Computational constraints on associative learning

Edmund Shing[1]
Cognitive Science Research Centre,
School of Computer Science, University of Birmingham
Birmingham B15 2TT, United Kingdom

Abstract

Due to the dynamic nature of the real world, learning in intelligent agents requires various processes of selection (‘attention’) of input features in order to enable computational tractability. This paper looks at associative learning and analyses the selection processes necessary for this to work effectively by avoiding the combinatorial explosion problem faced by an adaptive agent situated in a complex and dynamic world. Analysis suggests that adaptive agent architectures require selection processes in order to perform any “useful” learning. An agent design is constructed following a “broad and shallow” approach to meet both general (e.g. related to fundamental properties of the real world) and specific (e.g. related to the specific theory proposed) requirements, concentrating on learning and selection mechanisms in the implementation of reinforcement learning.

1 Introduction

Processes of learning and attention are necessary for intelligent agents such as people to function effectively in an ever-changing world. These processes allow both humans and animals to survive in a dynamic and potentially dangerous environment by learning both which external events are good predictors of other salient events and also which of their own actions can achieve desirable outcomes or prevent undesirable ones. At any point in time, all manner of unexpected events can affect an agent’s state and cause it to behave in such a way as to adapt to this change in the world state. This adaptation allows the agent to continue to strive towards its goals despite changes in the external environment.

This research looks at associative learning and examines the selection processes necessary for this to work effectively. Analysis suggests that adaptive agent architectures require attentional processes in order to perform any “useful” learning since otherwise they may be overwhelmed by the combinatorics. In addition, reinforcement learning coupled with certain simple selection, monitoring and evaluation mechanisms can achieve several seemingly more powerful forms of learning than is at first apparent.

1.1 Overview of approach

What types of architecture and mechanisms do biological agents employ, and by what are they constrained? I review the current models of learning and attention and conclude that they do not adequately analyse the interdependence of learning and attention in intelligent agents. A “design–based” approach (Sloman, 1994) is adopted that integrates work from various cognitive science disciplines with a rigorous methodology for designs for intelligent agent architectures. An application of this approach to examine the relationship between learning and selection processes in a computer simulation is then described.

According to the design–based approach there are several stages of analysis to perform, including specification of requirements for the proposed model, explorations of possible

[1] This paper was written under the supervision of Professor Aaron Sloman & funded by the Renaissance Trust and the UK Joint Research Council Initiative

designs that might satisfy these requirements and implementation of a subset of these designs in a computer simulation. Results of the simulation can then be fed back to revise the requirements and design; general principles which apply to the design of intelligent agents should be derivable from this iterative process. These principles may aid in the explanation of human and animal capabilities and limitations.

1.2 Definitions

Learning has been defined in different ways according to the different approaches to its study, although these definitions include references to changes made in a system or agent as the result of experience usually with the purpose of improving the agent's ability to achieve its goals. I look at a definition of learning in information-theoretic terms, as a many-to-few mapping in which the inputs are mapped to related outputs with maximal information preservation and introduction of the minimum of ambiguity (equivocation). Thornton (1992) looks at the definition of the similarity-based learning paradigm in these terms; I apply similar techniques to the definition of reinforcement learning, in particular habituation and classical conditioning.

A common element in definitions of attention is selection at one level or another; this is the essential element crucial for learning. There are several levels at which selection processes can operate: in visual attention alone, selection mechanisms operate in the control of head/body movements, eye movements, inhibitory beam and adaptation (Tsotos, 1990), let alone other modules in a cognitive architecture. I argue that selection processes are necessary for a classical conditioning learning mechanism embedded in an intelligent agent architecture.

1.3 The nature of the problem

Associative learning processes are potentially computationally intractable for an agent exposed to many stimuli and able to perform many possible actions due to both the generality of associations that could be learned and the inherently complex and dynamic nature of the real world. However biological agents manage to utilise this form of learning to adapt to the changes in the real world that occur in everyday life. Analysing the complexity of the task can both demonstrate this inherent intractability and allow us to investigate approximations for transforming associative learning into a tractable process.

The "competence" of animal conditioning mechanisms can be inferred from the "performance" of biological agents on conditioning tasks, as is done in work on language acquisition (Pinker, 1990). However, the animal conditioning literature affords only a piecemeal analysis of the performance limits inherent in biological associative conditioning mechanisms. Formal analyses of the nature both of the task and of the world in which the task is set allows much of the experimental work to be characterised explicitly in terms of real-world constraints and possible underlying mechanisms for such adaptive agents.

1.4 General requirements

There are several general requirements that a situated agent design should take into account (Beaudoin, 1994):

- An unpredictable multi-agent environment
- A fast-changing world
- Physical resource limits
- Multiple asynchronous goal processing
- Modularity and coarse grained parallelism

These requirements lead to several constraints on agent design such as incomplete world knowledge at any time, the need for interruptible processing (e.g. the interruptability of planning), and the ability to perform at any one time only a subset of the agent's possible actions. In addition, sensory and motor systems are constrained physically in situated agents; for instance, we can only see a portion of the world and also only be in one place and do so much with our hands at any one time. The cognitive architecture of situated agents therefore has to satisfy these "hard" constraints; selection mechanisms that enable agents such as people to cope with these constraints include covert and overt visual attention (Allport, 1989) and action selection mechanisms (Norman & Shallice, 1986).

