Problems of Attentional Behaviour in Autonomous Robotic Systems

Jason Garforth, Sue McHale, Anthony Meehan

Schools of Engineering, Psychology and Computing,
Sheffield Hallam University, Pond Street, Sheffield S1 1WB, UK.
jpg@janus.demon.co.uk; {a.s..meehan | s.mchale}@shu.ac.uk

Abstract

Behaviour based robots have problems related to inappropriate behaviours expressed by the machines. We identify two classes of problem, *capture errors* and *perseverative behaviour* which can cause a machine to fail to meet its real-time goals. We suggest that these errors may be moderated or eliminated by the use of an executive or Supervisory Attentional System, based on the neuropsychological model of attention developed by Norman & Shallice (1988)

Introduction

Autonomy is a problematic concept because it is a relative term. A wholly autonomous system is one which exists *for itself* and not just *in itself.* Robustness of such systems implies attributes such as self organisation, self repair, and even self reproduction (Courant, Hirsbrunner and Stoffel 1994). In practice, we look for dependable and reliable systems which will perform functions determined wholly by ourselves. These systems are more properly considered allonomous. Between these spectral poles exist the vast majority of 'autonomous' systems developed to date. These systems may be represented as allonomous in so far as we specify or select for a desired behaviour and autonomous in so far as the precise behaviour they exhibit is not specified for all states of a dynamic environment and so behaviours appear to be determined by the system itself.

Creating robotic systems which exhibit high degrees of autonomous behaviour is a complex and difficult task (Langton 1992; Meyer and Guillot 1991; Maes 1994). Whilst there has been some success in creating autonomous machines that exhibit low level behaviours (e.g. wall following, surface exploration, map building), machines that exhibit higher level behaviors have proven more difficult to develop (Lopes and Teixeira 2000). This short paper offers a framework within which to classify some of the 'behavioural' problems exhibited by such machines and proposes an approach which may resolve them.

We focus on robotic control systems which employ neuropsychological principles in their design. Despite their limited success to date, research based upon these principles continues, not least because such systems can exhibit self organisation and adaptation, are naturally fault tolerant - performance degrading gradually when damaged, neural controllers are inherently suited to the noisy sensory environments characteristic of the 'real world'. However, the 'problem' behaviours we consider are also exhibited by machines which are controlled by traditional 'algorithmic' controllers and so the problem is quite general.

Behavioural Problems in Robots

In exploring neurally-based control systems for robot systems developed at the AI lab at Sheffield Hallam University we, in common with others, have found two related behavioural problems: in the fist case, the robot persists in a behaviour - often a repetitive or cyclical behaviour. In the second, the robot seems to drift away from one goal towards another. These behavioural manifestations undermine the perceived autonomy of the machine. We should not be surprised that systems exhibit such behaviour (and especially not neural systems) as both these behaviours are exhibited by humans. For example, most of us can recall occasions when we become conscious that we are doing one task but intended to do another ('why am I driving to work when I want to go to the shops?').

In an attempt to understand and overcome these problems, we have sought accounts of similar behaviour in humans. Neuropsychologists describe two behaviour related pathologies of the prefrontal cortex which seem to correspond the behaviours we have identified: *perseverative* errors and *capture* errors (Roberts, Robbins and Weiskrantz 1998).

- *perseverative* behaviours occur when the system fails to revise its behaviour in a timely fashion.

- *capture errors* involve an inappropriate or undesirable switch to a different behaviour or goal.

These behaviours are associated with damage or abnormality in an area of the prefrontal cortex that seems to initiate plans and modulates higher level behaviour: functionally, it is referred to as the *executive*. The concept of an *executive* as a part of the brain is currently the subject of debate amongst both neurophysiologists and neuropsychologists and several models exist notably those of Struss and Benson (see Parkin 1996), Baddeley (Baddeley and Weiskrantz 1993) and Norman and Shallice (Shallice 1988). In the Struss and Benson model the executive is composed of processes which affect two functional systems: drive and sequencing which in turn control and influence other systems within the brain. Both the Baddeley and Norman and Shallice models contain an executive called the *Supervisory System* or *Supervisory Attentional System (SAS)*, respectively.

