The integration and control of behaviour: Insights from neuroscience and AI

David W. Glasspool
Advanced Computation Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, London, and Institute of Cognitive Neuroscience, University College London dg@acl.icnet.uk

Abstract

Clues to the way behaviour is integrated and controlled in the human mind have emerged from cognitive psychology and neuroscience. The picture which is emerging mirrors solutions (driven primarily by engineering concerns) to similar problems in the rather different domains of mobile robotics and intelligent agents in AI. I review both approaches and argue that the layered architectures which appear in each are formally similar. The higher layer of the psychological theory remains obscure, but it is possible to map its functions to an AI theory of executive control. This allows an outline model of Norman and Shallice's Supervisory Attentional System to be developed.

1 Introduction

Building a functional mind is an ambitious goal. How can the cognitive disciplines - artificial intelligence and cognitive psychology - contribute to such an undertaking? Both psychology and AI are well known for studying small areas of cognition and working with theories of single empirical phenomena. In a full scale cognitive theory two related issues must be addressed, those of *integration* (how are numerous cognitive modules organised into a coherent whole, rather than descending into behavioural chaos?) and *control* (how are the modules to be co-ordinated by an explicit goal?). In this paper I consider a set of theories from AI, neuropsychology and mobile robotics which are concerned with the integration and supervisory control of behaviour. These theories provide converging support for a form of cognitive architecture comprising layered control systems, the lower levels of which contain multiple simple, independent behavioural processes while higher levels are characterised by slower deliberative processes which exercise supervisory control.

A natural question is whether a convergence of this sort can benefit the individual disciplines involved by providing insights from other fields. There are potential benefits for both AI and psychology in this case. In the final part of the paper I describe an example of the way insights from AI, which has tended to concentrate on "higher level" cognitive processes, may benefit psychological theory, which tends not to be so well developed in these areas. Thus an AI theory of agent control can provide a model for higher level supervisory processes in a neuropsychological theory of behaviour control.

2 The organisation of action: A neuropsychological approach.

While a number of theories in psychology have addressed the organisation and control of behaviour, that of Norman and Shallice (1980; 1986) is perhaps the most dominant. The theory is informed both by the slips and lapses made by normal individuals in their everyday behaviour, and by the varieties of breakdown in the control of action exhibited following neurological injury.

2.1 Action lapses and slips

Reason (1984) has studied the slips and lapses made by normal individuals during routine behaviour. Errors in everyday behaviour turn out to be surprisingly common, but can be classified as belonging to a limited set of types. These include errors of place substitution (e.g. putting the kettle, rather than the milk, into the fridge after making coffee), errors of object substitution (e.g. opening a jar of jam, not the coffee jar, when intending to make coffee), errors of omission (e.g. pouring water into a tea pot without boiling it), and errors involving the "capture" of behaviour by a different routine (such as going upstairs to get changed but getting into bed). Interestingly Reason finds that the situations in which such slips and lapses occur share two properties in common: The action being performed is well-learned and routine, and attention is distracted, either by preoccupation or by some external event.

There are two points of interest here. Firstly it is clear that we can perform a wide range of often complex habitual actions without concentrating on them - the control of well-learned action can become automatic. Secondly, when we allow such behaviour to proceed without our conscious control it is susceptible to a specific range of characteristic errors. These observations provide one class of data which psychological theories of action control must address. Another important class of data is provided by the effects of neurological damage.

2.2 Neurological impairment of behaviour control

The breakdown of cognitive systems following neurological damage constitutes an important source of constraint on psychological theory. Cooper (2000) reviews a range of problems with the control of action which mainly follow damage to areas of prefrontal cortex. Here I briefly mention three syndromes of particular interest.

Patients with action disorganisation syndrome (ADS, Schwartz et al. 1991, Humphreys & Forde, 1998) make errors which are similar in type to those of normal individuals - errors in the sequencing of actions, the omission or insertion of actions, or the substitution of place or object. However their errors are far more frequent. For example patient HH of Schwartz et al. (1991) made 97 errors during 28 test sessions in which he made a cup of coffee.

Utilisation behaviour (Lhermitte, 1983) can be characterised as weakening of intentional control of action, so that irrelevant responses suggested by the environment may take control of behaviour. A neurological patient exhibiting utilisation behaviour may pick up and perform actions with items lying around on a table, for example, which are appropriate to the items but not relevant to the task in hand.

