

Learning Action Affordances and Action Schemas*

Richard Cooper (R.Cooper@psyc.bbk.ac.uk)
School of Psychology, Birkbeck College, University of London,
Malet St., London, WC1E 7HX

David Glasspool (dg@acl.icnet.uk)
Advanced Computation Laboratory, Imperial Cancer Research Fund,
Lincoln's Inn Fields, London, WC2A 3PX

Abstract

Several theories of action selection assume that the environment both suggests and constrains possible actions at each moment. We present an interactive activation model (based on an existing model of routine sequential action selection) in which actions are organised into partially ordered schemas for simple task elements, and the current state of the environment contributes to selecting actions and schemas. Previous versions of the model have not accounted for learning. We show that a simple reinforcement learning paradigm allows environment-action associations to be acquired through unguided exploration of the environment. The basic model is limited in the types of environment/action associations that it acquires. We explore ways in which greater diversity of learning (and behaviour) may be achieved, and suggest that, in order to acquire a broad range of environment/action associations, mechanisms for boredom avoidance or novelty seeking are required.

1 Introduction

The task of action selection is commonly argued to be mediated by constraints imposed by the environment. Thus, Thorndike (1898) argued that behaviour was structured into habit families, in which any situation (or stimulus) was associated with a hierarchically structured set of responses. Thorndike held that the response selected at any point in time was that at the top of the currently applicable habit family, and that habit families were modified (i.e., learnt) through the laws of exercise (promote or strengthen responses that are attempted) and effect (promote or strengthen responses that are successful but demote or weaken responses that fail or are inappropriate).

More recently, Gibson (1979) proposed that objects and/or situations might “afford” certain responses. Thus, a light switch might afford being flicked on or off, and a knife might afford cutting. While the mechanisms underlying affordances (notably so-called direct perception) are often portrayed as mysterious and anti-cognitivist, affordance-like notions have arisen in the cognitive literature. In the theory of automatic and willed action control proposed by Norman & Shallice (1980, 1986), for example, action schemas (corresponding to abstractions over individual actions or action sequences) participate in an interactive activation network, with action being controlled by the most highly active schema nodes. Environmental triggering is held to be a significant source of excitation within the network. Indeed, such triggering

*This work was supported in part by grant RSRG 20546 from the Royal Society to Richard Cooper and grant #R01 NS31824-05 from the National Institutes for Health to Myrna Schwartz.

is assumed to underly some common forms of action lapse (e.g., capture errors: cf. Reason, 1984) and behaviours of certain neurological patients (such as utilisation behaviour (Lhermitte, 1983), where patients appear to be unable to inhibit environmentally appropriate actions).

Similar requirements are apparent in approaches to action selection developed within Artificial Intelligence. Thus, excitation of nodes within Maes (1989) interactive activation network for sequential action control is partially dependent upon the satisfaction (by the current state of the environment) of action preconditions.

The common elements of the above theories are that, at any point in time, the environment is held to make a set of possible actions available, and that behaviour involves selecting and performing one action from this set. That is, the theories all assume environment/action associations (affordances, triggering conditions or preconditions). However, none can provide an account of the acquisition of these associations.

This incompleteness is particularly well illustrated in the case of Norman & Shallice's "Contention Scheduling" account of automatic action control. An implementation of the account has recently been developed (Cooper & Shallice, 2000; see also Cooper *et al.*, 1995). The model is able to produce complex well-formed action sequences, and is also able to account for a range of errors (including those seen in normals when distracted or fatigued and those characteristic of a number of classes of neurological patient). An important part of the model is the environmental triggering of schemas: following Norman & Shallice (1980, 1986), each schema has a triggering function that determines the extent to which it is excited (or inhibited) by the state of the environment at any point in time. A significant weakness of the model, however, is that it is unable to acquire these triggering functions — they must be specified by hand for all schemas within the model's repertoire. This raises both practical difficulties (specifying such triggering functions by hand is time-consuming and error-prone) and methodological difficulties (countless degrees of freedom are introduced into the model because the theory fails to specify the degree to which a given state of the environment triggers each action schema). Although the severity of these difficulties may be reduced by the incorporation of general principles governing the form of triggering functions, the question of acquisition remains.