2 Problem analyses

The more complex the world is, the more input features there are that can be associated by an associative learning system. This leads to a combinatorial explosion of processes searching for associations among sets of features (Ballard, 1992) as the power set of n elements will contain 2^n combinations (unordered); Hinton (1987) makes a similar point with weight revision algorithms in current connectionist models of learning.

The computational demands of the task of associative learning thus constrain the "space of possible designs" (Sloman, 1994), effectively requiring the implementation of a strategy for reducing the task to a computationally tractable form. I argue that this is effected in animals via the process of associative learning in which selection processes are implicit and via other selection mechanisms at various points in agents' architectures from perception through to action. Associative learning can thus be constrained so as to be tractable both by applying approximations and by optimising the resources devoted to associative learning. I go on to argue that the forms of selection used to heuristically prune the "learning space" can be justified both in information theoretic terms and also on the basis of simple assumptions about the nature of the real world.

2.1 Complexity considerations

Complexity-level analysis assesses the amount of computation required to solve a given problem and the number of elements (processors, connections, memory etc.) needed for its computation (Tsotos, 1990). The complexity measures used here relate to space and time requirements. Complexity-level analysis tries to match a proposed solution to a pre-specified set of resources. The two principles that Tsotos uses in his analysis of vision are complexity satisfaction and minimization of cost. Considerations about the computational complexity of the conditioning task are critical and lead directly to "hard" constraints on the architecture of adaptive systems, both biological and computational.

What is the inherent computational complexity of associative learning? In order to answer this, we first need to define the task in computational terms. The best studied form of associative learning in animals is classical conditioning. In classical conditioning, the aim is to associate conditioned stimuli (CSs) with the occurrence or absence of subsequent or co-occurrent unconditioned stimuli (USs); having formed this predictive association, a second association is formed between the conditioned stimulus/i and a set of appropriate responses (e.g. avoidance of a shock or running to a feeding box to receive a food pellet). A rat that is shocked a few seconds after hearing a buzzer sound will learn to associate the sound with the subsequent shock (a stimulus-stimulus association) despite all the other stimuli in the environment and will then learn to associate the sound with jumping over a barrier into a different area to avoid the shock – the Miller-Mowrer two-process theory of conditioning[2] (Walker, 1987). Classical conditioning[3] can thus be defined as the formation of two sets of mappings:

> The acquisition of a many-to-few mapping between one set of sensory patterns (representations of conditioned stimuli – CSs) and another (representations of unconditioned stimuli – USs) such that CS presentation can trigger the representation of a US (prediction learnt);
>
> The acquisition of a mapping between the representation of a US and a conditioned response (CR) which avoids a negative US or increases the frequency of a positive US (action response learnt).

There are thus two main features in classical conditioning; the first concerns maximising the predictability of USs on the basis of information in the environment, and the second concerns maximising the triggering of appropriate responses on the basis of these US predictions. These responses act to enable the agent receiver to avoid negative USs (such as shock) and to trigger positive USs (such as food delivery)[4].

Another question that needs to be answered before proceeding further with the complexity analysis concerns the exact phenomena being modelled: in what forms can classical conditioning be manifested in real animals?

In a typical classical conditioning scenario where a rat is in a partitioned cage. Unconditioned stimuli (USs) usually take the form of delivery of a food pellet (a positive reinforcer) or an electric shock applied to the feet (a negative reinforcer). Stimuli used as CSs to signal onset of a US are lights of various colours and sounds produced by buzzers or bells. In the simple Pavlovian case, a bell ringing is paired with food delivery; the dog will learn to salivate when the bell rings, even when food is not given. Thus the dog has paired the bell ringing (CS) with the representation of food (US) and produces the CR (salivation) when the representation is activated. There are several distinct phenomena observed in animal conditioning experiments which need to be accounted for in any model of classical conditioning:

1. Concurrent presentation of CS and US (as above);
2. Concurrent presentation of multiple CSs and a US, where the single CSs do not by themselves predict US occurrence but combinations do;
3. Single or multiple CS presentation predicting later US, as temporal contiguity is neither necessary nor sufficient for conditioning to take place (Garcia, Erwin & Koelling, 1966);
4. Several CSs presented alone predicting a US, where the concurrent presentation of these CSs does not predict US occurrence (cf. XOR);
5. Latent inhibition, where a CS–US association takes longer to learn if the CS has previously been presented unaccompanied by the US;
6. "Preparedness" (preferential processing) for certain survival-relevant classes of stimuli such as food (Garcia, Erwin & Koelling, 1966);
7. CSs predicting US non-occurrence, as non-occurrence of expected USs can serve to reinforce behaviour (Walker, 1987).
8. Degree of associative learning (strength change of an association) being proportional to the "surprisingness" of the unconditioned stimulus (Rescorla & Wagner, 1972).

