TOWARDS A FORMAL THEORY OF INTENTIONS*

Munindar P. Singh
Artificial Intelligence Laboratory
MCC
Austin, TX 78759
USA
msingh@mcc.com
msingh@cs.utexas.edu

Nicholas M. Asher
Universität Stuttgart
Keplerstrasse 17
D-7000 Stuttgart-1
FRG
ims%rus.uni-stuttgart.dbp.de
@relay.cs.net

Abstract

Intentions are an important concept in several subfields of Artificial Intelligence. We present a formal theory of intentions and beliefs based on Discourse Representation Theory that captures many of the important logical aspects of the functional roles of intentions and beliefs, and of relationships among intentions, and between intentions and beliefs. Unlike possible worlds approaches, this theory does not assume that agents are perfect reasoners, and gives a realistic view of their internal architecture; unlike most representational approaches, it has an objective semantics, and does not rely on an ad hoc labeling of the internal states of agents. We then describe a minimal logic for intentions and beliefs. We close with several additional inferences, and the constraints on the model that correspond to them.

1 Introduction

An understanding of intentions is important to several subfields of Artificial Intelligence (AI), especially, speech act theory [3, 4, 10, 11], discourse processing [17], planning [16], and plan recognition [2, 23, 27]. We present a formal theory of intentions and beliefs that is based on Discourse Representation Theory (DRT) [5, 7, 19]. Our theory involves a formal model of time and possibility, and also explicitly models the structure of the agents' internal states.

Before we turn to the presentation of our theory, we must closely examine what one desires, or ought to desire, from a theory of intentions from the standpoint of AI. This, of course, depends on what one might want to do with such a theory. A theory of intentions is needed at the foundational level of study in AI and Cognitive Science in order to complete an account of intelligent (and possibly, rational) agency [12, 33]. Intentions

*We are indebted to Allen Emerson and Rob Koons for several discussions and for comments on previous versions of this paper.
are important attitudes of intelligent agents and, for resource-bounded agents, cannot be reduced to simple considerations of the optimality of decisions. Their importance to cognition and their use in AI has been defended extensively in the literature [4, 8, 10, 11, 16, 17, 29]. The roles of intentions mentioned below are especially important when one is interested in agents who would not otherwise (because of their limitations) be able to make appropriate or rational decisions. Intentions, and therefore a theory of intentions, are needed so that

- Designers (respectively, analyzers) may abstractly characterize the behavior they expect from the agents they are designing (respectively, analyzing).
- Agents (so designed) may interact intelligently with each other; i.e., to cooperate with others, if they are cooperative, or to compete successfully against them, if they are not.

A theory of intentions is also necessary in order to make sense of talk in AI about “plans.” Plans are mostly treated operationally in AI. A theory of intentions can provide a principled substitute for areas like Natural Language Understanding, where the plans and intentions of agents must be understood, in order to

- Fully understand their utterances.
- Communicate effectively with them (i.e., generate felicitous utterances, say, in replying to their queries).
- Understand descriptions of their actions (this is called “story understanding”).
- Provide assistance to them (this is important in the design of user interfaces, and in Computer Assisted Instruction).

These applications of the theory of intentions impose certain requirements on it. A useful theory would

- Provide an abstract account of the architecture of intelligent agents, especially with regard to their beliefs and intentions; this account would serve as the foundation for the semantic model incorporated in the theory.
- Validate some general inferences involving intentions.
- Provide for several definitions of intentions, each corresponding to a different species of agent, as might be encountered in different applications.
- Provide a connection to events.
- Provide a connection to the structure of discourses.
- Provide a connection to plans, as traditionally construed in AI.