There is thus a need for a fully explicit computational account of the acquisition of environment/action associations. Thorndike's laws of exercise and effect provide a starting point for such an account. The remainder of this paper presents a model that uses reinforcement learning (effectively implementing Thorndike's law of effect) to acquire environment/action associations within the context of the Cooper & Shallice model of routine action selection. The basic premise of the model is that environment/action associations are learnt through trial and error with feedback, with the conditions under which successful actions are performed being reinforced beyond the conditions under which unsuccessful actions are performed. The basic model is only able to acquire a small set of environment/action associations. We therefore explore some simple modifications and extensions to the basic model. These explorations suggest that, in order to acquire appropriate associations relating to a broad range of actions, mechanisms for "boredom avoidance" or "novelty seeking" are required. We conclude by relating our findings to more general problems in learning to act, especially those concerning the learning of action sequences.

2 The Basic Model

An essential feature of this account of the acquisition of environment/action associations is that the shape of such associations is moulded by what is possible within the environment. Actions should only be associated with, or triggered by, states of the environment in which they may be successfully realised. It is therefore essential to develop a sufficiently realistic model of the environment in tandem with the model of the agent. The COGENT modelling environment (Cooper & Fox, 1998) was used to implement and evaluate such a system, with the agent model and the environment model encapsulated in separate, communicating, functional modules.

2.1 The Model of the Environment

The environment model is based on the coffee preparation domain used by Cooper *et al.* (1995) and Cooper & Shallice (2000). It consists of a set of objects, two hands, and a set of functions for manipulating objects with the hands. Each object has a set of features including size, shape, position, and relevant aspects of state. Thus, the world includes several containers and implements (an empty mug, a jug of water, a closed bowl of sugar, a teaspoon and a coffee stirrer). The manipulation functions correspond to basic actions (pickup, discard, open, close, pour, dip spoon, empty spoon and stir). Each includes a set of preconditions that specify the physical conditions under which the action may be successfully performed. The functions are called at various times by the agent, as described below. If a manipulation function's preconditions are satisfied by the state of the world when the function is called, then the world is modified according to that function and positive feedback is returned to the agent. If the preconditions are not satisfied, the world is not modified and negative feedback is returned.

2.2 The Model of the Agent

We assume that the agent is capable of a variety of basic actions, corresponding to (a subset of) the manipulation functions available within the environment. The agent's task is to learn to select actions through time that are compatible with the changing state of the environment. Essentially the agent must internalise action preconditions. Minimally, the agent comprises three processes: perceptual processes that transduce the state of the environment into internal representations; selection processes that select one action from those available to the agent on the basis of its internal representation of the environment, and a learning process that uses feedback from the environment to shape the selection process.

The perceptual processes are finessed by a set of rules that map each object in the environment to a feature vector (of width eight in the implementation reported here). Essentially the environment is modelled in symbolic terms, and the perceptual processes re-represent each object as a vector of features. The internal representation of the complete environment is therefore a set of such vectors. For present purposes such vectors include both perceptual and functional features, although the functional features employed (e.g., *implement*) may be taken as short-hand for simple combinations of perceptual features (e.g., *one primary axis* and *between 150 mm and 300 mm in length*). The approach is similar in spirit to that of Plaut & Shallice (1993).

Following Norman & Shallice (1980, 1986; see also Cooper *et al.*, 1995 and Cooper & Shallice, 2000), we use an interactive activation network to mediate the selection process. There is one node in this network for each action. Processes of lateral inhibition and self activation operate over the nodes to ensure that at most one node is highly active at any time. Nodes may also receive excitation (or inhibition) from the representation of the environment (as described below). Selection occurs when a node's activation exceeds a threshold.

Performance of an action entails selecting objects on which to act (and effectors with which to act). Thus, once a basic action such as pickup has been selected, it is necessary also to select an object to pick up and an effector to do the picking up. Not all actions require one object and one effector, and different actions have different argument requirements (e.g., pickup requires an effector that is not full, but discard requires an effector that is holding an object). Each action must therefore have an associated set of argument selection criteria.