We suggest that system control architectures which explicitly address attention are necessary to build robust systems that do not exhibit these problems (at least not to a to a problematic degree). We have chosen to explore this proposal by experimenting with a control architecture based upon the Norman and Shallice model of willed behaviour (Figure 1) and as further elaborated in (Shallice and Burgess 1998).

This model is composed of several functional components:

- *Perception Based Triggers*: a (database) mapping between sensory input patterns, the internal state of the agent including any goals that the agent currently has (cognitive subsystems) and behaviours (schemas).

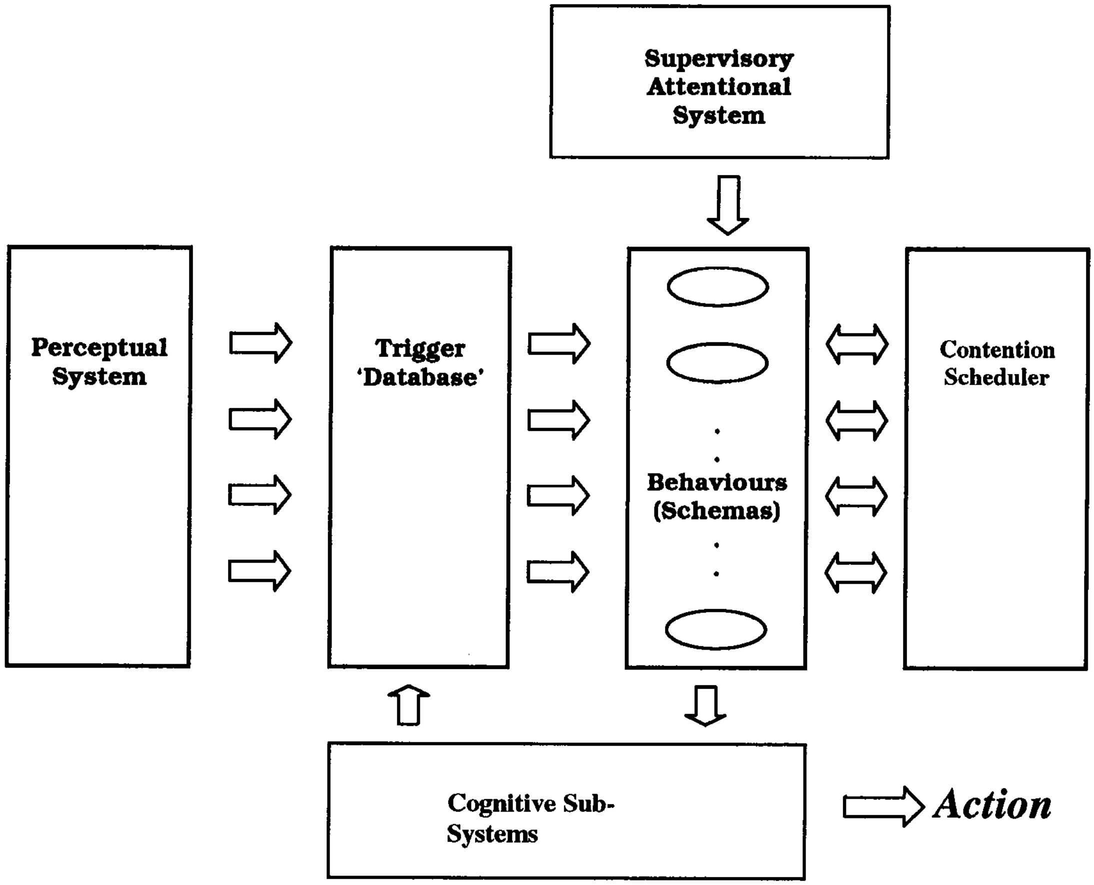

Figure 1. An adaptation of Norman & Shallice's model for willed and automatic control of behaviour which indicates where humans can interact with a (robotic) agent. The Trigger Database maps the perceived state of the world and of the system to a set of candidate behaviours. The Contention Scheduler deals with habitual (unconcious) 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, possibly novel, behaviour (see Styles 1997).

- A *Contention Scheduling* mechanism which is invoked when selected behaviours are incompatible: different behaviours are compared and adjusted dependant on the situation the agent finds itself in. (Comment: It seems to us that the model of the basal ganglia offered by (Prescott, Redgrave and Gurney 1999) relates directly to the function of the contention scheduler.)

