figure tracks the evolution of the context 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))`. We want to understand this dialogue by linking the representations and inference the agents use to the events in the dialogue on the one hand and the meanings of the agents' representations on the other. This puts us in the prototypical situation of a system-builder.

What happens, we argue, is that after D_1 is uttered, the intended action `addcr(rhombus(t))` takes its effect. D knows what his intended action was, so his private model of the context is updated correctly. M however comes to believe, erroneously, that `diamondJewel(t)` is in the context. By contrast, on the mutual belief model (or any model of shared attitudes), due to the misunderstanding, mutual belief is neither attained for `rhombus(t)` nor for `diamondJewel(t)`. Thus neither is in the true context, and *both* D and M are mistaken: D believes it mutually believed that `rhombus(t)`, as he intended, while M believes it mutually believed that `diamondJewel(t)`, as M interpreted.

Our story has three advantages. First, by using intended actions to characterize context, our model allows developers to think directly and exclusively in terms of the representations that a dialogue system needs independently for interpretation. Any framework needs an account of communicative intentions; we need them even to accurately characterize the potential for a misunderstanding like this one. And dialogue researchers have long argued that intentional action is fundamental to context change (Thomason, 1990; Poesio and Traum, 1997). Yet that's not what the mental state story says: there context is fundamentally about what interlocutors *think*, not about what they've *done*. In cases like Figure 2 the two notions come apart.

Second, our model better accounts for the meanings of utterances as used in misaligned contexts. When M says *ok* in M_2, intuitively, M accepts *whatever D meant* as part of the new context.[4] On the

[4]Compare *ok* as accepting *whatever M thinks D meant*, or *whatever D meant, but only provided M understood correctly*!

objective model, *ok* can presuppose this content now in the objective context and signal its acceptance, *whether or not that content was understood correctly*. However, on the mental state approach, an utterance of *ok* so interpreted cannot be a case of ordinary context-dependence. When M says *ok* at M_2, since the true context is mediated by M's erroneous private beliefs, what D meant is simply not part of the context. In fact, a mutual belief approach will more naturally say that D's intended update and M's acknowledgment simply do not take.

Third, our model lets us more clearly explain and specify the process of recovering from misunderstandings. In this example, 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 in the visual display. Upon detecting the problem, M can reinterpret D_1 and thus *correct* his private model of the context: M had at first thought the context was $[c]$ `diamondJewel(t)` whereas M now recognizes that the true, objective context was $[c]$ `rhombus(t)`. This allows D_3 to be interpreted as meaning that the target `rhombus` is `red`, as intended. Conversely, provided D did not foresee the ambiguity that led to the misunderstanding, D did nothing wrong in uttering *the diamond is red* in D_3. M signaled understanding of *a diamond* in D_1 by saying *ok*, so D naturally referred to the target as *the diamond* in D_3—again meaning `rhombus`—and felicitously asserted that the target `rhombus` was also `red`.

Compare the mental state 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.

In a sense, the important factor on either approach is whether the hearer recognizes the intended updates to the context. The real error is M's misinterpretation of D_1 as contributing `diamondJewel(t)`. With that in mind, it's striking how little leverage we get from a commitment to context as mutual belief. M cannot change the past and retroactively believe `rhombus(t)` (so that it was mutually believed when D_3 was uttered). In order to respect the mutual belief semantics, a developer must construe the 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. Meanwhile, it seems hard on this semantics to avoid the counterintuitive conclusion that D's utterance of D_3 is just a mistake: its contextual requirement `rhombus(t)` was simply not satisfied when D_3 was uttered.

4 Prospects for evaluation

Effective system-building and substantive evaluation of rich context representations seems to depend on three developments. First, we need a corpus of ground truth that can serve as a target for system-building and a benchmark for evaluation. Second, we need probabilistic methods that allow agents to proceed in dialogue despite the inevitable uncertainties in their linguistic experience. Third, we need to exploit these resources to derive context representations by clear and meaningful inference tasks. In this section, we show that our objective view of context makes significant progress on all three fronts.

4.1 Establishing a corpus for context

The objective model suggests that building a corpus of ground truth for context within a domain reduces to associating individual utterances with well-defined actions. It might be suspected that we are really in no better position to construct such a corpus than we were with the mutual belief model, since the relevant actions are *intended* actions, and a speaker's *intentions* might appear no more accessible than his *beliefs*. However, the intended actions on our model are very closely linked to real-world activity for which there is relatively clear evidence in dialogue. For example, in COREF's analysis of U_2 in Figure 1 as intending the actions in (3), the relevant decisions for annotation are whether U_2 initiates a clarification subtask (it clearly does) and whether the clarification subdialogue can be modeled as an embedded collaborative reference. This latter is a substantive question, to which clear evidence applies. In particular, a corpus of clarification subdialogues (Purver, 2004) can be inspected to evaluate the adequacy of the collaborative reference

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.

task model (which adds constraint networks to the context and uses a special purpose `addcr` action to manipulate them). The availability of adjudicating evidence makes questions about how to model intended actions and subtasks considerably less onerous than determining arbitrary private interlocutor beliefs. Understanding context in these terms helps make a "context corpus" comprehensible.

