Title
Limiting nested beliefs in cooperative dialogue

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 16(0)

Authors
Taylor, Jasper A.
Carletta, Jean C.

Publication Date
1994

Peer reviewed

Limiting nested beliefs in cooperative dialogue

Jasper A. Taylor
Human Communication Research Centre
2 Buccleuch Place
Edinburgh EH8 9LW, UK
jaspert@cogsci.ed.ac.uk

Jean C. Carletta
Human Communication Research Centre
jeanc@cogsci.ed.ac.uk

Abstract

Models of rationality typically rely on underlying logics that allow simulated agents to entertain beliefs about one another to any depth of nesting. We argue that representations of individual deeply nested beliefs are in principle unnecessary for any cooperative dialogue. We describe a simulation of such dialogues in a simple domain, and attempt to generalize the principles of this simulation, first to explain features of human dialogue in this domain, then those of cooperative dialogues in general.

We propose that for the purposes of cooperative interaction, the status of all deeply-nested beliefs about each concept can be conjoined into a single represented value, which will be affected by reasoning that might be expected to lead to conclusions in terms of deeply-nested beliefs. We concede that people are capable of using individual deeply-nested beliefs to some degree, but such beliefs need only be handled explicitly in dialogues involving secrecy or deception.

Introduction

Currently the most popular paradigm for the analysis of cooperative dialogues is that of viewing communication as plan-based rational action. This approach has been applied to straightforward instructional dialogues (Power, 1974; Cohen & Levesque, 1980), to the generation of appropriate referring expressions (Appelt, 1985), to helpful responses (Allen, 1983), to risk-taking and recovery (Carletta, 1992) and to indirect requests and deception (Taylor, 1994). All these systems model utterances as being components of plans required to achieve certain communicative effects.

Since one of the effects of any utterance is to change the state of mind of the addressee, a system for planning utterances must be capable of representing the beliefs of its conversant, i.e., it must be a model of an agent with *nested beliefs*. The comprehension of utterances can be modelled as plan recognition; if an agent's plan refers to nested beliefs, then another agent's recognition of this plan must refer to beliefs at the second level of nesting, i.e., the addressee's beliefs about what the speaker believes the addressee to believe. Taylor's CYNIC (1994) models deception by allowing agents to plan for their intentions to be recognized in a particular way; such examples employ further levels of belief nesting.

This paradigm encourages one to consider that an appropriate model for an agent engaged in dialogue must include a belief database capable of representing nested beliefs to any required degree. Hintikka's (1962) epistemic logic has been used as the basis of a number of languages of belief fulfilling this requirement (see Fagin and Halpern, 1988). The most flexible of these models are those based on nonmonotonic inference systems, e.g., Perrault's (1990), which are capable of describing what happens when an utterance fails to achieve its desired effect.

Such models, however, do not seem to correspond to what actually goes on during a typical cooperative dialogue. According to Grice (1957), proper comprehension of an utterance is only achieved when the speaker and hearer share a *mutual belief* about what the speaker intended to communicate. Mutual belief, according to Grice and others, is the state where two conversants each believe something, believe one another to believe it, believe one another to believe themselves to believe it, and so on to any degree of nesting. Clark and Marshall (1981, hereinafter referred to as C&M) discuss examples such as the following: Ann and Bob have spoken about the movie they believe will be showing, but later Ann pencil-marks an article in the paper announcing that the movie has been changed. Later still, Bob picks up the paper and looks at the marked article, and Ann sees him doing so but is not herself seen. Following this sequence of events, Ann knows that Bob knows that Ann knows that the movie has been changed, but they do not mutually know it, and in such cases Anne cannot refer to "tonight's movie" unambiguously when talking to Bob. C&M argue that since no finite number of nested beliefs are operationally equivalent to mutual belief, the latter must be established directly, through events in which both agents participate ("co-presence events", such as utterances in dialogues) or by awareness of a common background.