- If an agent encounters danger, requires a novel goal or recognises the selection of an inappropriate action, the agent must actively suppress one behaviour and exhibit another. This willed behaviour is achieved by the intervention of the *Supervisory Attentional System* (SAS).

The role of the SAS is distinct from a lower level action selection mechanism such as may exist in the basal ganglia. It provides an additional 'input' to the system which expresses a dynamic, 'willed' component which then influences action selection.

In humans the Norman & Shallice Supervisory Attentional System operates in five *situations*:

- Situations that involve planning or decision making.

- Situations that involve correcting or 'trouble shooting' behaviours.

- Situations where the responses are not well learned or contain novel sequences of actions

- Situations judged to be dangerous to the agent or technically difficult.

- Situations that require the overcoming of a strong habitual response or the resisting of temptation.

Though the SAS is now thought of as being a collection of at least three subsystems it is still classed as being a single functional system within the brain.

Direct evidence for the SAS is severely limited, but within neurophysiology the evidence for the existence of an executive such as the SAS or at least something that can modulate attention, is mounting. Studies of several specific brain diseases, notably Alzheimer's, Parkinson's and schizophrenia (Cohen, Braver and O'Reilly 1998), support the idea of an executive and/or supervisory attentional control.

Capture errors can be explained using Norman and Shallice's model as a temporary failure of the SAS to recognise a change in unconscious goal oriented behaviour. *Perseverative behaviour* is explained as a failure of the SAS to recognise the timely need for a change of goal or sub-goal.

Implementation

Using the Norman and Shallice model as a neural architecture requires a neural realisation of the functional components (although it is not necessary realise all components in this way to evaluate our proposals about the SAS).

The perception-based triggers for behaviour can be realised as an associative memory between sensors and a set of basis behaviours (Mataric 1996).

A 'winner-takes-all' model of the basal-ganglia (Prescott, Redgrave and Gurney 1999) forms the contention scheduler with the output for the effectors taken directly from it.

We have yet to determine the most appropriate neural architecture for the SAS. Currently, we propose to implement some of the hypothesised sub-systems, particularly those components associated with intention markers (Shallice and Burgess 1998).

An Experimental Context

To explore this and other proposals we have chosen a multi robot system for the behaviours relating sheep, sheepdogs and humans in a working environment. The problem of herding is one in which a robot is naturally faced with a choice between contradictory goals and behaviours. Sheepdogs exhibit herding behaviour and, in interacting with people, they obey just four basic working commands (Jones and Collins, 1987). The human input is mainly designed to allow us to control one channel of influence over the robots attention.

The robotic sheepdogs are receptive to sound, so that the human handler can control the machine in much the same way as they would a real sheepdog. The main goal is to try to influence the robotic sheepdogs attention so as to produce behavioural changes. Once this is a repeatable experiment, the SAS will either be suppressed or damaged and the experiment repeated to ascertain whether or not the attention of the robotic sheepdog is altered in a predictable way, consistent with reappearance of perseverative behaviours and capture errors, thereby validating the model. It should be possible to distinguish the functioning of the SAS from that of the contention scheduler (basal ganglia). Damage to the basal ganglia in real animals can lead to 'behavioural traps' where appropriate but alternating, ill-coordinated behaviours are expressed. In contrast, a damaged SAS in humans can lead to inappropriate, but coordinated behaviours and very frequent capture errors.

Progress

To support the research we have built a hardware and software laboratory, that includes:

- An experimental robot 'SheepOne' was built using several 8 bit micro-controllers and programmed using a custom written 'C' compiler. It exhibited basic behaviours. Whilst it successfully managed to wander around a room it exhibited severe perseverative errors for a given ultrasonic sensor configuration. The behavioural errors shown by the machine were identified as due to a lack of neural memory of recent history. This was overcome in simulation using Elman neural networks instead of feed forward multi layer perceptrons.

- The 'SheepTwo' robots: three wheeled robots, each having four processors. The design includes a servo based sensor head with low power CMOS cameras, two to three ultrasonic senor pairs and close-contact switches. Using ultrasonic range finders and contact switches the sheep can move fairly quickly around an environment. Their controllers run on version of Linux for embedded micro-controllers (uCLinux on an RT-Control uCsimm system on a stick). The sheep form Sheffield Hallam University's *flock* being built for research into emergent group behaviours and as a focus of attention for the sheepdogs. Two prototypes are currently operational. They exhibit a range of basis behaviours such as avoid, roam, approach, follow, aggregate (Mataric 1996).

