Mutual Beliefs in Conversational Systems : Their Role in Referring Expressions*

Gopalan Nadathur and Aravind K. Joshi

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, Pa 19104

ABSTRACT

Shared knowledge and beliefs affect conversational situations in various ways. One aspect in which they play a role is the choice of referring expressions. It is of interest to analyse this role since a natural language system must be able to decide when it can use a particular referring expression; or alternatively what a particular expression refers to. In this paper we attempt to formally characterise conditions for these. Specifically, we differ with the traditional notion of mutual knowledge and belief, state a conversational conjecture that convinces us to do so, express a weakened notion in a formal system for reasoning about knowledge, and show how this might be used to decide on satisfactory referring expressions. It is desirable to express a weakened notion of mutual belief that parallels that for mutual knowledge; this aspect is currently being investigated.

1.0 INTRODUCTION

Representing and reasoning about the knowledge and beliefs of the participants in an interaction is an important consideration for a system like a natural language interface to a knowledge base, designed to exhibit cooperative behaviour. The ability, for example, to

- understand the primary intent of a query or statement,
- predict the expectations of the participants so as to attempt to satisfy them, or
- understand or use referring expressions appropriately

depend crucially on this aspect of the system. The problem of designing an adequate representation and reasoning mechanism for knowledge and/or beliefs and of using this to produce the desired behaviour, however, is a non-trivial one. First it is not quite clear what an appropriate representation for knowledge and belief in general is. Then, even if one of these representations Is indeed satisfactory, it is not quite clear how the reasoning system is to be built so that it mirrors in some way the reasoning process of humans, rather than doing reasoning that is purely logical. This latter is an important criterion, since a system to be used in a conversational environment must be able to reason about the beliefs of others in the same manner that they would about them.

Insofar as producing cooperative behaviour impinges on the task of understanding and generating natural language, it is already evident that considerations of knowledge and belief are important for natural language processing. However, there are aspects of language processing that are affected more directly by beliefs and knowledge about others and of what beliefs and knowledge are shared with others. Consider for example the aspect of using or understanding a referring expression. In trying to refer to an object one takes recourse to the facts that one thinks or knows one's audience believes or knows; and a system that does not make the assumption that it communicates with people who know all that it knows must be able to do the same.

Thus it is clear that beliefs and knowledge do play a role in understanding and using referring expressions. However, it is not clear what exactly is the role they play. This aspect has been considered earlier in (Clark and Marshall, 1981) and (Perrault and Cohen, 1981). The intuition in these analyses seems to be that in using a definite description D to identify a particular object a, the speaker and his/her audience must mutually know (or believe) that the object a uniquely satisfies the description D.* In our opinion, however, this condition is too strong; in conversational situations, if the speaker knows (or believes) that the hearer knows (or believes) that an object a satisfies the description D uniquely, then he/she often conjectures that this is mutual knowledge (or a mutual belief) if there is no reason to doubt it. Thus a speaker may use the phrase 'the long haired Swede who does not shave during the course of the Wimbledon' to refer to Bjorn Borg even before

* This work was partially supported by the NSF Grant MCS 81-07290. We would like to thank Eric Mays for his comments and suggestions.

* Even though in (Perrault and Cohen, 1981) it is claimed that this need not be a mutual belief, the condition there stated is only a slight weakening of this condition.

he/she knows that it is mutual knowledge that Borg satisfies this description.

We attempt to give formal expression to this idea in this paper. In Section 2 we describe a representation system that includes an axioraatisation of mutual knowledge and of ignorance. In Section 3 we describe how this system can be used to give expression to the conjecture discussed above.