It is apparent that humans have difficulty reasoning about individual deeply nested beliefs. C&M's examples require careful thought, as does the "spies" example from CYNIC.[1] True, everyday conversation can be analysed in terms of planning and plan recognition as forms of theorem-proving in a fully expressive logic of belief. But as long as we are content to have our actual plans reco-

[1] Sergei: "The secret weapon is in Dneprpetrovsk." Natasha: "Ah, I guess the secret weapon *is* in Dneprpetrovsk. You wouldn't say that unless you wanted me to think it was in Krasnoyarsk!"

gnized, we need not concern ourselves with deeply nested beliefs. Only when practicing deception do we plan for our interlocutors to recognize some plan which we wish them to ascribe to us. In cooperative dialogues, the requirement to reason about plan recognition only arises when misunderstandings occur; in such cases, conversants can avert the need for deeply-nested reasoning by strategies such as directly expounding their own beliefs, and directly questioning the partner's beliefs. We contend that people also use such strategies in situations such as the C&M examples, where private deeply-nested beliefs can be inferred from what has taken place.

The argument of this paper is that the beliefs that play a part in cooperative dialogue can be represented using only three explicit levels of nesting, which together implicitly represent mutual belief. In the next section we shall describe the HCRC Map Task Corpus (Anderson *et al.*, 1991), and our simulation of the way people tackle the task. This is based on Carletta's JAM (1992), in which agents are modelled as having only beliefs down to a fixed level of nesting. Subsequently we will illustrate how even if the domain is extended to include potentially ambiguous maps, only three explicit levels of nesting are needed for agents in the simulation. Finally we shall argue that in domains in which the information available allows the inference of more deeply nested private beliefs, a model with three levels will give results characteristic of human performance in cooperative situations.

Simulation of Map Task dialogues

The domain of this model is the HCRC Map Task (Anderson *et al.*, 1991), an experimental task in which two participants are each given maps, on one of which is marked a route. The maps consist of labelled pictographical landmarks, and purport to represent the same terrain, but differ with respect to some of the landmarks and some of their labels. The participant whose map has the route marked (the instruction giver, or IG) attempts to communicate it to the other (the instruction follower, or IF) as accurately as possible. One of the objectives in the design of the task was to explore how people cope with situations in cooperative problem-solving in which the participants' beliefs are initially at variance. We feel that despite the relative simplicity of the problem domain, the Map Task is typical of cooperative dialogue domains in respect of the nestings of beliefs required during problem solving. Real world situations allow conversants to make use of a variety of experience-based and deductive techniques, but ascribing beliefs to one's partner is done through the same mechanisms of co-presence, shared world knowledge and nested reasoning (putting oneself in one's partner's shoes). These techniques underly systems that model much more complex domains, such as Kobsa's (1990) financial advisor.

The standard approach to modelling task-oriented dialogues involves planning and plan recognition based on nested beliefs, e.g., Pollack (1990), Grosz & Sidner (1990). Our simulation is based on Carletta's JAM (1992), in which agents reason about the content of the dialogue using nested beliefs. In the simulation, a participant need only explicitly represent propositions relating to:

- What is on the maps, i.e., what is on her own map plus what she has worked out to be on her partner's map (object level beliefs)
- What her partner thinks is on the maps (first level nested beliefs),
- What her partner thinks she thinks is on the maps (second level nested beliefs).

In our model, propositions of the last two types are implicitly taken to be themselves the subjects of mutual beliefs. For example, if Carol thinks Tom has a swamp, then Carol thinks Tom thinks Carol thinks Tom has a swamp, and so on. They correspond to what Kobsa (1985) calls "infinitely reflexive" beliefs, and can be combined with object level beliefs to produce full mutual beliefs about domain concepts. Kobsa's BGP-MS (1990) user model allows individual nested beliefs to be inherited from categories of nesting levels, but the three types described above are referred to as primitives and always distinguished from each other.