[2] This theory does not fit all the classical conditioning data but remains a useful approximation

[3] This definition is dependent on a representational approach to conditioning phenomena (Gallistel, 1990)

[4] Non-occurrence of expected punishment or reward can also act as positive and negative reinforcers

There are several variables to take account of in any mechanism aiming to model the above phenomena:

a Number of perceived stimuli;

b Time period between occurrence of conditioned and unconditioned stimuli;

c Occurrence/absence of unconditioned stimuli at any given time;

d Sign and strength of US-induced reinforcement;

e Number of self-generated actions possible (possible responses);

f Number of trials over which the association is learned;

Given that mammals seem to be able to solve both the AND and XOR problems (points 2 and 4), combinations of stimuli have to be represented separately from single stimulus representations so that they can be independently associated with occurrence/non-occurrence of unconditioned stimuli. Further, stimulus occurrences need to be stored for a certain period of time, as conditioned stimuli may precede unconditioned stimuli by the order of seconds or even longer for certain types of conditioned stimuli (points 3 and 7).

For n stimuli perceived at any point in time, there will be 2^n stimulus combinations. If the environment is sampled at least once a second (typically much more often than this), then for t timesteps there will be up to $t \times 2^n$ combinations. However, occasion setting experiments have demonstrated that rats can learn to associate temporally separated CSs with a US. Thus not only must we consider combinations of stimuli occurring on the same timestep, but we must also consider combinations of stimuli occurring at different timesteps. Thus potentially there are up to $2^{n \times t}$ combinations which might predict US onset.

The values that n and t can take appear to be substantial; pigeons can learn to distinguish between complex visual stimuli such as 160 photographs of real-world scenes and also between complex auditory stimuli such as different styles of classical music (Pearce, 1987). Even a bank of 20 lights yields 1,048,576 possible combinations; if, moreover, all stimulus occurrences are stored for 10 timesteps, then there are up to 2^{200} combinations to take account of. Both in terms of storage and time to process, this is too great a load to manage, especially when you realize that several of the classical conditioning phenomena mentioned above have not yet been considered.

So it would seem that a classical conditioning model requires selection mechanisms in order to heuristically prune the number of stimuli considered down to the minimum required for differentiation; indeed it would seem that animals employ a set of innately-programmed heuristics such as ‘unusualness’ (novelty), ‘similarity’ and temporal contiguity heuristics (Holyoak, Koh & Nisbett, 1989), as well as biases towards certain stimuli such as food or potentially dangerous stimuli.

2.2 Information theory and conditioning

An information-theoretic analysis of classical conditioning can shed light on what forms of selection processes enable effective classical conditioning to take place. Information theory works on the assumption that a message is being transmitted from a source to a receiver (communications theory); we can think of this as data about the external world being perceived and decoded by a situated agent. Information is a measure of the unpredictability of data. Thus a variable that never changes its value can be said to hold no information, whereas one that changes its value unpredictably does (Shannon & Weaver, 1949; Campbell, 1982). Shannon developed an information measure according to which the amount of information in a message was proportional to the degree of “surprisingness” of the message to the receiver. This is represented by the formula $Information(m) = -log_2 P(m)$, where m is the message and $P(m)$ is the probability of occurrence of that message.

We can define learning generally as a mapping from a large input set to a small output set (a many-to-few input to target mapping) in order to reduce uncertainty about the environment by increasing the receiver’s knowledge of the environment. However, in performing such a mapping there will be loss of information which can be seen as introduction of ambiguity, the “information/equivocation tradeoff” (Thornton, 1992). Thus learning should also reduce ambiguity (also called entropy) by preserving information at the same time as performing this many-to-few mapping and thus minimising the ambiguity introduced. According to this definition, then, selection can be seen as the process or collection of processes by which both to sort information-rich from information-poor perceptual stimulus features and also to identify information-rich patterns amongst a mass of information-poor features (as in vision). Indeed, the mammalian nervous system does this implicitly using large proportions of cells which respond only to changes in input and adapt to a continual signal, reducing firing rates in a negative exponential fashion.

The task of reducing a vast amount of incoming information about the world to a set of outputs representing US predictions represents an expensive set of computations in the form of updating set of correlations between CSs and USs and also between sets of CSs identified as predictive of a US and conditioned responses (CRs). How can the agent learn these two sets of mappings effectively, where the first translation effectively leads the agent to ignore large sections of the input data in favour of predictive CSs and then form a second mapping between these predictive CSs and CRs?

Now, if some stimuli appear much more frequently than others in the world, then if these stimuli also do not reliably predict anything about occurrence or absence of USs they should be effectively ignored as they have no information with respect to USs; conversely, stimuli which do play a part in predicting US occurrence/absence contain a high level of information and so should be concentrated on. If we return to the argument that the agent is subject to physical resource limits and so cannot receive all possible messages from the world, then the agent should then select to receive messages that have a high level of information in preference to those that have very low or zero levels. The potential information loss is thus kept to a minimum and the overall level of uncertainty regarding USs is reduced. From looking at the formula for information of a message, one can deduce that the information of a message asymptotes as the probability of that message tends to zero. It is thus in the agent’s interests to ignore messages which are frequently received and which are not relevant to its goals as they will have a relatively low information value and content. In actual biological agents, this is partially implemented in the form of habituation and is an effective means of reducing the set of input messages to be decoded. It is represented in animal learning theories of attention as favouring of novel over familiar non-correlated stimuli (Walker, 1987; Holyoak, Koh & Nisbett, 1989). In this way, uncertainty about the environment is reduced whilst satisfying the physical constraints requirement placed on a situated agent.

