

# Beliefs and socio-cognitive interaction: belief revision in multi-agent systems

Rineke Verbrugge  
Institute of Artificial Intelligence  
University of Groningen  
e-mail: rineke@ai.rug.nl

## 1 Introduction

I will report on research on group attitudes and communication in multi-agent systems, done in collaboration with Barbara Dunin-Kępicz, as well as research on cognitive limitations on reasoning about other agents with Petra Hendriks, Irene Krämer, and Lisette Mol.

## 2 Changing beliefs by reasoning about others

In everyday situations it is extremely important to reason about other people's minds: you need to reason about their knowledge in order to interpret what they say and to construct your own utterances so as to be understood; you need to reason about their beliefs and intentions in order to negotiate with them; for many situations you even need to reason about their beliefs about your beliefs about them, and about even higher-order beliefs.

Formal models of human reasoning, such as those in epistemic logic and game theory, assume that humans can faultlessly reason about other people's knowledge and about common knowledge, for example in card games such as happy families [9]. However, recent research in cognitive psychology reveals that adults do not always correctly use their theory of what others know in concrete situations [6], [5].

In Keysar's experiments, some adult subjects could not correctly reason in a practical situation about another person's lack of knowledge (first-order theory of mind reasoning) [6]. Hedden and Zhang, when describing their experiments involving a sequence of dyadic games, suggested that players generally began with first-order reasoning. When playing against first-order co-players, some began to use second-order reasoning, but most of them remained on the first level [5].

In recent experiments by AI Master's student Lisette Mol [8], it turns out that humans *can* learn to play a version of symmetric Mastermind involving natural language utterances such as "some colors are right". After mastering

the first task, namely to play the game according to its rules, many of them learn to perform a second task, namely to develop a winning strategy for the game by using a higher-order theory of mind: “Which sentences reveal the least information while still being true?” “What does the opponent think I am trying to make him think?”

## 2.1 Cognitive limitations and common knowledge

If even limited orders of theory of mind present such difficulties for humans, it seems that reasoning about common knowledge, which apparently involves an infinitude of levels, is impossible. From the time when the notion of common knowledge was first studied, there has been a puzzle about their establishment and assessment, the so-called *Mutual Knowledge Paradox*, most poignantly described in [1]. How can it be that to check whether one makes a felicitous reference when saying “Have you seen the movie showing at the Roxy tonight”, one has to check an infinitude of facts about reciprocal knowledge, but people seem to do this in a finite, indeed short, time?

## 3 Changing group beliefs by communication in multi-agent systems

Notions of knowledge about others and group knowledge also play an important role in *multi-agent systems*, where a number of computational agents work together in order to solve a problem that they cannot solve on their own. Indeed common knowledge is seen as the basis of coordination among agents. Halpern and Moses proved that common knowledge of certain facts is on the one hand necessary for coordination in well-known standard examples, while on the other hand, common knowledge cannot be established by communication if there is any uncertainty about the communication channel [3].

### 3.1 Belief and common belief

In practice in multi-agent systems, agents often make do with belief instead of knowledge for the following reasons. First, in multi-agent systems, perception provides the main background for beliefs. In a dynamic, unpredictable environment, the natural limits of perception may give rise to false beliefs or to beliefs that, while true, still cannot be fully justified by the agent. Second, communication channels may be of uncertain quality, so that even if a trustworthy sender knows a certain fact, the receiver may only believe it.

*Common belief* ( $C\text{-BEL}_G(\varphi)$ ) is the notion of group belief which is constructed in a similar way as common knowledge: everyone believes  $\varphi$ , everyone believes that everyone believes  $\varphi$ , and so on, ad infinitum. This is formalized as follows:

$$C1 \text{ E-BEL}_G(\varphi) \leftrightarrow \bigwedge_{i \in G} \text{BEL}(i, \varphi)$$

**C2**  $\text{C-BEL}_G(\varphi) \leftrightarrow \text{E-BEL}_G(\varphi \wedge \text{C-BEL}_G(\varphi))$

**RC1** From  $\varphi \rightarrow \text{E-BEL}_G(\psi \wedge \varphi)$  infer  $\varphi \rightarrow \text{C-BEL}_G(\psi)$  (Induction Rule)

**R2** From  $\varphi$  infer  $\text{BEL}(i, \varphi)$  (Belief Generalization)

Axiom **C2** is often called the fixed-point axiom, showing how  $\text{C-BEL}_G(\varphi)$  can be viewed as fixed point of the function  $f(x) = \text{E-BEL}_G(\varphi \wedge x)$ . Soundness of rule **R2** is proved by induction on  $k$ , to show from the antecedent  $\models \varphi \rightarrow \text{E-BEL}_G(\psi \wedge \varphi)$  that  $\models \varphi \rightarrow \text{E-BEL}_G^k(\psi \wedge \varphi)$ .