The system uses the notions of *concepts* and *vividness* to describe the information exchanged in map task interactions. These ideas were introduced in JAM, and were shown to be reducible to propositional information by Taylor (1993). Basically, items on the maps (landmarks or route sections) correspond to concepts in agents' internal states, and having a concept vivid means knowing where the item is (or would be) on one's own map, although actual positional information is not used. JAM also introduces the idea of *awareness* as a lower degree of acquaintance with a concept than vividness. Awareness is also reducible to propositional knowledge, namely that there exists an item to which an expression refers.

Agent architecture for the simulation

Our simulation is based on JAM, and retains the basic design in which agents are constrained to speak in turns, and respond to one another's utterances before applying their communicative strategies to initiate further exchanges. It has been altered to use an ATMS-based belief network and a principled representation of the effects of utterances (Taylor, 1993), but retains the fairly standard planning and plan recognition systems of the original. Utterances are actions with conditions in terms of the speaker's state of belief, but once they occur they are themselves subjects of mutual beliefs, and thus mutual beliefs about the speaker's state of belief can be derived from them, e.g., after an indiciative utterance, the agents mutually believe the proposition expressed. These inferences correspond to some of the effects of utterances as modelled in Kobsa's VIE-DPM (1985).

The simulation can be set up to allow the representation of distinct nested beliefs down to any given level; we have experimented with both fewer and more levels than the three listed above. When an utterance occurs, the resulting nested beliefs are added to the database down to the depth limit in force, and any other resulting belief revisions are propagated through the network. The planning and plan recognition components of the system can query the database about beliefs at any level of nesting, but when the level is deeper than the nesting limit the result will be extrapolated from shallower beliefs using the rule that propositions at the deepest two explicitly represented levels are taken to be mutually known.

Agents can hold disjunctions of beliefs (represented by different environments within the ATMS) and believe in logical implications (represented by justifications within the ATMS). Apart from the limitation on nested beliefs, the model follows a KD45 (weak S5) modal logic of belief.

Utterances in the simulation

Agents are characterized by their choice of communicative strategies. With regard to the introduction of referring expressions, the IG has two strategies to choose from. In the "safe" strategy, he mentions landmarks until enough have been established as mutually known to support descriptions of route sections or other landmarks. This produces dialogue such as the following:

Example 1 :
IG: "Do you have the palm beach?"
IF: "Yes."
IG: "Do you have the swamp?"
IF: "Yes."
IG: "The first section goes between the palm beach and the swamp."

This strategy never requires combinations of nested beliefs other than the types that are present in the initial situation. As long as all the proposed references are accepted, their mutual vividness is added directly for both agents, and descriptions using them are always understood. If a reference is proposed and not agreed, the IG knows the IF is only aware of it, and the IF knows the IG has it vivid, i.e., it has the same status as an undescribed route section.

Alternatively, the IG can describe sections using new landmarks and provide additional descriptions of these sections or landmarks if the first descriptions are not accepted. An example of dialogue produced by this "risky" strategy is:

Example 2 :
IG: "The first section goes to the left of the swamp."
IF: "I do not understand."
IG: "The swamp is below the waterfall"

These strategies can be mixed, and a characterization of the effects of meta-statements such as "I think we're in trouble here" might include bringing about a shift in strategy.

Generating cooperative responses

Both these strategies stick to the nested belief types already mentioned, but the latter may force the IG to keep track of a lot of different possibilities regarding the causes of a failure. Can utterances referring to more deeply nested beliefs by either agent improve the situation? In the simulation, one of the strategies available to the IF is to apply plan recognition to any utterance by the IG. There are two cases where this makes a difference to the dialogue. Firstly, when the IF rejects a description, he might recognize the IG's plan to get him to have the subject of the description vivid. The IF can then himself plan to get vivid the referents in the description with which he has a problem. This results in his asking questions about them, e.g.,

Example 3 :
IG: "The first section goes between the palm beach and the swamp."
IF: "I do not understand. Where is the swamp?"