Two temporal principles which are based on characteristics of the real world also serve to filter out those stimuli that are unlikely to be predictors of US occurrence, temporal contiguity (preferential linking of stimuli occurring close together in time) and forward causation (perceived cause usually precedes effects).

2.3 Other complexity-reducing strategies

Given that the number of possible features or combinations thereof perceived by the agent needs to be drastically reduced in order to make any form of learning computationally tractable, a number of techniques need to be employed to reduce the learning space while heuristically directing the agent towards stimuli that are more likely to have some connection with the agent's current goals.

One such strategy is the use of a form of stimulus representation that aids these heuristic selection processes. Indexical representation (Agre & Chapman, 1987), which encodes a relation between a particular object in the world and another world object or the agent, satisfies this requirement by associating with each perceived object a relation which can then be selected by the agent in terms of its relevance to the the collection of current goals. A visual attention mechanism is required to select the portion of the world to view at any one time; this reduces the visual stimuli available to the agent's learning processes. Such a mechanism (as part of a animate vision system) is also useful for generating indexical reference. If indexical reference is implemented for the representation of stimuli, then vision can be used as a retrieval mechanism for detailed information about the world; thus the world can be used as a "memory buffer" that can be accessed by visual behaviours (Ballard, 1991). This strategy then implements another form of implicit selection in that goal-relevant stimuli can be easily identified and preferentially processed (assuming that the agent uses goal representations).

If there are a large number of concurrent cognitive processes active in the brain, and a large proportion of these processes require detailed world knowledge, then the perceptual systems (especially the visual system) will be subject to impractical numbers of demands for information. But as the visual system in particular is physically limited, it is often going to be the case that it cannot satisfy all of these demands at any one time – a bottleneck in the agent's cognitive architecture. Similarly a large number of independent cognitive processes concerned with meeting the agent's bodily requirements (water, different kinds of food, mating, warmth etc.) are likely to generate inconsistent goals and will thus place the physically limited motor system under heavy demand; once again there will be occasions when there are mutually inconsistent demands, only one of which can be met by the motor system. These two bottlenecks at the input and output stages of the agent's cognitive architecture thus imply the utility of a limit in the number of concurrent cognitive processes, if only to reduce the number of mutually exclusive demands that the perceptual and motor systems are subject to. Selection mechanisms dealing with both perceptual input and effector output also aid in dealing with these bottlenecks; these take the form of visual attention and action selection mechanisms in biological agents.

I examine the relationship between learning and selection processes in a computer simulation, both looking at an example illustrating some difficulties of learning in the real world and then describing a simplified scenario for investigating possible agent architectures for satisfying these requirements.

3 The simulation domain

The simulation domain chosen should incorporate the same general requirements that a situated adaptive agent is subject to in the real world. The domain chosen is the "nursery", a domain used by members of the Cognition & Affect project at Birmingham University (Beaudoin, 1994). The nursery is divided into four or more separate rooms, each bounded by walls and ditches and with doorways from one room to adjoining rooms. It contains a number of babies that move about in an essentially random fashion; the nursemaid has control of a robot “hand” with which she can pick up and transport babies. The babies can be injured by falling into a ditch or by "thug" babies which act violently when in a crowded room. They can also fall ill, either to a contagious or a non-contagious illness. The nursemaid’s primary role is to keep these babies alive and well; in addition she has to remove babies from the nursery when they reach a certain age.

3.1 Agent requirements

There are several features of the real world that are also present in the nursery and generate requirements for the nursemaid’s design:

- Multiple goal processing – people try to satisfy many (often conflicting) goals in everyday life, such as for instance earning money to subsist and also taking time to relax and engaging in pleasurable pastimes. The nursemaid also has to process multiple goals, such as recharging a baby whose charge is running low and rescuing a baby who has fallen into a ditch.

- Complexity in the world – the real world is full of all kinds of different agents and objects; yet people can cope with this complexity and, in general, identify which agents or objects have or could have a bearing on the likelihood of satisfaction of one of the agent’s current goals. The nursery is through necessity a simple world, but complex enough to provide a challenging environment for the nursemaid to learn in and specifically to make the question of what stimuli to consider for learning nontrivial.

- Dynamic nature of the world – the state of the world is ever-changing and so has to be constantly monitored for goal-relevant changes. The nursery is dynamic in that there are several agents in the world (babies and perhaps others) whose behaviour is not predictable, forcing the nursemaid to constantly monitor the state of the nursery and in particular monitor the babies and other agents for changes that facilitate or inhibit the satisfaction of the nursemaid’s goals.

- Physical constraints – just as people are physically constrained by their bodies (e.g. the position and range of the arms and hands and the ability to look at only 100 degrees out of 360 of the world), so is the nursemaid constrained by the ability only to see a portion of the nursery at any time and the control of a single hand in the nursery with which to change its state.