In the model of Cooper & Shallice (2000), selection criteria were functions that mapped object representations to numeric values. The numeric value essentially indicated the appropriateness of an object or effector for the given argument role. In the current model these selection criteria must be learnt. We thus associate with each action a list of argument roles. The number and type of arguments for an action (e.g., two objects and one effector) is assumed to be intrinsic to the action (and hence hard-wired), but the mapping of objects and effectors to these argument roles is assumed to be subject to learning. Each argument role is therefore assumed to correspond to a vector of weights, with the appropriateness of an object for an argument role being determined by the dot product of the object's featural representation and the argument role's weight vector. Learning alters weight vectors, leading to more accurate measures of appropriateness.

While the specification of argument roles in terms of weight vectors addresses the issue of argument selection, the issue of action triggering by the representation of the environment remains. Separate triggering functions were employed in the model of Cooper & Shallice (2000), but such functions introduce additional degrees of freedom and are not required. The current model assumes that action nodes are triggered to the extent that their argument selection criteria are met by objects in the model's representation of the environment. Thus, selection criteria do double duty: once an action has been selected they work to select object representations to fill the action's argument roles, but prior to selection of an action they mediate excitation and inhibition of the action's node within the interactive activation network.

After the agent model selects and attempts an action, it receives feedback from the environment model. If the action succeeds the feedback is positive, and each argument and effector weight vector of the action is adjusted through delta rule learning so that its dot product with the representation of the actual object or effector used is nearer +1. If the action fails the feedback is negative, and each argument and effector weight vector of the action is adjusted through delta rule learning so that its dot product with the representation of the actual object or effector used is nearer -1. In all cases the node within the interactive activation network corresponding to the attempted action is subsequently inhibited, allowing other nodes to become active and hence other actions to be performed.

2.3 Functioning of the Model

In order to establish that appropriate selection restrictions would yield a model capable of generating random sequences of appropriate actions, the model's behaviour was first assessed with hand-coded selection restrictions. The selection restrictions in this "expert" model embodied all action preconditions. For example, the featural encoding of objects included several features indicating an object's position. One of these features was set (by perceptual processes) to +1 for objects that were held and -1 for objects that were not held. In the hand-coded selection restriction for the object argument role of pickup the corresponding feature was set to -1. Consequently, objects that were not held would tend to match the selection restrictions of pickup better than objects that were held. This matching would also lead to greater excitation of the pickup action by objects that were not held than by objects that were held.

This assessment was successful in that, with appropriate parameter settings (persistence = 0.85; self activation = 0.50; lateral inhibition = 0.50; environmental triggering = 0.20; selection threshold = 0.60; noise variance = 0.001), the hand-coded selection restrictions yielded well-formed, but somewhat aimless, sequences of actions. Figure 1 (left) shows one such sequence, and figure 1 (right) shows the activation profiles of node throughout this sequence.

To investigate learning, selection restrictions for all actions were initialised to random vectors near zero. (Specifically, the components of each vector were initialised to random values normally distributed around 0 with variance 0.01.) The same parameters were used in the learning model as in the expert model, with the exception of self activation and lateral inhibition, which were both increased to 0.60, and the addition of the learning rate parameter, which was set to 0.05. The increase in self activation and lateral inhibition was necessary to ensure robust functioning of the learning model during the early stages of learning. Prior to learning, little excitation is provided by the environment. Additional excitation within the action node network is needed to ensure that action nodes (including those corresponding to inappropriate actions) may become sufficiently active to allow action selection.

38: pickup stirrer with left hand
88: stir jug with stirrer
133: discard contents of left hand
144: pickup teaspoon with left hand
188: stir jug with teaspoon
211: dip teaspoon into jug
252: discard contents of left hand
272: pickup teaspoon with right hand

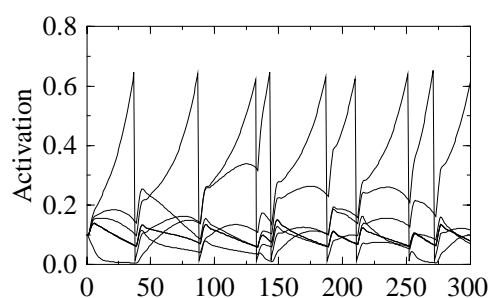


Figure 1: Left: A well-formed sequence of actions generated by the model when given hand-coded selection restrictions. The left column shows the time of action initiation, in terms of processing cycles. Right: Activation profiles of all action nodes throughout the sequence shown on the left. Time, in terms of processing cycles, is shown on the horizontal axis. Peaks in activation correspond to selected actions.

