



Who knows? Logic and artificial intelligence for reasoning about others

By Thomas Bolander

ABSTRACT:

Central to human collaboration and communication is our ability to reason about others, that is, to reason about the knowledge, beliefs, intentions and capabilities of those that we collaborate and communicate with. It is reasonable to expect that such abilities can become equally crucial for future robots to achieve efficient collaboration and communication with other robots and with humans. To be able to reason about others, one needs the ability to create mental models that include a representation of the mental models of these others. Such mental models can differ fundamentally in structure depending on whether they are formed by humans or computers. Mental models in computers can also differ fundamentally in structure, depending on which paradigm of mathematics is used to represent them. In my research, I have primarily focused on logical models representing the world in an explicit, symbolic way that easily lends itself to logical reasoning, planning, and verifiable, explainable behaviour. My talk will focus on the definition and use of such models to create artificial intelligence systems with the ability to engage in successful social behaviour via reasoning about others.

My own path into such models went through, first, a deep interest in mathematical aspects of self-reference (how structures can contain a representation of themselves), then models of introspection (how agents – humans and robots – can represent and reason about themselves), and then finally social aspects of artificial intelligence (AI): how AI systems can become socially competent by providing them with the ability to create mental models of others (Theory of Mind). In my talk, I will present some of the deep technical challenges involved, some of the possible solutions, and my expectations concerning the future paths of social aspects of AI.