These produce mutual belief that the IF is only aware of the queried landmark, as do rejections of individual mentions, thus reducing the number of disjuncts that need to be considered by the IG. This type of assistance does not need the belief representation to be extended; the IF is recognizing the IG's original plan, i.e., that the IF have the subject of the description vivid. This goal is a second-level belief as far as the IF is concerned, and if only two levels are represented explicitly, the database extrapolates the IF's actual lack of vividness about the subject as if it were mutually known, to conclude that the IG did indeed believe the IF not to have the subject vivid, and his plan was therefore to get the IF to have it vivid. The IF then adopts this plan, and asks about those referents of the original description which he does not have.

Secondly, the IF might indicate when a repair description has resulted in the original description becoming fully understood, thus removing the need for the IG to continue providing descriptions for the referents. Continuing the above example dialogue:

Example 4 :
IG: "The swamp is to the left of the waterfall."
IF: "OK. I am done talking about the first section."

For this case we need to extend the database to represent nested beliefs at the second level explicitly. This is because after the IF has received the repair description, he has the "first section" vivid but does not believe that the IG believes he has it vivid. In order to plan the last utterance to help the IG, the IF's lack of vividness at the second level must be kept distinct from his actual vividness about the subject.

Handling ambiguity

In some of the map task examples, one of the agents has two landmarks on her map with identical names and pictures, only one of which also appears on the other agent's map. On other pairs of maps, landmarks appear in the same place that are different but semantically related, e.g., "swamp" and "crocodiles". Is it possible that one of these map pairs might be better handled by agents using a larger set of nested belief states than that described above?

Reasoning about referring expressions

Let us start with potential ambiguity caused by pairs of identical landmarks. As soon as one agent mentions that she has two of anything, they mutually know that this possibility exists. The "safe" strategy described above must now be made safer, since if one agent asks if the other has a landmark and the other replies "yes", it may not be the same one. In such situations, people may augment their references with descriptions, e.g., "Do you have a swamp near the left hand side?" or "Are there some palm trees close to a wreck?". More usually, however, the participants rely on their mutual belief that any new referents will be to landmarks close to the current route section unless they are explicitly indicated to be otherwise.

The use of proximity or extra descriptions has not been modelled in the simulation, as they do not require any additional nested belief states. However, the possibility must now be taken into account that even when the agents have agreed on a reference, they might each have identified it with a landmark in a different place. So when one makes the first mention of a landmark, the other creates a new concept for it, expressing, e.g., "what my partner means by 'the swamp' ". Such a concept is not the subject of any nested beliefs; there is a first level nesting inherent in its semantics. What he does with this concept then depends on what he has on his own map and on his own strategy. It may be that he cannot identify the new concept with anything on his map, and rejects it or requests a clarification description. What happens in this case is the same as in the nonambiguous map case, except that possibly the clarification given will result in the IF identifying the new concept with a landmark already on his map.

If the addressee accepts the concept, there is still a need for both agents to be able to recover in the event that he has it wrong. So the concept becomes vivid only by virtue of being associated with another vivid landmark, e.g., something actually on his map. The association is itself the subject of a mutual belief, but if the addressee has made the wrong association, their mutual beliefs are wrong. This may not be detected at all, in which case no new types of belief are involved, just wrong ones of the types already described. But a wrong association may lead to an inconsistency, in which case it is first disbelieved at the object level by the agent noting the inconsistency, then renegotiated and either restored or thrown out altogether. This process is described briefly by Taylor (1993).

