18

INTELLIGENT AGENTS AND COMMON SENSE REASONING

John-Jules Meyer and Frank Veltman

1 INTRODUCTION

Modal logic plays an important role in the field of artificial intelligence (AI). This can be understood from the fact that AI tries to capture aspects of human intelligence by formalizing these in such a way that they can be implemented in an artificial system, particularly a computer-based system. This entails the need for formalization of mental attitudes such as beliefs and desires. In philosophical logic one has studied many of these attitudes using modal logic. Therefore it is only natural that AI researchers have resorted to modal logic for the formal description of the mental attitudes of their intended artifacts. (This is not to say that modal logic is the only way to represent these mental attitudes, since some choose to stick to classical predicate logic as closely as possible. Still, it is recognized widely within the AI community that modal logic presents a valuable tool!)

So, in this chapter we will be concerned with the use of modal logic techniques to formally describe mental attitudes of intelligent systems. We have chosen to split our

treatment into two parts. The first part deals with the use of modal logic for the description of so-called *intelligent agents*, an area of AI that emerged at the end of the 80s, dealing with the theory and practice of the construction of autonomous software or hardware entities that act intelligently (rationally). Typical issues are how to deal with motivational attitudes such as intentions and with informational attitudes such as beliefs. Special attention is paid to 'social' attitudes within a 'multi-agent system'. In such an 'agent society' it becomes important to analyze multi-agent informational and motivational attitudes such as common knowledge and collective intention.

The second part is related to this, but the topics discussed are older. An important problem in AI concerns the question as to how to formalize *commonsense reasoning*, the way humans reason 'in daily life', so to speak, as opposed to reasoning in formal sciences such as mathematics and logic. For example, here one is interested in reasoning patterns connected with defaults (rules of thumb) and counterfactuals ('if ... had been the case, then ... would have been the case'). The study of these reasoning mechanisms appeared to be much more difficult than originally anticipated, and has become a major subject of study within AI since the beginning of the 80s. It includes so-called *non-monotonic reasoning* (reasoning in which earlier conclusions can get lost when more premises become available) and *belief revision* (dealing with how the beliefs of a reasoner change when new information becomes available and is incorporated).

Of course, since also in commonsense reasoning notions such as knowledge and belief play an important role, there is a natural relation with the field of intelligent agents, but the emphasis is different. However, as artificial agents become more intelligent and will invade daily life (such as e.g. the application in so-called companion robots that are supposed to assist and entertain elderly people), undoubtedly there will be a moment where these agents should also employ some (possibly restricted) form of commonsense reasoning![1]

2 INTELLIGENT AGENTS

Intelligent agents have become a major field of research in AI. Although there is little consensus about the precise definition of an intelligent agent, it is generally held that agents are *autonomous* pieces of hardware/software, able to take initiative on behalf of a user or, more generally, to satisfy some goal. Agents are often held to possess *mental attitudes*; they are supposed to deal with information, and act upon this, based on motivation. This calls for a description in terms of the agent's beliefs/knowledge, desires, goals, intentions, commitments, obligations, etc. To describe these mental or cognitive attitudes one may fruitfully employ *modal logic*. Typically for the description of agents one needs an amalgam of modal operators/logics to cater for several of the mental attitudes as mentioned above. Moreover, since agents by definition act and display behavior, it is important to include the *dynamics* of these mental attitudes in the description. One might even maintain that the logics of some of these attitudes, such as goal directedness and *a fortiori* desire, have little interest *per se*: they are rather weak logics without exciting properties. What makes them interesting is their dynamics: their change over time in connection with each other! So, although (modal) logics for e.g. knowledge, belief,

[1] The first part (section 2–8) were written by John-Jules Meyer. Frank Veltman is responsible for part 2 (sections 9–11).