Shallice and Burgess (1991) report patients with "strategy application disorder" who are able to carry out individual tasks but have difficulty co-ordinating a number of simultaneous task demands. Such patients for example may be able to carry out individual food preparation tasks but are unable to plan and cook a meal. Their deficit appears to be in the ability to schedule multiple tasks over an extended period.

2.3 The Norman and Shallice framework for behaviour control

The challenge for a psychological account of the integration and control of behaviour is to explain data of the type outlined above. Norman and Shallice (1980; 1986) interpret the data as implying that two distinct systems operate to control the range of behaviour typically studied by psychologists. The systems are arranged in a layered manner as shown in Figure 1 (a). Over-learned or habitual action is held to be controlled by a set of *schemas* competing within a contention scheduling (CS) system for control of the motor system, while willed or attentional control of action is achieved by a supervisory attentional system (SAS) which can influence the CS system but has no direct access to motor control.

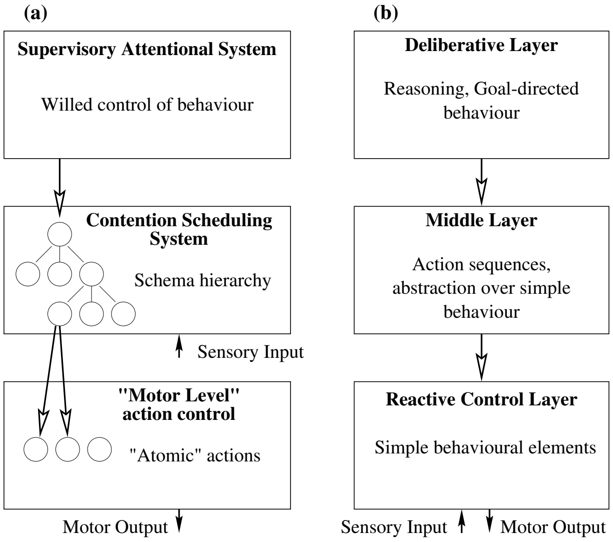

Figure 1: (a) Norman and Shallice's (1986) framework for action control augmented with Cooper and Shallice's (in press) distinction between cognitive and motor level action. (b) The three-layer architecture of Gat (1998) and colleagues.

Cooper and Shallice (in press) provide a number of arguments for distinguishing, on grounds of psychological data, two sub-levels of low-level behaviour. The lower sub-level, "motor" behaviour, comprises the individual motor commands required to carry out a simple action (extending and retracting individual muscle groups to grasp an item, for example). The higher sub-level, the "cognitive" level, operates with actions at the lowest level to which they are referred in everyday language - grasping, reaching etc. Norman and Shallice's CS component applies to cognitive level actions, which abstract over motor level actions. The theory does not directly address operations at the motor level.

The contention scheduling system comprises a hierarchy of schemas, defined as discrete actions or structures organising sets of actions or lower-level schemas. The schema hierarchy terminates in a set of "cognitive level" actions which are held to be carried out directly by motor systems. Actions at this level might include, for example, "pick up an item", "unscrew", or "stir". Higher level schemas might include "open jar", which would organise the actions of picking up, unscrewing a lid, and putting down. At a higher level still a "make coffee" schema might exist.

Schemas are connected in an interactive-activation network. They are activated from the top down by their parent schemas or by control from the SAS, and from the bottom up by input from the environment. They compete for execution on the basis of their activation level. A schema is triggered when its activation level is higher than any other schema and higher than a trigger threshold. A triggered schema feeds activation forward to its child schemas, and is inhibited after its goal has been achieved. Top-down activation can exert detailed control over behaviour or it can simply be used to specify goals, by activating high-level schemas. Such schemas may provide multiple ways for a goal to be achieved - coffee can be supplied in a jar or a packet, for example, so a schema for adding coffee to a mug can be indifferent to the particular lower level behaviour required to achieve its goal. Whichever suitable sub-schema best fits the current configuration of the environment will be selected.

Cooper and Shallice (in press) have simulated the CS system in detail. With a certain amount of background noise in the system, and a reduction in top-down input, the system makes occasional errors analogous to those made by normal individuals, when the wrong schema or sub-schema is triggered. By varying the parameters of the model - in particular the levels of top-down influence and environmental influence, utilisation behaviour and ADS can be simulated, as well as a number of other neuropsychological disorders of action control.