Figure 2 (left) shows the behaviour of the model prior to learning. Actions are attempted apparently at random, and generally with inappropriate arguments. In this example, only one of the first ten actions is well-formed (that of picking up the mug with the left hand). However, each time an action is attempted its selection restrictions are adjusted as described above. Consequently, the model rapidly discovers appropriate and inappropriate actions, internalising an approximation to each action’s preconditions. In general, 100 to 150 action attempts are required before the model’s approximate selection restrictions are sufficient for error-free action performance. Figure 2 (right) shows a fragment of such performance.

3 Towards Diversity

Although the action sequence of figure 2 (right) is error-free, it is also highly stereotyped. Two complementary actions are repeated *ad nauseum*. The model has not learnt selection restrictions for other actions, and examination of the selection restrictions it has learnt reveals that they are highly specialised (applying in the case shown in figure 2 (right) to the stirrer and the left hand, but not to other objects or to the right hand). Different runs of the model do yield different long-term behaviour (with different acquired selection restrictions), but that behaviour typically comprises performance of two complementary actions (such as pickup and discard, or open and close) with only a limited subset of objects and effectors.

In order to acquire general selection restrictions for the full range of actions it is necessary to successfully perform the full range of actions with a variety of arguments. Two factors prevent the model as described above from acquiring full generality. First, the only exogenous source of activation to the action network is due to environmental triggering by appropriate actions. Because reinforcement strengthens such triggering, actions whose selection restrictions are acquired first will dominate and be performed in favour of other actions, preventing attempts at such other actions and hence preventing the acquisition of triggering conditions for other actions. Second, argument selection is determined entirely through selection restrictions, so, once selected, an action will always be applied to the object that best matches its selection restrictions.

32: empty mug into stirrer*	
50: stir teaspoon with mug*	2978: pickup stirrer with left hand
58: pour mug into jug*	2989: discard contents of left hand
105: close jug with left hand*	3001: pickup stirrer with left hand
126: open bowl with right hand*	3010: discard contents of left hand
160: discard contents of right hand*	3022: pickup stirrer with left hand
187: pour stirrer into stirrer*	3031: discard contents of left hand
200: pickup mug with left hand	3043: pickup stirrer with left hand
217: open stirrer with left hand*	3053: discard contents of left hand
243: pickup mug with left hand*	

Figure 2: Left: A fully random action sequence, containing mostly errors, generated by the model prior to learning selection restrictions. Erroneous actions are marked with an asterisk. Right: An error-free action sequence following acquisition of approximate selection restrictions.

34868: discard contents of left hand
34889: pickup jug with left hand
34908: close jug using right hand
34924: open jug using right hand
34930: discard contents of left hand
34947: pickup jug with left hand
34972: close jug using right hand
34992: open jug using right hand
34995: discard contents of left hand

Figure 3: Action sequences generated when extreme inhibition (500 units) is applied to attempted actions.

The first difficulty appears to require further exogenous sources of excitation and inhibition to the action network. In particular, it appears that some mechanism is needed to inhibit or suppress actions whose triggering conditions have been acquired. We consider one such mechanism below.

The second difficulty is solved in the closely related model of routine action selection (without learning) of Cooper & Shallice (2000) through the incorporation of interactive activation networks for object representations and effectors. The networks, analogous to the action network, include one node for each represented object, and one node for each effector. Nodes receive excitation and inhibition from action nodes (mediated by selection restrictions) and (potentially) from attentional processes. In essence, the use of object representation and effector networks allows greater control over the arguments that are selected. Below we therefore consider augmenting the existing model with such networks.

3.1 Incorporating Action Inhibition

When the activation of an action schema node exceeds a threshold (0.6 in the above simulations) the action schema is selected, its argument roles are filled, and the resulting action is performed. The action schema node is then inhibited, allowing other action schema nodes to become active. Moderate inhibition (5 units in the above simulations) returns the selected action schema node's activation to approximately rest (a magnitude comparable to that of its competitors), allowing competition between all action schema nodes to proceed afresh.