The agent who originally made the reference can have beliefs about whether or not his partner thinks it refers to the right landmark. These beliefs refer to one of his own landmarks and a concept standing for "that which my partner thinks I am referring to". These are not nested beliefs, although there is a second level nesting inherent in the existence of the latter concept. It might appear that further concepts of this sort, e.g., "That which he thinks I think he is referring to", might be required, but any reasoning process which would change the association of such a concept would also change that of the corresponding concept at the first level, since any subsequent use of that referring expression by the agent who introduced it would have to take into account the fact that this is now the interpretation he expects his partner to make, and he never wants his partner to make the wrong interpretation. This is known by both agents, so neither need have representations of more deeply nested beliefs about referring expressions.

Extending the picture beyond the simulation

In order to argue successfully that more deeply nested beliefs than those described above are in principle unnecessary in descriptions of how humans behave in cooperative dialogue, we must show that the conversational strategies described above are always adequate for situations in which participants need to acquire information about one another's states of mind. The ultimate objective of cooperative dialogue is the establishment of fully mutual beliefs, so we expect people to use utterances that lead directly to full mutual beliefs whenever possible. However, if there is a possibility that a reference has been incorrectly interpreted, conversants may have to refer explicitly to one anothers' states of mind in order to resolve the problem. We contend that wherever a situation allows a misunderstanding concerning a second or deeper level nested belief, recovery strategies are used that attempt to re-establish mutual belief directly by an alternative route rather than patch up the existing set of beliefs.

Utterances referring to deeply nested beliefs

We can use Map Task examples to categorize the ways in which people's private beliefs may become less certain after a mutual belief has been established. If one participant introduces a description and her partner accepts it, they have mutual vividness about what is being referred to. Now the agent who accepted the reference may, due to subsequent inconsistencies, start to believe that she has it wrong. In this case, she is unsure about her private, object-level component of what is otherwise a mutual belief and can ask about it explicitly:

Example 5 : *"I'm not sure I have that swamp after all. Was it near the waterfall?"*

Alternatively, the one who offered the description originally might begin to suspect that the other may have got it wrong, for instance if she subsequently fails to make a connection that should be unproblematic. Here, the offerer has a private, first-level belief that differs from both her object-level belief and the earlier mutual belief. Again, she can explicitly ask about it:

Example 6 : *"Where do you think my palm beach is?"*

The example takes the form of a question as it is infelicitous to make assertions about an addressee's mental states (Searle, 1975).

Both the states of belief in these examples can be represented within the model we have described, but continuing the sequence produces an example involving private uncertainty about a second-level nested belief, which our model would conflate with uncertainty about all deeply-nested beliefs. This would happen if the agent who accepted the description subsequently became concerned that her behaviour might be leading her partner to suspect that she is not sure about the identity of the referent. The utterance for this situation corresponding to the above examples would be:

Example 7 : *"I'm not sure you believe I have that palm beach."*

which, taken literally, asserts that the speaker is unsure whether the addressee thinks the speaker has the concept of the palm beach vivid. There are two problems with this example; firstly, it is difficult to imagine a situation in which someone might come to believe that her conversant thinks she has misunderstood something without suspecting that she actually has misunderstood it, and secondly if the situation should arise, she can resolve the problem with a simpler utterance, reaffirming that she has understood the original referent, such as "I do have that palm beach".

The map task corpus is actually quite poor in examples in which references are incorrectly resolved; comparing proximity to the current section usually results in errors being resolved quickly, and where an error is detected after a significant interval there is a tendency to throw away all work done after the error rather than reevaluate any descriptions relating to the subject of the mistake.

Problems in which deeply nested beliefs arise

As we have said, the initial situation in the map task can be described in terms of belief states that are discriminated by our model. However, some real-life situations may afford the inference of private deeply-nested beliefs, such as happens in C&M's movie examples, in which people cannot make felicitous definite references, despite having series of nested beliefs about the identity of the referent. C&M propose that in making a reference, a condition of full mutual belief is checked, and that this condition is specifically asserted by co-presence events. Any private beliefs later acquired by one agent that differ from the original mutual belief will cause it to cease to be mutual, and make reference impossible. In practice, people in such a situation will often open with a question, such as "Did you know the movie's been changed?"