Just as important to the Norman and Shallice account of behaviour control is the SAS, which is held to take control of behaviour in non-routine situations (ie those where no appropriate well-learned schema exists) and in situations where performance is critical. The SAS exerts control by directly activating individual low-level actions, or by causing the selection of an existing schema which would not otherwise be selected in that situation. Internally, however, the SAS is poorly specified. Based largely on neuropsychological evidence but partially guided by *a priori* reasoning about the types of processes which must be involved in supervisory processing, Shallice and Burgess (1996) set out an outline of the processes involved in the SAS and their relationships during supervisory processing. They characterise the functioning of the SAS as centrally involving the construction and implementation of a temporary new schema, which can control lower level CS schemas so as to provide a procedure for dealing effectively with a novel situation.

Shallice and Burgess' characterisation of the SAS as modular, and their preliminary functional decomposition, provide a useful starting point for neuropsychological theory. However the picture remains unclear, with many processes under-specified. This is largely due to the difficulty of obtaining clear empirical data on such high-level processes. We return to the specification of the SAS later. For now however we can note that it is concerned with problem solving and planning, and delegates the control of routine behaviour to the CS system as long as things are running smoothly.

The Norman and Shallice theory provides a framework for the control of willed and automatic behaviour based on psychological and neuropsychological evidence. I now turn to an equivalent problem in artificial intelligence - the control of behaviour in autonomous robots.

3 The organisation of action in mobile robotics

Mobile robotics has long been seen as an important area for artificial intelligence research. It is an area where all aspects of an agent's behaviour and its interaction with its internal and external environment must be taken into account. Theories are forced to address, to some extent at least, the entire cognitive system from sensory input to motor output, and the interaction of the agent with its environment.

Early AI robotics projects (e.g. "Shakey", Nilsson 1984; the CART, Moravec, 1982) employed architectures centering on classical planning systems. Such systems typically involve three sequential steps in their control architectures: sensing, planning and acting. In the first step sensory information (e.g. from a video camera) is analysed and used to form a map of the robot's environment. In the second step a search-based planning system is applied to the map to find the most appropriate plan of actions to be followed in order to achieve a goal. Once a plan has been generated the robot can make a move. Such systems are often known as sense-plan-act (SPA) architectures.

There are a number of well-known problems with this approach. It requires search over a large state-space, leading to slow, resource-hungry operation. The plan which is generated is critically dependent on the reliability of the sensors and on the environment remaining static while the plan is formulated. Even with improvements in computing hardware and planning techniques robots based on this paradigm tend to remain slow, cumbersome and fragile in their operation.

In the mid 1980s Brooks developed an alternative approach to robot control in response to these problems, sometimes termed reactive control (or "reactive planning", Brooks 1991). This represents a break from the sense-plan-act cycle. Brook's paradigm largely does away with a central representation of the world and uses many simple, high-speed (reactive) processes coupling simple sensory systems directly to action, operating in a highly parallel manner. These reactive processes implement small, circumscribed elements of behaviour, and are usually referred to simply as "behaviours". The direct coupling of input to output and decomposition of behaviour into many simple, environmentally-driven "behaviours" allows small, fast, robust and flexible robot control systems to be built.

Rapid theoretical development followed Brook's initial work. It soon became apparent that, in its pure form, Brooks' reactive behaviour paradigm becomes difficult to program as more complex behaviour patterns are attempted. In practical applications the lack of any ability to carry out high-level planning and problem solving was also a concern. Gat and colleagues (Gat, 1998) have been in the vanguard of a second wave of development aimed at formalising reactive agent control systems to make them more robust and scalable. Much of this work centres on the idea that three distinct layers of control are required for a large-scale practical agent: a rapid but simple reactive low-level control system, an intermediate system capable of stringing together sequences of simple actions into useful behavioural elements, and a slow "deliberative" high level system capable of carrying out more complex planning and reasoning. Such schemes have been termed three-layer architectures (TLAs, Gat 1998) (Figure 1, b).

The lowest level in a TLA provides the responsive, flexible and robust low-level control of behaviour characteristic of Brooks' reactive approach. The top level provides a more traditional AI planning and problem-solving capability, allowing the robot's behaviour to be guided by long term, abstract goals. The middle layer interfaces between the two. It provides abstractions over lower level behaviours in two ways - by constructing more powerful behavioural elements through assembling sequences of simple behaviours, and by providing higher level goals which may be achieved by different lower level actions depending on prevailing circumstances. The top level system can interact with the robot through relatively abstract commands and need not specify every detail of the actions needed to implement its goals.