However, by dramatically increasing the inhibition on action schema nodes once they have been performed, such schema nodes can be effectively removed from competition for an extended period, giving greater opportunity for the selection of other schemas, and thus encouraging greater diversity of action. Figure 3 shows action sequences produced by the model when inhibition was increased to 500 units.

Although high levels of inhibition result in the model taking significantly longer to acquire selection restrictions that reliably yield appropriate action selection, greater diversity is apparent in the selection restrictions thus acquired. With inhibition at 500 units, reasonable approximations to selection restrictions for four distinct actions are acquired. This approach does not, however, lead to greater diversity in the objects to which actions are directed. This is reflected in the acquired selection restrictions,

which still converge to features of specific objects. Thus, the object role of pickup converges to match the favoured object (the open jug, in the case shown in figure 3), and not to match just the features of that object that are relevant to the action (e.g., that the object is not already in hand). Nevertheless, the approach does demonstrate that learning can be improved by manipulating the actions that are performed.

3.2 Incorporating Additional Networks

Incorporation of object representation and effector networks as described above into the model pose no difficulties. Activation flow within the additional networks is governed by the same principles (including self activation and lateral inhibition) as that within the action network. In addition, object representation nodes are excited or inhibited by action nodes in proportion to the product of the activation of the action node and the degree to which the corresponding object representation matches the action's argument selection restriction, and *vice versa* (i.e., action nodes are equivalently excited or inhibited by object nodes). Nodes in the effector network interact with the other networks in an analogous fashion.

On selection of an action, the action's arguments are filled by those objects and effectors that are both highly active and a good match for the action's selection restrictions. Arguments are selected to maximise the product of these two factors. Figure 4 shows an action sequence generated (after learning) by the augmented model. The model includes moderate inhibition of actions, object representations and effectors on each action attempt. The figure may thus be compared with figure 2 (right).

The introduction of object representation and effector networks results in considerable diversity of argument selection. Although only two action schemas are selected (cf. figure 4) both hands are employed to manipulate a variety of objects. The model has therefore acquired less restrictive, and more appropriate, selection restrictions for these two actions.

4 General Discussion

We have shown that a schema-based model of environmentally triggered action selection may acquire appropriate environment/action associations through trial and error. Diversity of behaviour arises when the basic model is augmented with mechanisms to

4131: pickup bowl with left hand
4139: pickup teaspoon with right hand
4144: discard contents of right hand
4150: pickup stirrer with right hand
4158: discard contents of left hand
4164: pickup jug with left hand
4167: discard contents of right hand
4173: discard contents of left hand
4178: pickup stirrer with right hand

Figure 4: Action sequences generated when object representation and effector networks are employed.

inhibit action reselection and to differentiate objects on the basis of activation. However, the use of extreme inhibition to encourage variety of action results in a system that is constantly switching between actions. As is apparent from figure 3, this dramatically slows learning. A more sophisticated mechanism might monitor learning and/or behaviour and manipulate excitation of both action schemas and their arguments in order to optimise learning. Such a mechanism might, for example, embody some notion of boredom detection, repeating behaviours until they are well learnt and then inhibiting them in favour of less well learnt behaviours. (Curiously, infants show just this kind of behaviour.)

The necessary distinction between novel and routine behaviour is already present in the Norman & Shallice (1980, 1986) framework. The former is handled by a supervisory attentional system (SAS) that responds to a novel situation by generating a strategy for dealing with it. If an initially novel situation is repeated the required strategies eventually become well learnt and may be automatically executed by contention scheduling (CS), the routine action selection system, without the intervention of SAS. At this point the situation has effectively become routine. Mechanisms thus already exist for (i) detecting whether a well-learnt schema exists in CS to deal with the current situation and (ii) supervising the learning of a new action schema if required. In its normal mode of operation the CS/SAS system employs existing schemas wherever possible, and only uses the limited resources of the SAS when necessary. We can however envisage a different mode of operation (“exploration” or “play”) in which this precedence is reversed so that situations that stimulate the SAS are deliberately sought, and activity is abandoned when it becomes routine. We hypothesise that such an operational mode is essential to the acquisition of general selection restrictions for a broad range of actions.