However, people may use expressions in the absence of full mutual knowledge of their referent if they are confident enough that they know what shared knowledge does exist. If one knows that a cinema programme has been changed, it would seem unreasonable to discount the possibility that one's conversant knows about the change even if there is no direct reason to think that they do. On the other hand, in a situation in which one person knows that the "true" referent of an expression has changed, but can be confident that their conversant does not know this, they can continue to use the expression to denote the previous referent.

For instance, Ann or Bob could say: "The latest Ford Escort does 40 miles per gallon" with no risk of misunderstanding, even if Ann works at Ford and knows there is a top secret prototype of a newer Escort that does 50 miles per gallon. Her confidence in the secrecy of the existence of the alternative referent allows her to use, or infer the use of, an "incorrect" referring expression without problems. In our belief model Ann has an object-level belief that differs from her first- and second-level nested beliefs, which are derived from cultural co-presence of the newest Escort in the showrooms.

Now let us say that unbeknownst to Ann, Bob happens to be an industrial spy and has recently paid a clandestine visit to the Ford factory, on which he discovered the existence of the prototype and Ann's involvement with it. Confident that he has not been detected, he can understand Ann's reference as meaning the currently available Escort, and could use that reference the same way himself. We now model him as having object-level and first-level nested beliefs that the prototype Escort is the newest, but using his second-level nested belief that the showroom Escort is newest to construct or understand references to it.

But if Ann has watched Bob's visit on a video from a hidden security camera, our model cannot distinguish the second-level nested belief thus made available (she knows Bob knows she knows about the prototype car) from any remaining nested belief that she does not know about it. In our model, such a situation would create an uncertainty at the deepest level of nesting. The result is that Ann cannot refer to either car simply as "the new Escort", and if Bob uses that expression she has to ask him for clarification. In a cooperative dialogue situation, this behaviour is plausible for actual humans; explicitly asking for clarification, or using a more specific referring expression, takes less effort than the reasoning required to establish that "the new Escort" must still mean the showroom model.

Situations involving deception

If the sequence of events in the last example above happened in a non-cooperative situation, the two agents might have goals of not giving away what they knew about one another, and would then use "the new Escort" to refer to the showroom model whilst watching for the slightest sign that the other was in any doubt about its meaning. Such behaviour would require distinct representation of deeply nested private beliefs. The belief state hierarchy in BGP-MS (Kobsa, 1990) allows these beliefs to be distinguished when necessary; we claim that the criterion for their being necessary is the existence of goals involving secrecy or deception on the part of either conversant.

Sometimes, even cooperative agents need to deceive their partners; for example, teachers may oversimplify the truth deliberately in order to keep their tuition simple enough for their students to understand. Physics teachers describe Newtonian mechanics even when they know about Einstein's general relativity, and programming can be taught using a "virtual machine" without mentioning that the actual machine is engaged in a lot of other tasks at the same time. If the pupil in such a situation is none the wiser, our model handles it correctly; the only difference from the non-deceptive situation is that the teacher has private object-level beliefs that differ from what is being taught. We categorize such situations as cooperative because the teacher has no commitment to the deception. If the pupil suspects she is not being taught the whole truth, the teacher can come clean and proceed with the lesson with the mutual knowledge that what is being taught is a simplification.

Conclusion

Belief models based on epistemic logics capable of representing nested beliefs to an arbitrary degree appear to be superfluous in the modelling of cooperative dialogues. In particular, the structure of the map task domain is such that agents taking part in the task do not need to bring to bear on it any nested beliefs at more than one level of nesting, except for full mutual belief. This explains how JAM is able to generate a range of simulated map task dialogues without modelling beliefs at more than two levels of nesting — the second level of nesting in JAM is equivalent to mutual belief between the participants.

