Motivated Behaviour for Goal Adoption

Michael Luck*  Mark d’Inverno†

* Department of Computer Science, University of Warwick, CV4 7AL, UK
mikeluck@dcs.warwick.ac.uk
† Cavendish School of Computer Science, Westminster University, London W1M 8JS, UK
dinverm@westminster.ac.uk

Abstract. Social behaviour arises as a result of individual agents cooperating with each other so as to exploit the resources available in a rich and dynamic multi-agent domain. If agents are to make use of others to help them in their tasks, such social behaviour is critical. Underlying this cooperation is the transfer or adoption of goals from one agent to another, a subtle and complex process that depends on the nature of the agents involved. In this paper we analyse this process by building upon a hierarchy previously constructed to define objects, agents, and autonomous agents. We describe the motivated self-generation of goals that defines agent autonomy and the adoption of goals between agents that enables social behaviour. Then we consider three classes of goal adoption by objects, agents, and autonomous agents. The first of these is merely a question of instantiation, the second requires an understanding of the relationship of the agent to others that are engaging it, and the third amounts to a question of negotiation or persuasion.

1 Introduction

The notion of autonomy has associated with it many variations of meaning. According to Steels, autonomous systems must be automatic systems and, in addition, they must have the capacity to form and adapt their behaviour while operating in the environment. Thus traditional AI systems and most robots are automatic but not autonomous — they are not independent of the control of their designers [29]. Autonomous systems are independent and exercise self-control. To do this, it is argued, they must be motivated.

Autonomous agents possess goals which are generated within rather than adopted from other agents. These goals are generated from motivations which are higher-level non-derivative components characterizing the nature of the agent. For example, consider the motivation greed. This is not a goal in the classical artificial intelligence sense since it does not specify a state of affairs to be achieved, nor is it describable in terms of the environment. However, it may, if other motivations permit, give rise to the generation of a goal to rob a bank. The distinction between the motivation of greed and the goal of robbing a bank is clear, with the former providing a reason to do the latter, and the latter specifying what must be done.

This view of autonomous agents is based on the generation and transfer of goals between entities. Specifically, an entity is an agent if it can be viewed as satisfying a goal. This goal must first be created and then, if necessary and appropriate, transferred to another entity. It is this adoption of goals that changes an entity from an object to an agent, and it is the self-generation of goals that is responsible for its autonomy.
Key to understanding the nature and behaviour both of individual agents, and of any interactions between them, is this notion of autonomy. In a series of papers, we have described and formally specified an extended theory of agent interaction, based on goals and motivations, which takes exactly this standpoint. The theory describes a framework for categorising different agents [14], which has been used as a basis for investigating aspects of the relationships between agents [16], providing an operational account of their invocation and destruction [6] and analysing their complexity [8], as well as for reformulating existing systems and theories [3,4,7]. In all this, however, one aspect has either been omitted or only briefly alluded to, namely a detailed account of the generation of goals from motivations, and goal adoption between agents. This paper addresses that omission, by showing how the formal framework may be used to provide a detailed operational account of the processes of goal generation and adoption.

First, we provide some context for the concept of motivation and its use in directing reasoning and behaviour with a short review of related work, and then consider the role of motivation in autonomous behaviour in more detail. Section 4 provides a brief outline of the formal agent framework, giving a selection of Z schemas that describe salient aspects, so that a reasonable context is available within which to situate this work. Then we analyse how motivations are used in goal generation and subsequently goal adoption. At each point we describe the processes involved both informally and formally using the Z notation. Finally, we summarise and present concluding remarks.

2 What and Why Motivation?

According to Halliday, the word motivation does not refer to a specific set of readily identified processes [9]. It is frequently discussed in terms of drive and incentive. Drives are related to physiological states such as the deprivation of food, hormones, etc, while incentives refer to external stimuli that affect motivation such as the presence of food, as an incentive to eat. Research on motivation is currently being pursued from a variety of perspectives including psychology and ethology. Our focus, however, is on providing an effective control mechanism for governing the behaviour and reasoning of autonomous agents through the use of motivations. Though we focus on a computational approach, in this section we will discuss related work.

Some psychological research has recognised the role of motivations in reasoning in a similar way to that suggested here. Kunda [12] informally defines motivation to be, “any wish, desire, or preference that concerns the outcome of a given reasoning task” and suggests that motivation affects reasoning in a variety of ways including the accessing, constructing and evaluating of beliefs and evidence, and decision making. Such arguments are supported by a large body of experimental research, but no attempt is made to address the issue of how motivations may be represented or applied in a computational context.

Computational work has also recognised the role of motivations. Simon [25] takes motivation to be “that which controls attention at any given time,” and explores the relation of motivation to information-processing behaviour, but from a cognitive perspective. Sloman [27,26] has elaborated on Simon’s work, showing how motivations are relevant to emotions and the development of a computational theory of mind.