4.2 Private uncertainty about context

In the literature on context, it has been common for discrepancies between the contexts believed to obtain by two interlocutors to be marginalized.[5] Yet in computational models of interpretation, some degree of uncertainty about what an utterance means is the *norm*, so some discrepancies will be unavoidable. We thus aim for a model of context that supports a clear understanding of both misunderstandings and uncertainty. In practical applications, alternative interpretations are typically assigned probabilities, 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:

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

The probabilities capture M's uncertainty about how D's intended action in D_1 changed the context. Evidence appearing across multiple utterances, such as the recognition of a misunderstanding by M after D_3 in Figure 2, can now reduce uncertainty about the true context in a straightforward manner.

[5] 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 argues that interlocutors are sufficiently motivated to avoid misunderstandings that nondefective contexts can be treated as the normal case.

On a mutual belief approach, however, there is simply no route to an internal model of context analogous to (4). In modeling uncertain mutual belief, we can either treat belief as bivalent throughout or probabilistic throughout. If belief is bivalent, then upon hearing D_1, 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 $P(\mathsf{MB}_{D,M}\ \texttt{rhombus(t)}) = 0$, since M's own lack of belief rules it out. M then ends up with this private model:

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

Thus M awkwardly assigns probability 0.4 to the "no mutual belief" scenario, even though this corresponds neither to what D intended nor to how M interpreted D.

If belief is probabilistic throughout, then this presumably applies not only to M's first order beliefs but also to the entire infinite hierarchy of beliefs in (1). So context can never be summarized as in (4). Instead we are in the methodologically hopeless position of trying to assign probabilities to a vast array of higher-order attitudes: such things as M's belief that D believes that M believes `rhombus(t)` after D_1.

4.3 Learning from the context

The objective view of context provides a normative target that can help developers frame clear learning problems for their agents. In a game of mental chess, suppose player p receives an email containing a hitherto unfamiliar chess move by his opponent: `castle(whiteRook1)`. Though p has not seen this move before, p trusts that the true game state is a matter of fact. This allows p to put his uncertainty about what the move has done into straightforward contact with his future experience. For example, if p finds his queen subsequently captured by his opponent's rook, this reduces p's uncertainty about where the move left the rook. Thus, *p learns* something about what this new move does by *playing chess*.

The objective view brings the same clarity to utterance context. Returning to the dialogue excerpt in Figure 2, suppose now that M understands D_1

correctly, as `addcr(rhombus(t))`, but is unsure exactly what the effects of the `addcr` action are. In particular, *M* is not sure whether `addcr(rhombus(t))` adds `rhombus(t)` to the conversational record, making it available for presupposition. Again, *M* can trust that there is some matter of fact what the utterance context is, and use subsequent experience to reduce remaining any uncertainty about it. For example, when *D* utters D_3, *the diamond is red*, which *M* can only interpret as presupposing `rhombus(t)`, *M* gains evidence that `addcr(rhombus(t))` did in fact make `rhombus(t)` available for presupposition. Thus, *M learns* something about the action `addcr(rhombus(t))` by *using language*.

On a mutual belief view, the corresponding learning problems are more complicated to describe, harder to interpret, and implicate private beliefs that remain methodologically inaccessible.

5 Conclusion

Computational linguistics needs a methodology for the effective development and evaluation of dialogue systems that exploit rich models of context. We have argued that achieving this goal requires that we abandon the mutual belief model of context. Instead, we have shown that context can and should be seen as an objective, normative product of interlocutor action. In dialogue, 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 implement agents that proceed on sound footing in any case, allowing developers to accumulate evidence that can be used to improve contextual models, frame clear learning problems, and evaluate their progress.

References

J. Allen, G. Ferguson, and A. Stent. 2001. An architecture for more realistic conversational systems. In *IUI*.

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. R. 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.

J. Ginzburg and R. Cooper. 2004. Clarification, ellipsis and the nature of contextual updates in dialogue. *Linguistics and Philosophy*, 27(3):297–365.

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

B. J. Grosz and C. L. Sidner. 1990. Plans for discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 417–444. MIT.

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

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

C. R. Perrault. 1990. An application of default logic to speech act theory. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, pages 161–185. MIT.

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

M. Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, King's College.

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

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

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

R. H. Thomason and P. Jordan. 1995. Empirical methods in discourse: Limits and prospects. In *1995 AAAI Workshop on Empirical Methods in Discourse*.

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

D. R. Traum and J. F. Allen. 1994. Discourse obligations in dialogue processing. In *ACL*, pages 1–8.

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

D. R. Traum. 1999. Computational models of grounding in collaborative systems. In S. E. Brennan, A. Giboin, and D. Traum, editors, *AAAI Fall Symposium on Psychological Models of Communication*, pages 124–131.