2.0 REPRESENTATION OF KNOWLEDGE

We shall use the representation proposed in (Konolige, 1981) which is based on Sato's K4 (Sato, 1976) and extended to include an axiomatisation of ignorance. K4 is a family of languages parameterised by the choice of propositional letters Pr and agents Sp. For a particular choice of Pr and Sp, K4 is the propositional calculus over Pr together with a set of indexed unary modal operators [S], S £ Sp. The intended meaning of [S]A for any sentence A is that S knows A. There Is a special, fictitious agent 0 £ Sp, used to axiomatise common knowledge; for any sentence A the intended meaning of [0]A is that A is common knowledge. The axiom schemata for K4 are

(Al) All propositional tautologies
(A2) [S]a => a
(A3) [S]a -> [S][S]a
(A4) [S](a -> b) => ([S]a -> [S]b)
(A5) [0]a => [0][S]a

where a and b are arbitrary sentences. K4 includes two rules of inference: modus ponens (I.e. from a and a -> b infer b) and necessitation (i.e. from a Infer [S]a for any S E Sp).

KA as defined above has been shown to be decidable. Konolige uses this fact to explicitly introduce the notion of provability in KA into the logic. The language of KA is extended to include a family of unary modal operators indexed by sentences of KA; these are written as <a> where a is a sentence of KA, and the intended meaning of <a>b, where a and b are sentences of K4, is that a => b is provable in KA. The new language is called K14. The axioms of K1A are then defined to be those of KA together with the schemata

(A6) <a>b where KA I- a «> b
(A7) ~<a>b where KA I- a -> b

The two rules of inference are retained. Since KA is decIdable, (A6) and (A7) form a recursive set. Further for every a and b that are sentences of KA, either <a>b or ~<a>b is a theorem of KIA. Given this and that KA is decidable, it is easy to see that KIA is also decIdable.

The new logical system is of Interest since in it it is possible to partially characterise an agent's ignorance. We can partition an agent's knowledge into two sets - the first corresponds to his/her 'core' knowledge, and the second to the knowledge that is derived from the core knowledge. Let B be the conjunction of the set of sentences that form the core knowledge of the agent S. If now A is [S]B, then if ~<A>C is provable in KI4 it can be concluded that the agent S is ignorant of C. Note that this is only a partial characterisation of S's ignorance; the claim that S knows D if <A>D is deducible in the system is problematic since an agent need not know all the logical consequences of his/her core knowledge.

In the rest of this paper we shall be interested in interactions that involve two participants only. These will be identified as the speaker S and the hearer H. Any reference to KI4 made henceforth will refer to the class of those theories for which Sp = {0,H,S}.

3.0 USING AND UNDERSTANDING REFERRING EXPRESSIONS

In this discussion we shall restrict our attention to one kind of referring expression - definite descriptions. A speaker who uses a definite description referentially may be thought of as doing so to enable his/her audience to pick out an object that he/she is talking about. Our objective is to characterise the conditions under which it is safe to use a description D to refer to an object a.

Let P stand for the proposition : "the description D In the context C is uniquely satisfied by a". In (Clark and Marshall, 1981) it is claimed that in order for the speaker to use D to refer to a, it must be the case that P is mutual knowledge; with respect to KI4 this amounts to saying that [0]P is provable. It is argued in (Perrault and Cohen, 1981) that this condition is too strong; they present a series of examples from which they conclude (ignoring for the moment that they are dealing with belief rather than knowledge) that what is necessary is that in

[S]P ^ [S][H]P A [S][H][S]P A ...

only a finite number of conjuncts should be false.

Neither of these analyses is quite satisfactory. First, despite the claim in (Perrault and Cohen, 1981), it seems that the speaker must know (or believe) that the hearer knows (or believes) that P is true in order to use D to refer to a. If this were not the case there is no assurance for the speaker that the hearer will understand what he/she is referring to by D and hence will make the proper connection; in fact this is our intuition regarding the examples in (Perrault and Cohen, 1981). Second, and most Important, in insisting that P be mutual knowledge (or 'almost' a mutual belief as in (Perrault and Cohen, 1981)) an important conversational convention is being ignored - even If P is not mutual knowledge (or a mutual belief) it is often conjectured to be so given some positive evidence and no negative evidence.

This consideration leads us to state that

<T>[S][H]P A ~<T>[S]-[0]P,

where T is [S] B and B is the conjunction of the

set of sentences that form the core knowledge of S, is a sufficient* condition for using D to refer to a. Intuitively, this condition says that S knows that H knows that P holds; moreover there is no reason to believe that S knows P is not mutual knowledge.