4 Converging architectures?

The Norman and Shallice framework and the TLA paradigm address similar issues of control and integration of an agent's behaviour in two rather different domains. While the original Norman and Shallice theory speaks to only two layers of control - CS and SAS - the inclusion of Cooper and Shallice's "motor" action level yields a three-layer framework. The correspondence with the TLA is striking (Figure 1). Might the resemblance simply be superficial, though? We need to compare the way the layers are specified in each approach.

Shallice and Burgess describe the SAS as corresponding to frontal-lobe processes "critically involved in coping with novel situations as opposed to routine ones" (1996, p.1406). They specify its functions in terms of goal-setting, problem solving and schema generation (planning). Gat (1998) describes the topmost TLA system as "the locus of time-consuming computations. Usually this means such things as planning and other exponential search-based algorithms [...] It can produce plans for the [middle layer] to implement, or it can respond to specific queries from the [middle layer]". In other words the main functions are generating new plans of action and dealing with situations for which no pre-existing procedure exists in lower levels, i.e. novel situations. Despite the language differences - an inevitable consequence of comparison across disciplines - the two architectures apparently ascribe essentially the same functions to their highest level systems.

Turning to the lowest level of behaviour control, on Cooper and Shallice's (in press) account this corresponds to "motor level" actions. These operations are the preserve of motor systems and are not susceptible to the types of errors typically made at the "cognitive" level. On the Norman & Shallice / Cooper & Shallice framework the distinction between the lowest (motor) level and middle (CS) level is well defined. It is not clear that the corresponding distinction in the TLA approach is well defined, however. Gat (1998) describes the processes at the lowest TLA level as "designed to produce simple primitive behaviours that can be composed to produce more complex task-achieving behaviour". The composition of simple behaviours into complex behaviour is a function of the middle layer. It is not entirely clear at what point a simple behaviour becomes a complex one (although Gat does give a number of guidelines for the type of behaviour to be considered simple, including keeping internal state to a minimum and using only input-output transfer functions which are continuous with respect to internal state). If the idea were simply that actions which are, from the point of view of higher level systems, atomic should be included this level would correspond well with Cooper and Shallice's motor level. However the notion of reactive control - tight sensory-to-motor coupling - is an important part of the TLA definition of this layer. The triggering of action by environmental input is not prominent in Cooper and Shallice's characterisation (although reflex and sensory-motor feedback certainly play an important part in low-level human motor control). This type of control is however certainly part of the definition of CS. Cooper and Glasspool (in submission), for example, treat the environmental triggering conditions of schemas in CS as "affordances" for action, priming appropriate behaviour in response to learned environmental configurations. It is thus possible that the lowest level layer in the TLA account corresponds to a combination of the motor layer and the lowest level action representations in CS. Higher order schemas in CS would then correspond to the middle TLA layer.

In the TLA account, a primary function of the middle layer is to organise primitive behaviours into behaviour sequences which perform two functions: they form a more compact and convenient representation of behaviour for use by higher level processes (i.e. sequences of behaviour which are often needed are "chunked" together), and they provide abstraction - alternative means may be specified for achieving a goal, providing low-level flexibility and avoiding the need to specify behaviour in detail. Both of these functions are central to the Norman and Shallice CS system. Schemas represent well-learned fragments of behaviour and provide a goal-based representation - sub-schemas for achieving the same goal compete to service a higher-order schema's requirements. Functionally, the CS corresponds well to the TLA middle layer.

In this connection it is important to note an early attempt to overcome some of the problems of "pure" reactive robotic control by Maes (1989). Maes' scheme has a range of alternative behaviours (specified at a level typical of the TLA "middle layer") competing for control of resources (robot effectors) in an interactive activation network under the influence of environmental input. The similarities with Contention Scheduling are striking, especially given the very different provenance of the theories. The approach has not been followed up, apparently because of a view that in real-world cases robot control systems can be made simple enough that flexible, on-line resource allocation and conflict resolution are not necessary. That this appears to be a primary function of intermediate-level behaviour control in humans suggests that this view may be challenged as robotic systems are scaled up to more complex tasks.

