

Higher-order theory of mind, from the logical point of view Rineke Verbrugge

Theory of mind, a particularly important aspect of social cognition, is the cognitive capacity by which one understands and predicts external behavior of others by attributing internal states, such as knowledge, beliefs, and intentions. Everyone uses this type of social cognition daily: when leading a team, deciding whether to send e-mail to a group with the ‘bcc’ option, or deciding how to formulate feedback tactfully. Social cognition and cooperation are essential to success in human life and increasingly essential to modern computer science.

Formal logics of human social cognition tend to assume that our abilities are infinite: that when it comes to attributing mental states to other people, we are capable of an arbitrary amount of recursion. This implies that I can not only think about what you know, but also about what you think I know, about what you want me to think you know... and so onwards, for ever higher orders of theory of mind. However, in practice, even adults tend to lose track of such reasoning after only a few levels. The assumptions made by formal logics are therefore idealisations of our actual skills. In the future, however, such idealisations are likely to prove problematic.

Artificial assistants, in the form of software agents and robots, are often built with formal logics such as epistemic logic as their reasoning systems. They use the rules and axioms of their logics to derive knowledge about the world, and to decide what they should do next. If such artificial assistants are to work together with human teammates, or to support them in complex negotiations, it is important that they do not overestimate the social cognition of their human counterparts. Moreover, when it comes to higher-order social cognition, there is much left to discover. When do people use it? How does it help them make better decisions? What exactly are their limitations?

The talk will include an overview of current research on higher-order theory of mind from cognitive science and logic, including results of studies carried out together with P. Hendriks, L. Mol, L. Flobbe, E. van der Vaart and I. Krämer. Finally, some ideas for future research will be presented, based on a combination of three methods: logic, computational cognitive modeling, and agent-based modeling.