Adjustably Autonomous Robotic Agents

Jason Garforth, Sue McHale, Anthony Meehan

Sheffield Hallam University, Pond Street, Sheffield S1 1WB, UK.
{a.meehan, s.l.mchale}@shu.ac.uk; jpg@janus.demon.co.uk

Extended Abstract

Building autonomous robotic agents that interact with the real-world is a complex and difficult task and several distinct paradigms are used to develop such machines. In the past decade there has been an exploration of approaches inspired by evolutionary and/or neurodynamical principles. [1, 2, 3, 4]. In earlier work, we have examined architectures which support agent-based design of such robotic systems [5]. But, whilst there have been successes in developing systems that exhibit low levels of performance e.g. wall following, safe exploration, map building, it is proving more difficult to develop systems that succeed in high level tasks which depend upon machines being capable of exhibiting (many) different behaviours.

One of the motivations for building autonomous robots in which the control system is designed using neurological principles is that such systems are fault tolerant, their performance degrading gradually. However, we have found that one problem with such systems is the initiation of a behaviour in an 'unintended' fashion. In the psychological literature, these events are called 'capture errors'; they may arise through incomplete training of neurally-based control system(s) or through 'damage' to the control system so that a learnt association is altered [6] (c.f. a stroke in humans).

In seeking to explore architectures that help manage capture errors we attempt to identify how humans might help the autonomous robot recover from the capture error (without, of course, resorting to repair). We have drawn upon the model of Norman and Shallice which was developed to explain how people control attention and avoid (frequent) capture errors. This model (Figure 1) proposes an architecture for understanding (neurologically-based) agent systems which are capable of 'high level behaviours which:

- require planning or decision making
- involve trouble shooting
- are ill-learned or contain novel sequences
- are dangerous or technically difficult
- require overcoming a strong habitual response [7].

At the heart of the model is *Contention Scheduling* which involves the routine selection of routine behaviours. In the Norman and Shallice model this is 'managed' through a *Trigger Database* and a *Supervisory Attentional System* (SAS). The structure of the model has suggested to us the means by which humans could interact with autonomous agents in order to adjust, momentarily or otherwise, the autonomy they possess; these means are via Perception-based Triggers, and the SAS:

1. *Perception-based Triggers* represent associations/mappings between a sensor mediated view of the world (including the current goal/state of the autonomous robot) and one or more behaviours. The autonomy of the robot may be modified either by the suppression, or by the simulated stimulation, of one or more perceptions.

2. The *Supervisory Attentional System* (SAS) plays a different role in that it is responsible for 'willed actions' that involve bringing under conscious control behaviours that would normally run unconsciously or by suppressing unwanted actions (such as capture errors) and facilitating wanted actions.

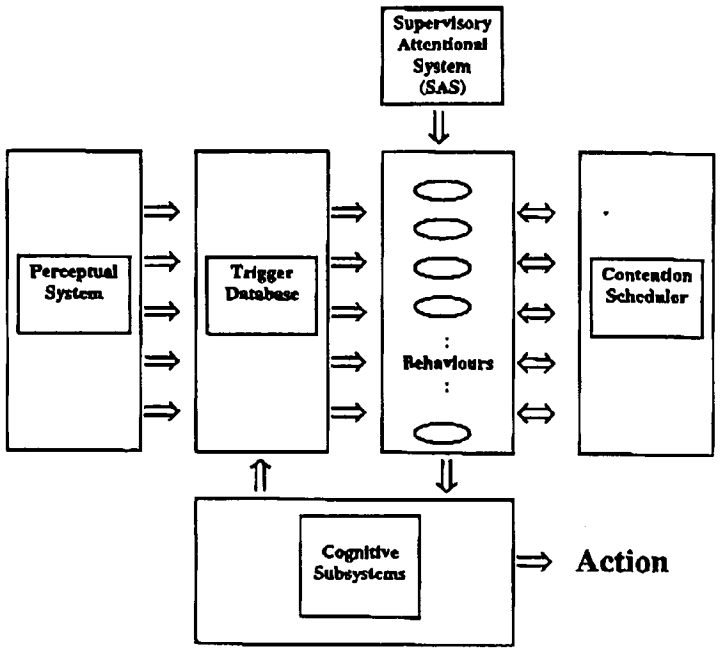

Figure 1. An adaptation of Norman & Shallice's model for willed and automatic control of behaviour which indicates where humans can interact with a (robotic) agent. The Trigger Database maps the perceived state of the world and of the system to behaviours. The Contention Scheduler deals with habitual behaviours that need no attentional control. The Supervisory Attentional System exerts "conscious" control to suppress unwanted behaviours such as capture errors or to 'force' the expression of a willed behaviour.

Implementing Adjustable Autonomy

There are many approaches to implementing this form of adjustable autonomy. To intervene at the level of agent perception one can, for some period of time, 'switch off' a set of sensors. Alternatively one can artificially stimulate a sensor. This form of human machine interaction has something of the character of the relation of a human to a prosthetic limb. To intervene at the level of the SAS one momentarily overrides the currently expressed behaviour to induce another This is not to extinguish the autonomy of the robotic agent, rather, the relationship has more of the character of a human who works with a working dog such as a sheep dog. These and other approaches to human interaction with autonomous robots continue to be the subject of experimental work at the AI Laboratory of Sheffield Hallam University.

References

1. Beer, R., Chiel, H. and Sterling, L. A biological perspective on autonomous agent design. Robotics and Autonomous Systems Vol 6, 1990, 169-186
2. Cliff, D. Computational neuroethology: a provisional manifesto. CSRP 162, University of Sussex School of Computing and Cognitive Sciences, 1990
3. Cliff, D. AI and Artificial Life: Special Theme Editor. AISBQ No. 87, Spring 1994. SSAISB Brighton, UK
4. Dyer, M.G. Towards Synthesising Artificial Neural Networks that Exhibit Co-operative Intelligent Behaviour: Some Open Issues in Artificial Life. Artificial Life Vol 1, 1990, 111-134.
5. Garforth, J., Brammer, K., Meehan, A.S., A Flexible and Extendible architecture for Autonomous Robots. Proc. 5th International Workshop on Advanced Robotics and Intelligent Machines. 25-27 March 1997 Salford, UK
6. Shallice, T. From neuropsychology to mental structure. Cambridge University Press, 1988, Ch. 14.
7. Styles, E.A. The Psychology of Attention, Psychology Press 1997, Ch 9.
8. Hinton, G.E, Shallice, T. Lesioning an attractor network: Investigations of acquired dyslexia, Psychological Review, Vol 98, No 1, 1991, 74-95