It thus appears that the similarity between TLAs and the SAS/CS framework is more than superficial and may represent a true convergence of theory in two distinct areas. Whether this is the case would be clearer with a more detailed specification of the Norman and Shallice framework. The CS component is well specified and has been modelled in detail by Cooper and Shallice (in press). The motor level and the SAS are less clearly specified. The SAS in particular is only characterised in outline by Shallice and Burgess (1996). However, an implementation of the SAS, even if only in outline, would provide a valuable first step in fully formalising the theory as well as enabling a number of issues concerning the interface between SAS and CS to be addressed. In the remainder of this paper I therefore describe a first step towards a computational model of the SAS.

5 Modelling the SAS

The shadowy nature of the SAS is testament to the difficulty of "reverse engineering" processes of such scope and complexity in human psychology. However, while the SAS is a construct posed at an unusually high level for psychological theory, it does address processes at the same general level as many theories in AI. This may allow psychological theory to benefit from the alternative perspective of AI, with its greater emphasis on engineering intelligent systems from first principles. Shallice and Burgess (1996) identify three stages in the operation of the SAS in its typical role of reacting to an unanticipated situation:

1. The construction of a temporary new schema. This is held to involve a problem orientation phase during which goals are set, followed by the generation of a candidate schema for achieving these goals.

2. The implementation of the temporary schema. This requires sequential activation of existing schemas in CS corresponding to its component actions.

3. The monitoring of schema execution. Since the situation and the temporary schema are both novel processing must be monitored to ensure that the schema is effective.

The domino model of Fox and colleagues (Das, Fox, Elsdon & Hammond, 1997; see also Fox and Cooper, this symposium) provides a framework for processes of goal-setting, problem solving and plan execution which gives a promising initial fit to Shallice and Burgess's outline. It specifies seven types of process operating on six types of information. The domino framework is shown in Figure 2 (broken lines). Starting from a database of beliefs about its environment the agent raises goals in response to events requiring action. Such goals lead to problem solving in order to find candidate solutions. Alternative solutions are assessed and one is adopted, leading to new beliefs and possibly to the implementation of a plan of action, which is decomposed into individual actions in the world. The processes are similar to those specified by Shallice and Burgess: goal setting, solution generation and evaluation, decision making, planning, acting and monitoring the effects of action. A set of well understood and well specified formal semantics can be associated with the framework to render it computationally implementable. The domino thus provides an appropriate starting point for an SAS model. Figure 2 shows that the processes identified by Shallice and Burgess (1996) can be mapped cleanly onto the domino framework. The "candidate solution generation" process of the domino framework corresponds to the generation of a "strategy" in SAS - a generalised plan of action which is subsequently implemented as a concrete schema for execution by the CS system.

5.1 Architecture and operation

For the purposes of modelling a target task is required. A standard test of frontal lobe (and *ex hypothesi* of SAS) function in neuropsychology is the Wisconsin card-sorting test (WCST). The subject is given a set of cards which vary in the number, shape and colour of the symbols they show (thus a card might show two green squares, or four red triangles). The experimenter lays out four "stimulus" cards, and the subject is asked to sort the cards into piles corresponding to these, but they are not told the criterion for sorting. They might sort cards by the number of symbols, their colour or their shape. After each card is placed the experimenter indicates whether it was correctly sorted. Once the subject has worked out the sorting criterion the experimenter is using they are allowed to place ten cards correctly, then the experimenter changes to another sorting criterion without warning. Neurologically intact individuals typically catch on to the procedure quickly and make few errors, these being immediately after the change of criterion. Patients with frontal lobe damage make many errors, typically involving the inability to discover the sorting strategy or inability to change strategies despite repeated negative feedback.

Sorting objects according to their features is the type of well-learned behaviour we would expect to find as a high-level schema in CS. The CS/SAS model would most