These features serve to generate certain requirements for the nursemaid’s design; the nursemaid will generally have to make decisions on the basis of incomplete knowledge about the world, will often be constrained by the time pressures inherent in a situation and will therefore have to make quick decisions, and will also have to make decisions about what actions to perform (given that all not all possible actions can be performed by a single hand). In addition, there will be cases where concurrent stimuli signal impending disasters or opportunities for achieving goals, necessitating the association of combinations of features with the disaster/opportunity.

I do not intend to simulate in detail all manner of cognitive processes that might be found in people; for instance, the nursemaid’s visual perception system makes simplifying assumptions about the visual world as I am only interested in visual perception insofar as it bears on attentional and learning processes.

4 The agent design

The "broad and shallow" approach (Bates, Loyall & Reilly, 1991) taken in the design of the agent architecture means that it is composed of several functionally diverse processing modules which act in parallel (actually simulated parallelism via update in timesteps), but which are not implemented in great detail.

4.1 The overall architecture

The core of the model is a set of interleaved production systems with shared access to short-term memory components; these memory components have a loose hierarchical structure (stimuli, goals and action plans are stored in the short-term memory; stimuli can trigger goals, and goals can trigger plans) and so the system could be described essentially as a form of parallel blackboard system. In addition there are specific modules for simple perceptual and effector processing (in the form of communication from the nursemaid to the hand).

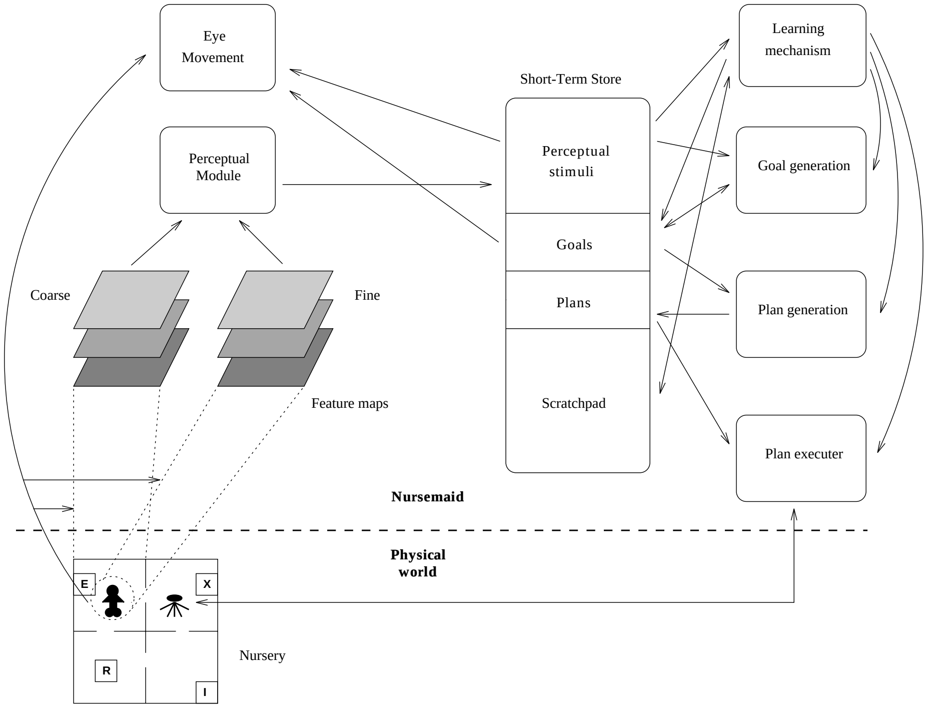

Figure 1: The basic agent architecture

Short-term store (STS) This stores perceptual, goal and plan information; also has a scratch-pad for other information (primarily for learning). This has a limited storage capacity as per working memory components in traditional production systems:

a The short-term memory is stored eventually in a physical structure, and therefore must have a storage limit;

b The capacity limit reduces the space to be searched: remember that the world can be used as a detailed long term store for world information, and that indexical representation is being implemented (only objects in the world which are relevant to the agent's current goals are taken account of). In addition, limited storage forces selection of "interesting/important" data.

c Much empirical work on short-term memory and on language comprehension has indicated that people have a very limited short-term memory span.

Feature maps There are both coarse-grained and fine-grained feature maps which represent specific low-level visual features such as specific colours and line segments as per feature integration theory based on visual search experiments (Treisman, 1986); in addition there is a hierarchical structure of feature maps, maps at a higher level integrating information from lower level maps. This builds up to object recognition; this mechanism is very over-simplified, but as with all the components in the architecture, gross simplifications are necessary in order to reduce the implementation task so that only our central problem need be addressed. There is a similar (but even cruder) setup for sound processing.

Perceptual module This takes information from these feature maps and places positive instances in the short-term store; information tags are assigned to visual 'objects' at this stage. In addition, perceptual features that have changed in some way are flagged for possible priority in further processing.

Visual guidance This controls the current locus of visual processing (both coarse and fine resolution) at any given moment; this can be controlled either bottom-up (stimulus-driven) or top-down (goal-driven).

Goal triggering This is a production memory that triggers goals based on perceptual and existing goal information in the STS. Several such goal rules are prestored in the module and are assumed to be innate.

