© 1996 E. Sandewall
ECAI 96. *12th European Conference on Artificial Intelligence*
Edited by W. Wahlster
Published in 1996 by John Wiley & Sons, Ltd.

Getting Robots to Know Exactly What They Are Doing Extended abstract

Erik Sandewall[1]

In order to qualify as intelligent, robots have to be able to reason about their own actions and about events in their environment. In order to be reliable as well, they must be able to know what they are doing. In other words, they need to have *true, justified belief* about the results of their own actions. In order to design robots that know what they are doing, *we* must be able to verify the justifications of the beliefs of those robots.

I will propose that it is an appropriate long-term goal for A.I. research that we should be able to make a complete and coherent validation of a robot's cognitive system. Certainly we are not able to do that today, although quite a number of relevant techniques for solving some of the constituent problems have been developed. They are also the topic of the present lecture.

For the present purpose, I assume that belief is expressed using formulae in logic. What should then be meant by the *truth* of a robot's belief about its own immediate future, when it is able to take actions affecting that future? In order to make sense of this, I make the assumption that the robot maintains an *agenda* for its own actions in the near future. There is no point in assigning truth or falsity to the statements about its future actions, but truth or falsity can be assigned to statements about the future states of the world on the condition that the agenda is executed.

However, belief that just accidentally happens to be true does not qualify as knowledge; knowledge is *justified* true belief. In my view, such justifications must ultimately be based on accurate descriptions of the physical characteristics of the robot itself, its sensors and actuators, and the world surrounding the robot. Establishing a strict justification of the robot's beliefs about its world and its future requires a considerable number of steps, in two dimensions: several kinds of modelling are required for the various types of phenomena that are involved, and each phenomenon may need to be modelled on several levels, with successive transformations between those levels. For example, the robot's environment may contain hybrid continuous/discrete systems whose continuous aspects are modelled using partial differential equations; the use of those models may involve extended bond graphs and timed transitions systems before arriving to the level of formulae in logic.

Some of the relevant techniques have been developed in fields other than A.I., often with considerable sophistication, Therefore, the development of validation techniques for robotic knowledge is an interdisciplinary problem, which requires the cooperation of researchers from several backgrounds.

The lecture will describe and relate a number of problems which have to be addressed in this line of research. It will also describe a selection of approaches and solutions to those problems, with an emphasis on work in our department at Linköping University. Publications describing these contributions can be accessed via the following WWW page:

`http://www.ida.liu.se/~erisa/valrok`

This WWW page will also contain a reference to the forthcoming full text of the present paper.

[1] Department of Computer and Information Science,Linköping University, Sweden, E-mail `ejs@ida.liu.se`

© 1996 E. Sandewall

ECAI 96. *12th European Conference on Artificial Intelligence*
Edited by W. Wahlster
Published in 1996 by John Wiley & Sons, Ltd.