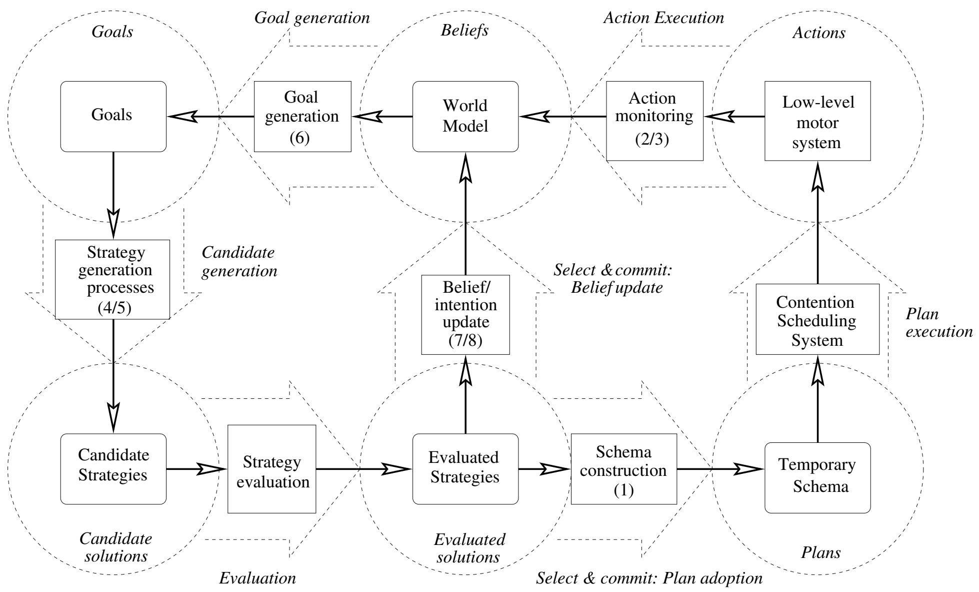

Figure 2: The SAS outline of Shallice and Burgess (1996) mapped on to the Domino framework of Das et al. (1997) (broken lines). Numbers in brackets refer to processes as identified by Shallice and Burgess.

straightforwardly address the WCST on the basis that SAS is involved in initial generation of a sorting strategy and configuration of CS, which would then carry out that strategy with subsequent cards unless negative feedback was received, when SAS is again required to generate an alternative strategy. Figure 3 shows a minimal implementation of the system of Figure 2 in the COGENT computational modelling environment which allows the boxes in such "box and arrow" diagrams to be fleshed out with computational specifications so that the model may be executed.

Bearing in mind that at this stage the requirement is simply for an outline model to demonstrate the principle of an SAS implementation, the implementation of Figure 3 is simplified to include only the essential elements of Figure 2. Following Figure 3 in a clockwise direction operation is as follows: "Current beliefs" maintains information from the environment provided by sensory processes. A "Novelty detection" process triggers the generation of a new goal in response to an unexpected situation, which may be the result of novel circumstances or of the failure of an automatised behaviour in CS. The presence of a goal triggers strategy generation processes. A number of such processes may operate in parallel on the problem posed by the goal, potentially yielding more than one candidate solution. A solution evaluation process provides a means of ranking these candidates, yielding the fourth domino "dot", Evaluated Strategies. At this point the highest ranked candidate is selected for implementation. The "current beliefs" are updated to reflect the candidate strategy. Simultaneously, the strategy is enacted via the CS system. This may simply require the activation of an existing CS schema or may involve the construction and implementation of a new temporary schema. A single process (Schema Implementation) is assumed to be responsible for either, resulting in a temporary schema specification which sends activation to existing CS schemas.

Shallice and Burgess suggest a number of procedures for strategy generation in response to a goal, the simplest of which is "spontaneous schema generation" - the propensity of a suitable strategy to simply come to mind in response to a simple problem. In the current implementation a process of this type is simulated by a rule in the "strategy generation" process which may be paraphrased as: If the goal is to sort an item into a category, and the item has distinguishable features, the item may be sorted according to one of those features. Cards are defined as having the features symbol, number and colour, so this rule will always generate three corresponding sorting strategies. The "strategy evaluation" process ranks strategies according to two rules: Strategies which have recently been attempted are ranked lower, and strategies which have recently proved successful are also ranked lower. A strategy which has recently been attempted and has been successful will thus be ranked lowest of all. This simple scheme leads to appropriate strategy-testing behaviour during the WCST task.

The Contention Scheduling system is simulated in the current model by a simple set of processes; a full computational simulation is available which could be used for more detailed modelling (Cooper & Shallice, in press). A single well-learned schema ("match_to_feature") is assumed to be present for placing a held item next to a stimulus matching on a specified feature. This schema may be activated by the SAS simulation along with a token representing the feature to be matched (colour, shape or number).

Performance of the WCST task starts with a request from an external "experimenter" process to sort a card.