Plan triggering This is another production memory that triggers action plans (serial list of atomic actions to be sent to the hand for execution) based on goal information in the STS. Again, several such plan rules are pre-programmed and are assumed to be innate.

Plan selection/execution Although the nursemaid can consider several competing plans, only one can be executed at a time (there is only one effector); this module thus selects a single action plan (based on computed priority values) and then passes 'atomic' actions serially to the hand. When one action is successful (detected by the hand), then next action in the plan is sent. The plans are interruptible; a new more 'pressing' plan can displace the one adopted for execution, although a biasing value to prefer the currently executed plan is included.

4.2 The learning mechanism

The learning mechanism is based on the rule-based model of classical conditioning described by Holyoak, Koh & Nisbett (1989) and is implemented in the form of a production system with interleaved production memories and several function-specific short-term memory stores (using Poplog Pop-11). There are five rule bases integral to the classical conditioning mechanism:

1. Learning rules – these determine when and how learning should happen, i.e. when to create a new rule and when to change the weight of an existing rule.

2. Selection rules – these define the basis on which stimuli are selected/filtered out for/from consideration for learning about (heuristics).

3. Prediction rules – these define which stimuli or combinations of stimuli are expected to predict the occurrence/absence of an unconditioned stimulus (reinforced positively or negatively) and which then allow evaluation of 'surprising' (unpredicted) stimuli (or absence thereof).

4. Unconditioned stimulus rules – these define which stimuli are unconditioned stimuli (USs) and therefore elicit positive or negative reinforcement (triggering unconditioned responses); in addition they store the reinforcement strengths associated with each of these USs.

5. Familiarity rules – these rules encode the familiarity of stimuli over a period of time beyond the life of information in the short-term memory

The short-term memory in the production system stores information about the state of the world (stimuli) and “unusual” events, as well as other results of triggered production rules. In addition to these five rule sets, I am initially using two rule sets to represent stimulus-response rules (agent behaviour) and world changes.

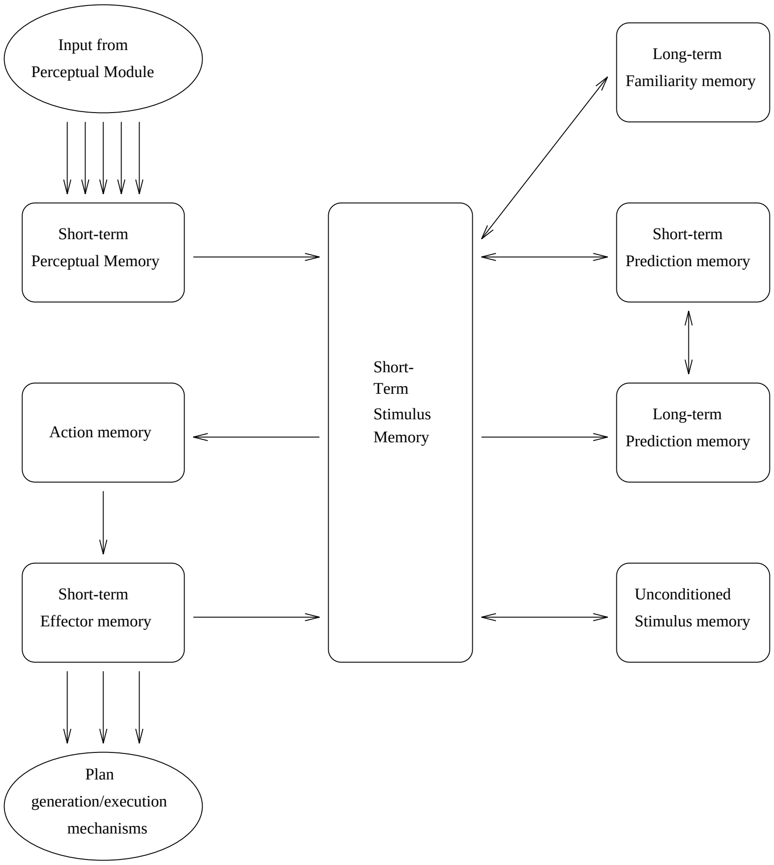

Figure 2: Data flow diagram for the classical conditioning mechanism

Selection Rules In this relatively simple reinforcement learning mechanism there is a rule set defining the following heuristic selection rules (based on the animal conditioning literature):

- Representative heuristic – links stimuli that are similar in salient respects;
- Temporal causality – based on the fact that causes generally precede effects;
- Temporal contiguity – preferentially links stimuli that occur relatively close together in time;
- Unusualness – selects novel stimuli as predictors of later occurrences that require explanation.

Based on Holyoak, Koh & Nisbett (1989)

Without these (or possibly other) selection rules, there is no way of deciding on the basis of information likely to be available, which stimuli to include in new prediction rules (i.e. deciding which stimuli predict the US and therefore reinforcement). This reflects the forms of selection implicit in theories of “attention” which try to explain how the relevant stimuli are chosen for association in animal learning. Such theories include those of Mackintosh, which posits selection on the basis of previous predictive success), and Pearce & Hall, which suggests selection on the basis of novelty (Walker, 1987).