We can see the implication of this condition with reference to the example in Section 1. While the earlier analyses would permit the use of the description for Borg only in case the association between these two was mutual knowledge, this analysis permits the use of the description for Borg given that the speaker knows that the hearer knows the association.

A similar analysis yields the following condition that we think is sufficient for understanding that D refers to a

$$<T'>[H]P \ A \ \sim<T'>[H]\sim[0]P$$

where T' is [H]B' and B' is the conjunction of the set of sentences that form the core knowledge of H.

4.0 ON USING BELIEF

In trying to decide whether a definite description may be used to refer, it is necessary to make the distinction between 'successful reference', where the speaker has all the right Intentions and has conformed to all the right conventions In using a referring expression, and a 'fully consummated reference', where uptake is also secured; the criterion to use is that of 'successful reference', as pointed out in (Perrault and Cohen, 1981). One implication of this is that what is of interest are the beliefs of the speaker, and not his/her knowledge.

A set of axiom scheraas similar to that of K4 has been developed for belief. In order to introduce the notion of provability into the logic (something that Is needed to express the notion of ignorance), it must be shown that the underlying logic is decidable. Since the logic of Interest is a propositional one, we conjecture that it is. However this must be proved. We are currently pursuing this aspect. We believe that in this new formal system the ideas expressed in section 3 can be used to state conditions for the use and understanding of referring expressions that are Intuitively more satisfying.

* This condition is not necessary, because some situations can be constructed where reference may be successful even if it Is not the case that <T>[S1[H]P. (The examples In (Perrault and Cohen, 1981), however, do not seem to be such cases). This conjunct is needed, since otherwise it does not seem warranted to conjecture P as a mutual belief. If this condition is sufficient, then the condition in (Perrault and Cohen, 1981) is too strong, i.e. is not necessary.

5.0 CONCLUSION

In this paper we have considered an issue in which knowledge and belief play an important role, and we have attempted to express certain intuitions in a formal manner. We grant that there are limitations in the formal expression, not the least of them being that we have used a crucial feature of propositional modal logics, that they are decidable. The point however is not so much in the solution advocated here, but in that there is a weaker notion of mutual knowledge (and of mutual belief) that is operative in practice; and It is this that a system that reasons about knowledge and belief in order to produce cooperative behaviour must give expression to.

Although we have limited our attention to only referring expressions, there are other situations in which mutual knowledge and beliefs play a role. One such situation is that where in the course of an interaction the two participants discover that their beliefs are incompatible, and enter into a dialogue to correct this situation. Such interactions have been analysed in (Joshi, 1982) in the context of a question-answer system. The assumption there was that in the end one of the participants beliefs should become a mutual belief as a result of the interaction. This requires very strong assumptions on the nature of discourse. A matter to be studied is whether the weakened notion of mutual belief is sufficient for most situations where the notion of mutual belief is needed at all; and if so, whether this makes more realistic the assumptions required to ensure that a clarification dialogue succeeds.

REFERENCES

[1] Clark, H.H., and Marshall, C, 'Definite Reference and Mutual Knowledge', in Elements of Discourse Understanding (eds. A.K. Joshi, B.L. Webber and I.A. Sag), Cambridge University Press (1981), pp 10-64.

[2] Joshi, A.K., 'Mutual Beliefs in Question Answer Systems', in Mutual Knowledge (ed. N. Smith), Academic Press (1982), pp 181-197.

[3] Konolige, K., 'Circumscriptive Ignorance', Proc. AAAI, Pittsburgh, August 1982, pp 202-204.

[4] Perrault, C.R., and Cohen, P.R., 'It's for your own good: a note on Inaccurate Reference', in Elements of Discourse Understanding (eds. A.K. Joshi, B.L.Webber, and I.A. Sag), Cambridge University Press (1981), pp 217-231.

[5] Sato, M., 'A Study of Kripke-type Models for some Modal Logics by Gentzen's Sequential Method', Research Institute for Mathematical Sciences, Kyoto University, Kyoto, Japan, July 1976.