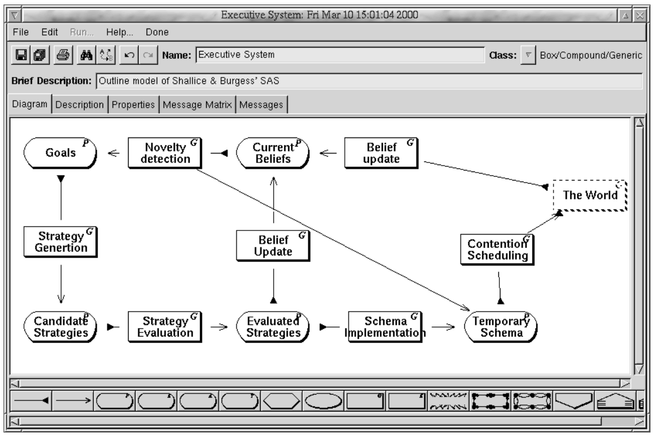

Figure 3: An outline implementation of the Shallice and Burgess SAS in the COGENT modelling system. Rounded boxes are buffers, square boxes are processes. "The world" is an external world representation.

This is placed in "current beliefs" and is treated as a novel event. A goal is thus set to serve this request. This triggers strategy generation which produces three candidate strategies, sort by colour, shape or number. Initially all are equally ranked so one is selected at random for execution. This leads to update of beliefs (with the new current strategy) and to execution of the strategy, which involves activation of the "match_to_feature" schema along with the corresponding feature token in the CS simulation. This schema executes in CS causing the current card to be matched according to the chosen feature. If this action receives positive feedback from the experimenter the SAS takes no further action - as further cards are produced by the experimenter the "match_to_feature" schema remains active and immediately responds by sorting them appropriately. If the experimenter gives negative feedback (which may occur immediately if the wrong sorting strategy has been attempted first, or may occur after a number of correct sorts when the experimenter changes the sorting criterion) the SAS treats this as a novel situation and again raises a goal to find a sorting strategy. Recently tried and recently successful strategies are both ranked lower than untried strategies ensuring a that successful new strategy is rapidly found.

The simulation raises a problem at this point, however. While the SAS simulation is determining a new strategy the CS simulation still has the old strategy active and proceeds to sort the next card despite the negative feedback. Evidently an additional control signal is required to halt automatic behaviour in CS when unexpected feedback is received. Intuitively this seems reasonable: animals have a "startle" reflex which achieves much this result in situations where the habitual response needs to be suppressed. A connection is accordingly added to the simulation (between "novelty detection" and the temporary schema in Figure 3) which removes the current temporary schema when triggered. This in turn removes activation input from the currently active CS schemas and halts automatic behaviour. Table 1 shows sample output from a short run of the WCST simulation.

Table 1: Sample output from a short run of the WCST simulation. The experimenter's criterion is initially to sort by shape, but changes to sort by colour after three correct responses.

Card to sort	Model's response	Feedback
4 blue squares	place with 4s	wrong
2 green triangles	place with triangles	correct
1 red square	place with squares	correct
3 blue circles	place with circles	correct
2 green circles	place with circles	wrong
1 red triangle	place with reds	correct
2 blue squares	place with blues	correct

5.2 Discussion

While the model described here is certainly highly simplified and just as certainly incomplete, it represents a first step towards a psychologically plausible simulation of major aspects of the SAS. The current simulation is not detailed enough to allow very specific claims to be made about the origin of errors in the WCST following frontal-lobe damage, but some general points can be raised. The best known error type, perseverative responding (i.e. failure to adjust to a new sorting strategy when the experimenter changes the sorting criterion) may implicate a number of systems. For example, negative feedback may fail to result in the generation of a goal to change behaviour; candidate strategies may not be correctly weighted, so that the previously successful strategy is chosen again despite having been recently used and having elicited negative feedback; or the process of de-selecting the current schema in CS may be defective. Perseverative behaviour can be simulated in the model in any of these ways and a more detailed simulation, including a full simulation of contention scheduling, may provide a better basis for disambiguating these possibilities.

A more general benefit of an SAS simulation is the possibility of investigating the interface between SAS and CS. Learning is one important target for investigation. The CS system is held to acquire new schemas as a result of repeated application of the same strategy by SAS in similar situations. Once a schema has been acquired the SAS is able to delegate operation to it without having to explicitly control behaviour. A number of processes are implicated in this SAS-to-CS transfer which cannot be studied without adequate characterisations of the two systems.

Another aspect of the interaction between SAS and CS is the need to remove the temporary schema (and possibly also deselect CS schemas) in response to novelty. Interestingly such behaviour is also found in robot control systems where a sufficiently powerful top-level executive system is present. For example an autonomous spacecraft control system demonstrated recently by NASA (Muscettola et al. 1998) includes a process which puts the spacecraft into a "standby" mode - suspending routine operations - when an anomalous event occurs. Operation resumes when the anomaly has been analysed by executive systems and a new plan of action generated to deal with it. The need to add this behaviour to the model illustrates the advantage of simulation in the analysis of large-scale agent models. The interactions of multiple systems controlling behaviour with each other, with the agent as a whole and with its environment can be difficult to analyse in the abstract.