A lot of detailed AI work has been done on each of the topics involved such as planning, vision and machine learning, far more than could possibly be implemented in a single agent within a reasonable time span. However, few syntheses of these aspects have been done examining learning and selection processes. Utilising the design-based and approach to modelling allows a subset of these mechanisms to be implemented in order to produce valid results without falling prey to the danger of making untenable assumptions with respect to the problem tackled.

5 Results

In this section I discuss experimental testing of the learning mechanism. This allows the demonstration of both the combinatorial explosion problem that is inherent in unconstrained associative conditioning and also the ways in which different selection heuristic mechanisms can be implemented in order to prioritise or filter those sets of stimuli to be selectively processed as a subset of the possible stimulus combinations at any given time.

To this end, a set of learning mechanisms of incremental sophistication was tested against an increasingly complex set of conditioning phenomena in order to determine the minimal set-up of a learning mechanism required for each conditioning phenomenon. These mechanisms range from two simple unlimited look-up tables to selection mechanisms based on notions of novelty, past success and other relevant criteria as described in section 4.2. The classical conditioning scenarios tested mimic several of the phenomena mentioned in section 2.1:

1. Contiguous single CS–US association.
2. Single CS–US association over time period.
3. Multiple contiguous CS–US association.
4. Multiple CS–US association over time period.
5. Single and multiple CS–US association with multiple non-relevant stimuli

The levels of learning mechanism implemented are:

1. A simple look-up table of stimulus-stimulus associations;
2. A simple look-up table processing a randomly selected subset of perceived stimuli;
3. A look-up table coupled with a novelty filter and success weighting;
4. As for (3) but with the addition of temporal causation heuristics.

These mechanisms can then be tested over a series of conditioning tasks of incremental difficulty, based on phenomena reported in the animal conditioning literature. These phenomena are based on the following experimental variables:

- Number of stimuli that must co-occur/occur in a certain pattern in order to act as conditioned stimuli (CSs), predicting an unconditioned stimulus (US)

- Time elapsed between conditioned stimulus offset and unconditioned stimulus onset (if not temporally contiguous); also time period over which to keep records of stimulus occurrence e.g. time of stimulus onset and duration of stimulus - can depend on the type of stimulus
- Strength and valence (i.e. positively- or negatively reinforcing effect) of the US
- "Surprisingness" of US occurrence
- Presence or absence of US
- Number of possible responses (actions) in a given situation
- Number of stimulus features present in the agent's environment at any time - difficult to calculate empirically as it depends on the agent's modes of representation (including cross-modal associations of sensory stimuli)

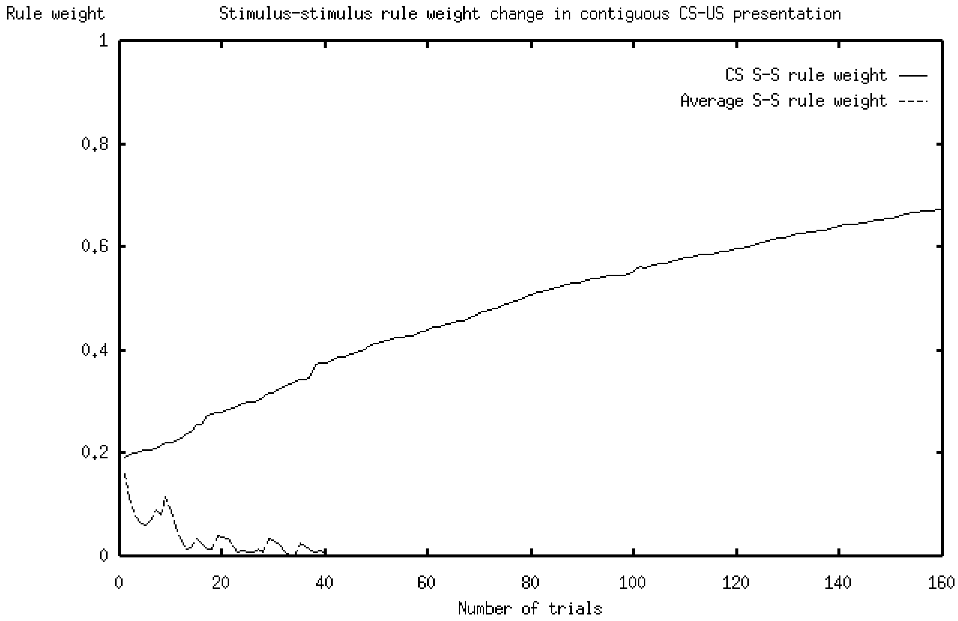

Figure 3: Results of the look-up table mechanism in the single contiguous CS-US condition

The look-up table mechanism converges on the correct solution in the first test scenario for small numbers of nonpredictive stimuli (see figure 3). However, when the numbers of stimuli are increased from 10 to 15 or 20, the mechanism proves extremely slow to perform the necessary computations as the number of stimulus-stimulus associations to be processed at each timestep increases exponentially with the number of stimuli. Given that the number of possible rules is 2^{n+1} where n is the total number of stimuli, this is to be expected. The random selection mechanism slows down the rate of learning to the point that it fails to learn to correctly predict unconditioned stimulus occurrence when the number of stimuli is either 15 or more. The simple look-up table mechanism is able to converge on the correct solution even for the more difficult scenarios as every possible combination of stimuli generated is checked/tested; however the number of trials needed increases to the order of thousands (from 100-150 for the first scenario).