- Sheepdogs: six wheeled robots substantially larger and more powerful. These machines use a flexible communications architecture to connect processors together. The superstructure is constructed from a 3U 19" sub-rack frame mounted on top of the chassis containing the drive mechanism and batteries. Capable of supporting several on-board processing units and large amounts of I/O these machines will have a simple vision system as well as more conventional sensors. As in the Sheep the Sheepdogs machines use embedded Linux (uCLinux) as the operating system, enabling rapid application development and deployment. Two choices for the CPU's are being evaluated: an RT-Control provided uCdimm (available Q1 2001) and an AXIS Etrax development board.

- An evolvable neural architecture used to build the controllers.

- SheepWorld: a Java-based simulation for the sheep and sheepdogs built to validate the neural controllers prior to porting to the physical machines.

Summary

This research project is focused on one of the problems facing the builders of behaviour based autonomous robotic agents namely the occurrence of what we classify as *Perseverative behaviours* invoked through an habitual response which at some point in time become inappropriate and *Capture errors* where an inappropriate behaviour is not actively suppressed by the agent.

We suggested that the robustness of autonomous robots can be improved by implementing a neuropsychological model of attentional control (*Supervisory Attentional System*).

References

Beer, R., Chiel, H. and Sterling, L. 1990 A biological perspective on autonomous agent design. *Robotics and Autonomous Systems* Vol 6, 169-186

Baddeley, A., Weiskrantz, L. 1993. *Attention: Selection, Awareness, & Control.* Oxford University Press. 152-170.

Cohen, J.D., Braver, T.S. and O'Reilly, R.C. 1998. A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges in Roberts, A. C., Robbins T. W., Weiskrantz, L. (eds.) *The Prefrontal Cortex.* Oxford University Press. 195-220.

Courant, M., Hirsbrunner, B. and Stoffel, K. 1994. Managing entities for Autonomous Behaviour in Thalmann, N. and Thalmann, D. (eds) *Artificial Life and Virtual Reality.* Wiley. 83-95

Jones G. H. and Collins B. C. 1987. *A Way of Life: Sheepdog Training Handling and Trialling.* Farming Press.

Langton, C.G. 1992 Introduction. In Langton, C.G. et al (eds.) *Artificial Life II: proceedings of the Workshop on Artificial Life (February, 1990)* Santa Fe, New Mexico, Addison-Wesley.

Lopes, L.S. and Teixeira, A. 2000. Teaching behavioural and task knowledge to robots through spoken dialogue. *AAAI Spring Symposium 2000,* Stanford University, Ca. USA.

Maes P. 1994. Modeling Adaptive Autonomous Agents. *Artificial Life* 1: 135-162.

Mataric, M.J. 1996. Designing and Understanding Adaptive Group Behaviour. *Adaptive Behaviour* 4:1 51-80.

Meyer, J. A. and Guillot, A. (1991) Simulation of adaptive behaviour in animats: review and prospects. In Meyer, J. A. & Wilson, S.W. (eds.) *From animals to animats, Proceedings of the First International Conference on the Simulation of Adaptive Behaviour.* Cambridge, MA: MIT Press/Bradford Books.

Parkin A. J. 1996. *Explorations in Cognitive Neuropsychology*. Psychology Press.

Prescott, T.J., Redgrave, P and Gurney, K. 1999. Layered control architectures in robots and vertebrates. *Adaptive Behaviour* 7(1) 99-127. International Society for Adaptive Behaviour.

Shallice, T. *From Neuropsychology to Mental Structure.* Cambridge University Press, 1988, Ch. 14.

Shallice, T. and Burgess, P. (1998) 'The domain of supervisory processes and temporal organisation of behaviour' in Roberts, A. C., Robbins T. W., Weiskrantz, L. (eds.) *The Prefrontal Cortex*. Oxford University Press. 22-35.

Styles, E.A. (1997) *The Psychology of Attention.* Psychology Press 1997, Ch 9.