6 Conclusions

I have argued that architectures for the integration and control of behaviour which have emerged from the study of neuropsychological data and from essentially engineering research into the efficient control of mobile robots are formally similar. While competing positions exist in both fields the apparent convergence of independent work in different domains indicates that this class of mechanism is worth investigation as a candidate architecture for a functional model of mind. Within cognitive psychology a major problem is the obscurity of higher-level processes. I have suggested that theories in AI, which are typically more focussed on higher cognitive functions, may point the way to appropriate decompositions of such opaque processes, and I have offered a preliminary model of the Norman and Shallice SAS as an example.

Theories have been constructed in AI and in cognitive psychology which address the same types of cognitive process, and both disciplines have made great progress in recent years in adding detail to these theories. It seems that both have now reached a level where we can expect each to begin providing useful insights for the other. A dialogue between AI and neuroscience on the problem of the control and integration of behaviour should benefit both fields. Approaches from AI and robotics may shed light on the structure of obscure higher processes in psychology. In turn the increasingly detailed picture of human executive function emerging from neuropsychology can provide a rich context for theories of behaviour integration and control in AI.

Acknowledgements

I am most grateful to Richard Cooper, John Fox, Tim Shallice and Heather Rhodes for numerous discussions on the material in this paper.

References

R. A. Brooks. Intelligence without representation. *Artificial Intelligence 47*, 139-160. 1991.

R. Cooper. The control of routine action: modelling normal and impaired functioning. To appear in G. Houghton (ed.) *Connectionist Modelling in Psychology*. Psychology Press. 2000.

R. Cooper & D. W. Glasspool. *Learning to act.* In submission.

R. Cooper & T. Shallice. Contention Scheduling and the control of routine activities. *Cognitive Neuropsychology.* In press.

S. K. Das, J. Fox, D. Elsdon & P. Hammond A Flexible architecture for autonomous agents. *Journal of Experimental and Theoretical Artificial Intelligence,* 9, 407-440. 1997.

E. Gat. On three layer architectures. In D. Kortenkamp, R. P. Bonnasso and R. Murphey, eds. *Artificial Intelligence and Mobile Robots.* AAAI Press. 1998.

G. W. Humphreys & E. M. E. Forde. Disordered action schema and action disorganisation syndrome. *Cognitive Neuropsychology 15,* 771-811. 1998.

F. Lhermitte. Utilisation behaviour and its relation to lesions of the frontal lobes. *Brain*, 106, 237-255. 1983.

P. Maes. How to do the right thing. *Connection Science*, 1, 291-323. 1989.

H. P. Moravec. The Stanford Cart and the CMU Rover. *Proceedings of the IEEE,* 71(7), 872-884. 1982.

N. Muscettola, P. Nayak, B. Pell & B. Williams. Remote Agent: to boldly go where no AI system has gone before. *Artificial Intelligence* 103(1-2):5-48. 1998.

N. J. Nilsson (Ed.). *Shakey the robot.* SRI AI Center technical note 323. 1984.

D. A. Norman & T. Shallice. *Attention to action: Willed and automatic control of behaviour. Center for Human Information Processing* (Technical Report No. 99). University of California, San Diego. 1980.

D. A. Norman & T. Shallice. Attention to action: Willed and automatic control of behaviour. Reprinted in revised form in R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.) 1986. *Consciousness and self-regulation, Vol. 4* (pp. 1-18). New York: Plenum Press. 1986.

J. T. Reason. Lapses of attention in everyday life. In W. Parasuraman and R. Davies (eds) *Varieties of Attention*, pp. 515-549. Orlando, FL: Academic Press. 1984.

M. F. Schwartz, E. S. Reed, M. W. Montgomery, C. Palmer & N. H. Mayer. The quantitative description of action disorganisation after brain damage: a case study. *Cognitive Neuropsychology*, 8, 381-414. 1991.

T. Shallice & P. Burgess. The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London B.* 351, 1405-1412. 1996.

T. Shallice & P. Burgess. Deficits in strategy application following frontal lobe lesions. *Brain*, 114, 727-741. 1991.