Implementation of temporal and novelty heuristics to guide selective rule formation reduces the complexity of the problem as it drastically reduces the number of stimulus features considered upon unconditioned stimulus occurrence. If, as with the novelty filter, features are only considered if they pass a threshold weighted according to these selection heuristics, then the number of stimulus features to consider upon detection of US occurrence is typically a fraction of the total. Even though these heuristics can lead to incorrect credit assignment on certain trials, the mechanism converges on the correct association between conditioned and unconditioned stimuli within a few trials. The correct CS-US predictive rule is learned much faster than for the simple look-up table mechanism, especially when the number of stimuli to consider is greater than 10.

Addition of relatively simple temporal, success- and novelty-based selection heuristics to a look-up table thus significantly improves performance in terms not only of processing time required per timestep but also in terms of speed of convergence on the correct solution.

6 Discussion

Learning performance in a simple mechanism omitting the implicit selection mechanisms in Holyoak et al.'s classical conditioning model demonstrates the necessity of selection mechanisms in order to constrain the types of association learned; in addition, the nature of the real world requires that the agent architecture includes further explicit selection mechanisms in order to control information flow in the perception-action processing cycle and to enable negative feedback loops in the control of the field of vision and of physical effectors.

The design-based approach to a "broad and shallow" agent architecture gives a rigorous methodology in which to ground exploration of the interaction between selection and learning in intelligent agents. However the external validity of any results obtained from a model designed according to this approach are subject to criticism of the initial design requirements. There may well be other requirements that are important in the real-world problem that have not been included in the design requirements; equally well, the micro-world chosen for the model may not embody the essential characteristics of the real-world problem domain.

Further work with this model involves empirical testing of the hypothesis that reinforcement learning coupled with certain simple selection, monitoring and evaluation mechanisms can achieve several seemingly more complex forms of learning in a dynamic domain. This assertion is based on observations of the complexity of the forms of learning achievable observed in animal conditioning research; it would be interesting to see if these types of results could be replicated to some degree with this model. In addition, current work is being done to correlate the animal conditioning data with human psychological results as has been assumed within this paper.

References

Agre, P. E. & Chapman, D. (1987). Pengi: An implementation of a theory of situated action. In *Proceedings of AAAI-87*.

Allport, D. A. (1989). Visual attention. In Posner, M. I. (Ed.), *Foundations of Cognitive Science*. Cambridge, Mass: MIT Press.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, 48:57–86.

Ballard, D. H. (1992). Personal communication with Aaron Sloman.

Bates, J., Loyall, B., & Reilly, W. (1991). Broad agents. Notes for AAAI Spring Symposium on Integrated Agent Architectures. Also SIGART Bulletin, 2 (4), pages 38–40.

Beaudoin, L. (1994). *Goal Processing in Autonomous Agents*. PhD thesis, School of Computer Science, University of Birmingham, Edgbaston, Birmingham.

Campbell, J. (1982). *Grammatical Man: Information, Entropy, Language and LIfe*. Allen Lane.

Gallistel, C. R. (1990). *The Organisation of Learning*. Cambridge, MA: MIT Press.

Garcia, J., Erwin, F. R., & Koelling, R. A. (1966). Learning with prolonged delay of reinforcement. *Psychonomic Science*, 5:121–2.

Hinton, G. E. (1987). Connectionist learning procedures. Technical Report CMU-CS-87-115, Computer Science Department, Carnegie Mellon University. From Holyak, Koh & Nisbett (1989).

Holyoak, K. J., Koh, K., & Nisbett, R. E. (1989). A theory of conditioning: Inductive learning within rule- based default hierarchies. *Psychological Review*, 96(2):315–340.

Norman, D. A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In Davidson, R. J., Schwartz, G. E., & Shapiro, D. (Eds.), *Consciousness and Self-Regulation. Advances in research and theory*, volume 4, pages 1–18. New York: Plenum Press.

Pearce, J. M. (1987). *Introduction to Animal Cognition*. Lawrence Erlbaum Associates.

Pinker, S. (1990). Language acquisition. In Osherson, D. N. & Lasnik, H. (Eds.), *Invitation to Cognitive Science: Language*, volume 1, chapter 8, pages 199–241. Cambridge, MA: MIT Press.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. & Prokasy, W. F. (Eds.), *Classical conditioning II: Current research and theory*, pages 64–99. New York: Appleton-Century-Crofts.

Shannon, C. & Weaver, W. (1949). *The Mathematical Theory of Information*. Urbana: University of Illinois Press.

Sloman, A. (1994). Exploration in design space. In Cohn, A. (Ed.), *Proceedings of ECAI'94*. Wiley and Sons.

Thornton, C. J. (1992). *Techniques in Computational Learning: An Introduction*. London: Chapman & Hall.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, pages 106–115.

Tsotos, J. K. (1990). Analysing vision at the complexity level. *Behavioral and Brain Sciences*, 13:423–469.

Walker, S. F. (1987). *Animal Learning: An Introduction*. London: Routledge, Kegan and Paul.