Map task situations not handled by JAM, namely maps with ambiguous landmarks or different names for the same landmark, do not require explicit nested beliefs beyond the second level either. This follows from the fact that in cooperative dialogue, agents use referring expressions only when they believe they will be understood correctly, so the referred object is the same in deeply nested beliefs as in shallower ones.

This principle appears to hold in the world of cooperative dialogue in general. Where it is possible to derive individual deeply nested beliefs about the domain, such as in C&M's examples, these tend to be ignored in favour of other possible ways of conveying information, as if there is no requirement for secrecy, conversants can always talk about their actual object-level beliefs.

Successful reference requires that the conversants be co-present at an event that makes mutual knowledge of the referent possible, but it does not take another copresence event to alter the possibility of using a reference. In some circumstances referring expressions can be generated and understood when agents have object-level or first-level nested knowledge that differs from that conferred by the co-presence event. However, only when they are concerned with deception do people differentiate between second-level and more deeply nested beliefs. In other cases, a contradictory second-level nested belief is enough to interfere with the use of a referring expression.

Acknowledgements

The work described here is funded by JCI grant #SPG9200139, running for three years from November 1992. The contributions of the other members of the project team, Chris Mellish and Keith Stenning, are gratefully acknowledged.

References

Allen, James F. (1983). Recognizing intentions from natural language utterances. In Brady and Berwick, editors, *Computational Models of Discourse*, pages 107–166. MIT Press, Cambridge, MA.

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry Thompson, and Regina Weinert. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4):351–366.

Appelt, D. E. (1985). *Planning English sentences.* Cambridge University Press.

Carletta, Jean C. (1992). *Risk-taking and Recovery in Task-Oriented Dialogue.* PhD thesis, Edinburgh University Department of Artificial Intelligence.

Clark, Herbert H. and Catherine R. Marshall (1981). Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, editors, *Elements of Discourse Understanding*, pages 10,63. Cambridge University Press.

Cohen, Philip R. and Hector J. Levesque. (1980). Speech acts and the recognition of shared plans. In J. R. Sampson, editor, *Proceedings of the third biennial conference, Canadian Society for Computational Studies of Intelligence*, pages 263–271, Toronto.

Fagin, Ronald and Joseph Y. Halpern. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76.

Grice, H. Paul. (1957). Meaning. *Philosophical Review*, 66:377–88.

Grosz, Barbara J. and Candace L. Sidner. (1990). Plans for discourse. In Phillip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 417–444. MIT Press, Cambridge, MA.

Hintikka, Jaakko. (1962). *Knowledge and Belief.* Cornell University Press, New York.

Kobsa, Alfred (1985). VIE-DPM: A user model in a natural language dialogue system. In J. Laubsch, editor, *GWAI-84, 8th German Workshop on Artificial Intelligence*, Berlin: Springer.

Kobsa, Alfred. (1990). Modeling the user's conceptual knowledge in BGP-MS, a user modelling shell system. *Computational Intelligence*, 6:193–208.

Perrault, C. Raymond. (1990). An application of default logic to speech act theory. In Phillip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 161–185. MIT Press, Cambridge, MA.

Pollack, Martha. (1990). Plans as complex mental attitudes. In Phillip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, pages 77–103. MIT Press, Cambridge, MA.

Power, R. J. D. (1974). *A Computer Model of Conversation.* PhD thesis, Edinburgh University Department of Artificial Intelligence.

Searle, John R. (1975). Indirect speech acts. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics 3: Speech Acts*, pages 59–82. Academic Press, New York.

Taylor, Jasper (1993). An ATMS-based belief model for dialogue simulation. Technical Report HCRC/RP-45, Human Communication Research Centre, University of Edinburgh, Edinburgh, UK.

Taylor, Jasper (1994). *A Multi-Agent Planner for Modelling Dialogue.* PhD thesis, School of Cognitive and Computing Sciences, University of Sussex, Brighton, UK.