While the selection restrictions acquired by the model may be viewed as serving the function of Gibsonian affordances, they are not true affordances in the Gibsonian sense. To Gibson, the concept of an affordance went hand in hand with that of direct perception. Affordances were held to be directly available from the environment, without the integration of perceived features into object representations. Because selection restrictions within the model are object-based, the model requires that features are first bound together as objects before they can affect an action schema’s activation. Within the framework adopted here, the binding of features into objects is a necessary precondition for action triggering. The model thus sides against direct perception.

The model is also intended only as a fragment of a model of intelligent action selection. While the model does acquire action preconditions, it lacks goal directedness. The mechanisms of action selection embodied within the model are compatible, however, with those required by CS/SAS theory, and consistent with those of Cooper & Shallice’s (2000) implementation of CS. In fact, the model has implications for the CS implementation, suggesting that triggering functions and selection restrictions, currently separate elements of CS, might be merged.

Contention scheduling is a general mechanism for schema selection that may operate at multiple levels over a hierarchically structured sets of schemas. The model developed here illustrates learning only at the lowest level. Two additional mechanisms are required if the current learning mechanisms are to generalise to higher-order action schemas. First, the model must be instructable via higher-order supervisory processes.

This is relatively straightforward: such a process may, through selective excitation and inhibition of nodes in the various networks, exert detailed control over behaviour. Second, the model must be able to acquire and represent higher-order action schemas, including appropriate sequential relations within those schemas. Two factors drive sequential behaviour in the current model: The changing state of the environment, and the inhibition of previous actions which prevents their immediate repetition. An existing class of models of sequential behaviour, Competitive Queueing models (e.g., Houghton, 1990; Glasspool, 1998), have successfully accounted for many aspects of serial behaviour in similar terms. In these models actions are given a *gradient* of activations, such that the earlier an action is to be performed the more active it is. Sequential behaviour is achieved by repeatedly performing and inhibiting the most active action. Since appropriate selection and inhibition mechanisms already exist in the current model, the learning of simple sequential behaviour might involve nothing more than the acquisition of an appropriate activation gradient over actions. This may again be achieved by reinforcement learning, with actions that occur too early receiving negative feedback and actions that occur too late receiving positive feedback. Further work will explore this possibility.

References

- Cooper, R., & Fox, J. (1998). COGENT: A visual design environment for cognitive modeling. *Behavior Research Methods, Instruments, and Computers*, *30*, 553–564.
- Cooper, R., & Shallice, T. (2000). Contention Scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*, 297–338.
- Cooper, R., Shallice, T., & Farrington, J. (1995). Symbolic and continuous processes in the automatic selection of actions. In Hallam, J. (Ed.), *Hybrid Problems, Hybrid Solutions*, pp. 27–37. IOS Press, Amsterdam.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin, Boston, MA.
- Glasspool, D. W. (1998). *Modelling Serial Order in Behaviour: Studies of Spelling*. Ph.D. thesis, Department of Psychology, University College, London, UK.
- Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In Dale, R., Mellish, C., & Zock, M. (Eds.), *Current Research in Natural Language Generation*, chap. 11, pp. 287–319. Academic Press, London, UK.
- Lhermitte, F. (1983). Utilisation behaviour and its relation to lesions of the frontal lobes. *Brain*, *106*, 237–255.
- Maes, P. (1989). How to do the right thing. *Connection Science*, *1*, 291–323.
- Norman, D. A., & Shallice, T. (1980). Attention to action: Willed and automatic control of behavior. Chip report 99, University of California, San Diego.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In Davidson, R., Schwartz, G., & Shapiro, D. (Eds.), *Consciousness and Self Regulation: Advances in Research and Theory, Volume 4*, pp. 1–18. Plenum, New York, NY.
- Plaut, D. C., & Shallice, T. (1993). Perseverative and semantic influences on visual object naming errors in optic aphasia. *Journal of Cognitive Neuroscience*, *5*, 89–117.
- Reason, J. T. (1984). Lapses of attention in everyday life. In Parasuraman, W., & Davies, R. (Eds.), *Varieties of Attention*, chap. 14, pp. 515–549. Academic Press, Orlando, FL.
- Thorndike, E. L. (1898). Animal intelligence: An experimental study of associative processes in animals. *Psychological Monographs*, *2*.