Note that, in contrast to common knowledge, which is always sure, common belief need not be truthful, thus in some situations  $\text{C-BEL}_G(\varphi)$  may be a common illusion. The axiom system governing individual and common belief is called  $KD45_n^C$  (see [3, 9, 7] for more about these logics).

### 3.2 Problems in creating common knowledge

Halpern and Moses [4] proved a surprising result in the eighties: under some very natural assumptions, namely that processors do not change their local states simultaneously, common knowledge does not increase over a run (sequence of time steps) in a distributed system. The well-known example of the two generals who do not manage to reach common knowledge about the time of attack, even if a messenger brings any number of acknowledgments back and forth, is an example of this result. If there is any uncertainty about the messenger making it to the other general, even about whether he may be delayed, common knowledge cannot be reached [3, 9]. In multi-agent systems, there is almost always uncertainty about messages reaching the other party.

Note that Halpern and Moses' result does not carry over to common belief. Their proof hinges on the fact that if processors do not change their local states simultaneously, then any two global states in a sequence of time-steps are accessible to each other by a sequence of knowledge-accessibility relations. This is in turn based on the fact that other global states with the same local state are always knowledge-accessible for a processor, a fact that need not hold for belief-accessibility.

### 3.3 A protocol for creating common belief

Even though common knowledge cannot in general be established by communication, we have shown that common belief can [2]. In this context, it turns out to be an advantage that belief, in contrast to knowledge, need not be true. Thus, Halpern and Moses' impossibility results about the growth of common knowledge do not carry over to common belief. Even stronger, it is possible to give a procedure that can, under some assumptions, establish common beliefs.

In multi-agent systems literature, it is usually assumed as a simplification that public announcements are always successful: announcements reach all group members, and in the end their content is commonly believed by the group. Such an assumption takes for granted that the communication medium

is perfect and that no messages are lost, which is not the case in practice. We think it is important to relax this strong assumption of a perfect communication medium.

In the talk, we will informally present such a procedure for creating a common belief in a group, essentially by one initiator broadcasting an appropriate message to all agents in the group [2]. The reason that this procedure can establish common belief, whereas common knowledge can never be established, is exactly that common beliefs need not be true.

The effectiveness of the procedure for creating common belief is based on some weak assumptions about the communication channels, trust among group members and temporary persistence of some relevant beliefs.

## References

- [1] H. H. Clark and C. Marshall. Definite reference and mutual knowledge. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, 1981.
- [2] B. Dunin-Kępicz and R. Verbrugge. Creating common beliefs in rescue situations. In A. Jankowski, A. Skowron, and M. Szczuka, editors, *Proceedings of Monitoring, Security and Rescue Techniques in Multiagent Systems (MSRAS)*, Berlin, 2004, to appear. Springer.
- [3] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, MA, 1995.
- [4] Joseph Y. Halpern and Yoram Moses. Knowledge and common knowledge in a distributed environment. In *Symposium on Principles of Distributed Computing*, pages 50–61, 1984.
- [5] T. Hedden and J. Zhang. What do you think I think you think? Strategic reasoning in matrix games. *Cognition*, 85:1–36, 2002.
- [6] B. Keysar, S. Lin, and D.J. Barr. Limits on theory of mind use in adults. *Cognition*, 89:25–41, 2003.
- [7] J.-J. Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Theoretical Computer Science*. Cambridge University Press, Cambridge, 1995.
- [8] L. Mol. Learning to reason about other people’s minds. Technical report, Institute of Artificial Intelligence, University of Groningen, 2004. Master’s thesis, to appear.
- [9] W. van der Hoek and R. Verbrugge. Epistemic logic: A survey. In L.A. Petrosjan and V.V. Mazalov, editors, *Game Theory and Applications*, pages 53–94. Nova Science Publishers, vol. 8, New York, 2002.